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 #: 1891

De.2. **Measure Title:** Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization

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

De.3. **Brief Description of Measure:** The measure estimates a hospital-level 30-day, all-cause, risk-standardized readmission rate (RSRR) for patients discharged from the hospital with either a principal discharge diagnosis of COPD or a principal discharge diagnosis of respiratory failure with a secondary diagnosis of acute exacerbation of COPD. The outcome (readmission) is defined as unplanned readmission for any cause within 30 days of the discharge date for the index admission (the admission included in the measure cohort). A specified set of planned readmissions do not count in the readmission outcome. CMS annually reports the measure for patients who are 65 years or older, are enrolled in fee-for-service (FFS) Medicare, and hospitalized in non-federal hospitals.

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 30-day, all-cause, risk-standardized readmission rates following hospitalization for COPD. 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 their patient case mix, and therefore promote hospital quality improvement and better inform consumers about care quality.

COPD readmission is a priority area for outcomes 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 readmission. We define readmission as an inpatient admission for any cause, with the exception of certain planned readmissions, within 30 days from the date of discharge from the index admission for patients discharged from the hospital with a principal discharge diagnosis of COPD or principal discharge diagnosis of respiratory failure with a secondary discharge diagnosis of acute exacerbation of COPD. If a patient has more than one unplanned admission (for any reason) within 30 days after discharge from the index admission, only the first one is counted as a readmission. The measure looks for a dichotomous yes or no outcome of whether each admitted patient has an unplanned readmission within 30 days. However, if the first readmission after discharge is considered planned, any subsequent unplanned readmission is not counted as an outcome for that index admission because the unplanned readmission could be related to care provided during the intervening planned readmission rather than during the index admission.

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 40 years or older. We have explicitly tested the measure in both age groups.

The cohort includes admissions for patients discharged from the hospital with either a principal discharge diagnosis of COPD (see codes below) OR a principal discharge diagnosis of respiratory failure (see codes below) with a secondary discharge diagnosis of acute exacerbation of COPD (see codes below) and with a complete claims history for the 12 months prior to admission. The measure is currently publicly reported by CMS for those patients 65 years and older who are Medicare FFS beneficiaries admitted to non-federal hospitals.

Additional details are provided in S.9 Denominator Details.
S.10. **Denominator Exclusions:** The readmission measures exclude index admissions for patients:

1. Without at least 30 days post-discharge enrollment in FFS Medicare.
2. Discharged against medical advice (AMA);
3. Admitted within 30 days of a prior index admission.

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Administrative claims
S.26. **Level of Analysis:** Facility

IF Endorsement Maintenance – Original Endorsement Date: Mar 06, 2013 **Most Recent Endorsement Date:** Mar 06, 2013 (Pulmonary Project)

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 paired with a measure of hospital-level 30-day, all-cause, risk-standardized mortality (RSMR) following chronic obstructive pulmonary disease (COPD) hospitalization.

Maintenance of Endorsement -- 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. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

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.

Summary of evidence provided by the developer:.

- As a rationale for measuring this health outcome, the developer suggests that hospitals are able to influence readmission rates through a broad range of clinical activities including communication between providers, prevention of, and response to, complications, patient safety and coordinated transitions to the outpatient environment.
- The developer notes that there are no updates to the evidence since the last review.

Question for the Committee:

- *Does the Committee agree that the underlying rationale for the measure remains reasonable and there is no need for repeat discussion and vote on evidence?*

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.

- The developer provides performance data from four measurement periods, covering a total of 925,315 admissions.

- The data show that during the measurement period of 07/2011–06/2014, COPD readmission rates ranged from a minimum of 15.5% to a maximum of 26.6%, with the 10th percentile at 18.9%, the 50th percentile at 20.2%, and the 90th percentile at 21.7%.

Disparities

- To help in assessment of potential disparities, the developers also provide performance scores (using 2011-2014 data) for hospitals serving a low proportion of dual eligible patients vs. those serving a high proportion of dual eligible patients, performance scores for hospitals serving a low proportion of African-American patients vs. those serving a high proportion of African-American patients, and performance scores for hospitals serving a low proportion of patients with AHRQ SES Index Score index score equal to or below 45.9 vs. those serving a high proportion of patients with an AHRQ SES index score equal to or below 45.9.

- Hospitals serving a low proportion (=13.5%) Dual Eligible patients had a slightly lower median readmission rates (-0.4%) compared to hospitals serving a high proportion (=28.2%) Dual Eligible patients. Hospitals serving a low proportion (=0.0%) African-American patients had a slightly lower median readmissions rates (-0.4%) compared to hospitals serving a high proportion (=9.2%) African-American patients. Finally, hospitals serving a low proportion of patients below AHRQ SES index score of 45.0 had slightly lower median readmissions rates (-0.3%) compared to hospitals serving a high proportion of patients below AHRQ SRS index score of 45.0.

- By proportion of **Dual Eligible Patients**:

  // Low proportion (=13.5%) Dual Eligible patients//Hospitals with a high proportion (=28.2%) Dual Eligible patients
  Number of Measured Hospitals// 960 // 958
  Number of Patients// 241,848 patients in low-proportion hospitals/ 173,032 in high-proportion hospitals
  Maximum// 25.3// 26.6
  90th percentile// 21.5// 22.0
  75th percentile// 20.8 // 21.1
  Median (50th percentile)// 20.0// 20.4
  25th percentile// 19.4// 19.6
  10th percentile// 18.7// 19.1
  Minimum // 15.5 // 17.3

- By proportion of **African-American Patients**:

  // Low proportion (=0.0%) African-American patients//Hospitals with a high proportion (=9.2%) African-American patients
  Number of Measured Hospitals// 1,182 // 960
  Number of Patients// 119,954 patients in low-proportion hospitals/ 269,532in high-proportion hospitals
  Maximum// 25.1// 25.8
  90th percentile// 21.3// 22.2
  75th percentile// 20.7// 21.2
  Median (50th percentile)// 20.0// 20.4
  25th percentile// 19.4// 19.6
  10th percentile// 18.8// 19.0
  Minimum // 15.5// 16.7

- By Proportion of Patients with AHRQ SES Index Scores Equal or Below 45.9:

  // Low proportion of patients with AHRQ SES index score equal to or below 45.0 (=2.8%)// Hospitals with a high proportion of patients with AHRQ SES index score equal to or below 45.0 (=43.0%)
  Number of Measures Hospitals// 959 // 960
  Number of Patients// 210,243 patients in hospitals with low proportion of patients with AHRQ SES index score equal

to or below 45.0 //189,314 patients in hospitals with high proportion of patients with AHRQ SES index score equal to or below 45.0
Maximum// 25.3// 26.6
90th percentile// 21.9// 22.1
75th percentile// 20.9// 21.1
Median (50th percentile)// 20.1// 20.4
25th percentile// 19.4// 19.7
10th percentile// 18.8// 19.1
Minimum // 16.9// 16.8

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)

1. Importance to Measure and Report

1a. Evidence to Support Measure Focus

Comments: **Health outcome: Hospital 30-day, all-cause, risk standardized readmission rate following COPD hospitalization. Multiple aspects of care such as communication among providers, prevention of and response to complications, patient safety and coordination of care during transitions contribute to the outcome measure.

1b. Performance Gap

Comments: **During the measurement period of 7/11-6/14, COPD readmission rates ranged from 15.5% to 26.6%. 10th percentile 18.9%; 50th 20.2%; 90th at 21.7%.

Hospitals serving a low proportion (=13.5%) Dual Eligible patients had a slightly lower median readmission rates (-0.4%) compared to hospitals serving a high proportion (=28.2%) Dual Eligible patients. Hospitals serving a low proportion (=0.0%) African-American patients had a slightly lower median readmissions rates (-0.4%) compared to hospitals serving a high proportion (=9.2%) African-American patients. Finally, hospitals serving a low proportion of patients below AHRQ SES index score of 45.0 had slightly lower median readmissions rates (-0.3%) compared to hospitals serving a high proportion of patients below AHRQ SRS index score of 45.0.

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

Comments: **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.

- This measure calculates 30-day readmissions for patients hospitalized with chronic obstructive pulmonary disorder (COPD).
- The measure produces a risk-standardized readmission rate (RSRR), which is calculated as the ratio of the number of "predicted" to the number of "expected" readmission at a given hospital, multiplied by the national observed readmission rate.
- This claims-based measure can be used in either of two patient cohorts: (1) patients aged 65 years or older or (2) patients aged 40 years or older.

- The numerator includes patients who were readmitted for any cause, with the exception of certain planned readmissions, within 30 days from the date of discharge from the index COPD admission.
- The denominator population is defined using ICD-9 and ICD-10 codes; a list of applicable codes is included in the submission.
- The data sources for this measure may include Medicare Part A and B claims, the Medicare Enrollment Database (EDB),the American Community Survey, and all-payer data sources such as the California Patient Discharge Database.
- The measure's time window can be specified from one to three years.
- The measure is risk-adjusted using a statistical risk model (see details below).

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.

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

- Datasets used for testing included Medicare Parts A and B claims as well as the Medicare Enrollment Database (EDB). Additionally, census data were used to assess socio-demographic factors.
- **Data element reliability**:
  - With regard to data element reliability, the developer notes that the measure has been developed to avoid the use of claims data elements that are thought to be coded inconsistently across hospitals or providers, instead using fields that are consequential for payment and which are audited by CMS.
  - In addition, the developer compared frequencies and odds ratios of variables from their risk model across three years of data in order to assess the consistency of those variables over time.
- **Performance score reliability:**
  - The developer defines performance score reliability as the degree to which repeated measurements of the same entity agree with each other.
  - In line with this thinking, the developer's approach to assessing score-level reliability was to consider the extent to which assessments of a hospital using different but randomly-selected subsets of patients produce similar measures of hospital performance. The developers refer to this as a "test-retest" approach; it may also be called a "split-half" method. This is generally considered to be an appropriate method of testing reliability.

Results of reliability testing

- **Data element reliability**:
  - Summarizing the results of this analysis, the developer notes that the frequency of some model variables increased and others decreased between 2011 and 2014, which may reflect an increased or decreased rate of specific comorbidities in the FFS population.
  - The developer states that examination of the odds ratios for each risk variable in the model shows that, overall, the odds ratios for individual risk variables remained relatively constant across the three years.
  - The developer interprets the stability of the risk factor odds ratios over time as suggesting that the underlying data elements are reliable.

- **Performance score reliability:**
    - A total of 925,315 admissions over a 3-year period were examined, with 461,505 in one sample and 463,810 in the other randomly-selected sample. Two risk-standardized readmission rates (RSRR) were calculated for each hospital: one from each of the two separate samples.
        - The agreement between the two RSRRs for each hospital (as measured by an intra-class correlation coefficient (ICC)) was 0.48; the developer states that according to the conventional interpretation, this is considered a "moderate" level of agreement.
        - The developer notes that this analysis was limited to hospitals with 12 or more cases in each split sample, and that splitting the total population into two samples resulted in a sample equivalent of only 1.5 years of data, whereas the measure is reported with the full three years of data. [Note: It is unclear whether the measure itself is limited to hospitals with 12 or more cases; if it is not, then testing was not consistent with the measure as specified.]

Guidance from the Reliability Algorithm

- Question 1. Submitted specifications are precise, unambiguous, and complete. Measure can be consistently implemented.
- Question 2. Empirical reliability testing was conducted using statistical tests with the measure as specified.
- Question 3. Empirical validity testing of patient-level data was conducted.
- Question 4. Reliability testing was conducted with computed performance measure scores for each measured entity.
- Question 5. Random split-half correlation was used to assess the proportion of variability due to real differences among the measured entities.
- Question 6. The ICC was 0.48 which is considered a moderate level of agreement.

Questions for the Committee:

- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*
- *Does the measure testing match the measure specifications?*

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**

- This measure estimates 30-day all-cause readmission rates for patients discharged from an acute care hospital with a diagnosis of COPD using a risk-standardized readmission rate (RSRR), which is calculated as the ratio of the number of "predicted" to the number of "expected" readmission at a given hospital, multiplied by the national observed readmission rate
- As a rationale for measuring this health outcome, the developers suggest that hospitals are able to influence readmission rates through a broad range of clinical activities, including prevention of complications, improving communication among providers involved at care transition, discharge planning, management of care transitions, patient education, and encouraging strategies that promote disease management.

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 developer demonstrated measure validity through prior validity testing done on their claims-based measures, through use of established measure development guidelines, and by systematic assessment of measure face validity by a Technical Expert Panel (TEP).
    - Validity of Claims-Based Measures:
        - The developer states that they have demonstrated for a number of other readmission measures the validity of claims-based measures by comparing either the measure result or the individual data elements against medical records.
        - Claims model validation was conducted by building comparable models using abstracted medical chart data for risk adjustment. 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.
    - Validity Indicated by Established Measure Development Guidelines
        - The developer states that this measure was developed in consultation with national guidelines for publicly reported outcomes measures, with outside experts, and with the public.
    - Validity as Assessed by External Groups:
        - Input was obtained through regular discussions with an advisory working group, a TEP, and a 30-day public comment period.
    - Face Validity as Determined by TEP:
        - The developer asked members of the TEP to note their agreement that the RSRRs obtained from the COPD readmission measure will provide an accurate reflection of quality.
        - Of the TEP members who responded, 90% agreed (70% moderately or strongly agreed) that the measure will provide an accurate reflection of quality.
        - The developer interpreted this as a moderate level of agreement.
- The developer assessed the areas under the receiver operating characteristic (ROC) curve for the two models, the predictive ability comparing readmission rates in the lowest predicted decile and the highest predicted decile. However, this is generally considered to be a test of reliability rather than validity.
    - The developer notes the performance of the development and validation samples were similar and areas under the receiver operating characteristic (ROC) curve were 0.627 and 0.629, respectively, for the two models.

Describe any updates to validity testing

- The developer did additional testing to convert the measure from ICD-9 to ICD-10 codes.
- The goal was to convert this measure to a new code set, fully consistent with the intent of the original measure. Details of the conversion process are noted below.

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**

Questions for the Committee:

- *Do the empirical results from other readmissions measures demonstrate sufficient validity so that conclusions about quality can be made for this measure?*
- *Do the face validity results sufficiently demonstrate the validity of the measure?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- Patients in the following categories are excluded from the measure:
  - Without at least 30 days post-discharge enrollment in FFS Medicare.
  - Discharged against medical advice (AMA);
  - COPD admission within 30 days of a prior COPD index admission
- To determine the impact of exclusions, the developer examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion.
- The number and percentage of patients excluded for each criterion are as follows:
  1. Without at least 30 days post-discharge enrollment in FFS Medicare for index admissions: **5,173 (.51%)**
  2. Discharged against medical advice (AMA): **5,966 (.59%)**
  3. COPD admission within 30 days of a prior COPD index admission: **75,166 (7.43%)**

Questions for the Committee:

- *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: **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 measure employs a hierarchical logistic regression model (a form of hierarchical generalized linear model [HGLM]) to create a hospital-level 30-day risk-standardized readmission rate (RSRR).
- The developer suggests that 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.
- The developer notes that this approach simultaneously models data at the patient and hospital levels to account for the variance in patient outcomes both within and between hospitals.
- Variables considered for inclusion in the model 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 covariates are defined using condition categories (CCs), which are clinically-meaningful groupings of more than 15,000 ICD-9-CM diagnosis codes.
- The measure does not adjust for CCs that were possible adverse events of care and that were only recorded in the index admission.
- The final set of 41 risk-adjustment variables is included in the testing attachment; the odds ratio associated with each variable is also provided.

- The developers also considered a number of variables related to sociodemographic status (SDS) for potential inclusion in the risk-adjustment model. Candidate SDS variables were selected for examination based on a review of literature and national data sources.
- **Conceptual analysis of the need for SDS adjustment:**
    - The developers note there is a large body of literature linking various SES factors and African-American race to worse health status and higher readmission risk with income, education, and occupational level being the most commonly examined variables. The developers state that the literature directly examining how SES factors or race might influence the likelihood of older, insured, Medicare patient of being readmitted within 30 days of an admission for heart failure is more limited.
    - The developers state that few studies directly address causal pathways for SDS factors to affect 30-day readmission rates or examine the role of the hospital in these pathways.
    - There are at least four potential pathways for SDS factors to affect 30-day readmission rates:
        - One potential pathway is the relationship to health status at the time of admission. SDS factors 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. The developers note that this pathway should be largely accounted for by their clinical risk-adjustment model.
        - The next potential path way is that patients with low income and African-American patient are more likely to be seen in lower quality hospitals, which can contribute to increased risk of readmission.
        - The third major pathway is that a patient's race or SDS status cause them to experience differential, lower quality care or may not receive the differentiated care they require.
        - Finally, some SES risk factors may affect the likelihood of readmission without directly affecting health status at admission or the quality of care received during the hospitalization. Patients may have worse outcomes due to competing economic priorities or a lack of access to care outside the hospital.
- **Empirical analysis of SDS factors:**
    - The developers considered African-American race, dual-eligible status-i.e. enrolled in both Medicare and Medicaid, and AHRQ SES index score. The developers assessed the relationship between the SES variables and race with the outcome and examined the incremental effect in a multivariable mode.
    - The developer stated that they examined all patient-level indicators of both SES and race/ethnicity that are reliably available for all Medicare beneficiaries and linkable to claims data and selected those that are most valid.
    - The developer noted that the AHRQ-validated SES index score is a widely-used variable that describes the average socioeconomic status of people living in defined geographic areas. The developer notes that its value as a proxy for patient-level SDS is depend on having the most granular level data.
        - These variables are linked to patients by zip code and census block; however, the data are only linked at a 5-digit zip code level—nine-digit zip code data, which may provide a more granular view of patient sociodemographic status, were not available.
        - However, the developers note they are currently performing analyses at the census block level (the most granular level possible in this dataset) and hope to present the results of this analysis to the committee.
    - The developer assessed the relationship between the SDS variables and the 30-day COPD readmission rate and examined the incremental effect of SDS in a multivariable model, evaluating the extent to which the addition of any one of these variables improved model performance or changed hospital results.
    - The developer notes that 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. Therefore, the developers performed a decomposition analysis to assess the independent effects of the SES and race variables at the patient level and the hospital level.

- The developers' analysis found that the prevalence of SDS factors in the COPD cohort does vary across measured entities.
- With regard to the empirical association of each SDS variable with the outcome (univariate), the analysis found that patient-level observed COPD readmission rate for dual-eligible patients was higher, at 22.8% compared with 19.6% for all other patients. The readmission rate for African-American patients was also higher at 22.1% compared with 20.1% for patients of all other races. Similarly the readmission rate for patients with an AHRQ SES index score equal to or below 45.0 was 20.9% compared with 20.0% for patients with an AHRQ SES index score above 45.0.
- With regard to the strength and significance of the SDS variables in the context of a multivariable model, the developers' analysis found that the effect size of each of these variables is small, the c-statistic (i.e., predictive value) is unchanged with the addition of any of these variables into the model, and the addition of any of these variables into the model has little to no effect on hospital performance.
    - The median absolute change in hospitals' RSRRs when adding a dual eligibility indicator is -0.005% (interquartile range [IQR] -0.027% – 0.032%, minimum -0.348% – maximum 0.213%) with a correlation coefficient between RSRRs for each hospital with and without dual eligibility added of 0.99888.
    - The median absolute change in hospitals' RSRRs when adding a race indicator is 0.007% (IQR -0.005% – 0.016%, minimum -0.305% – maximum 0.044%) with a correlation coefficient between RSRRs for each hospital with and without race added of 0.99973.
    - The median absolute change in hospitals' RSRRs when adding a low AHRQ Index of SES score indicator to the model is 0.017% (IQR -0.054% – 0.068%, minimum -1.209% – maximum 0.914%) with a correlation coefficient between RSRRs for each hospital with and without an indicator for a low AHRQ Index of SES score is 0. 9919.
- The developers state that the patient-level and hospital-level dual eligible, race, and low AHRQ SES Index effects were significantly associated with COPD readmission in the decomposition analysis. The developers note that if the dual eligible, race, or low AHRQ SES Index variables are 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 developers state that given these findings and complex pathways that could explain any relationship between SDS and readmission, they did not incorporate SDS variables into the measure.

- **Risk Model Diagnostics:**
    - To assess the overall performance of their risk-adjustment model, the developers computed three summary statistics, including:
        - *Area under the receiver operating characteristic (ROC) curve* (also known as a c-statistic, which measures the probability that the model's prediction of the outcome is better than chance)
        - *Predictive ability* (the model's ability to distinguish high-risk subjects from low-risk subjects)
        - *Over-fitting indices (model calibration)* (to ensure that the model is not only describing the relationship between predictive variables and outcome in the development dataset but also providing valid predictions in new patients)
    - For the current measure cohort, the findings from this analysis are as follows:
        - C-statistic: **0.64**
            - A c-statistic of 0.64 means that for 64% of all possible pairs of patients—one who was readmitted and one who was not—the model correctly assigned a higher probability to those who were readmitted. Generally, a c-statistic of at least 0.70 is considered

acceptable.
- The developers interpret this as 'fair' model discrimination.

- Predictive ability (lowest decile %, highest decile %): (**10.1%, 36.5%**)
  - The developers state that this indicates a wide range between the lowest decile and highest decile, indicating the ability to distinguish high-risk subjects from low-risk subjects.
- Overfitting indices (model calibration) [presented as (γ0, γ1)]:
  - The developer states that if the γ0 in the validation samples are substantially far from zero and the γ1 is substantially far from one, there is potential evidence of over-fitting. The calibration value of close to 0 at one end and close to 1 to the other end indicates good calibration of the model.
    - 1st half of split sample: Calibration: (-0.034, 0.970)
    - 2nd half of split sample: Calibration: (0.004, 0.994)

- The developer's overall interpretation of the results of their analysis is that the findings demonstrate the risk-adjustment model adequately controls for differences in patient characteristics (case mix).
- The developer also conducted additional analyses to determine whether the measure could be applied to a population of patients aged 18+ using all-payer data.
- The developers report that this testing was conducted prior to specifying the measure for patients age 40 and over. The developers note that cohort to age 40 and over, however is not likely to affect the results given that only 1.5% of patients were between the ages of 18 and 39 and they believe the measure can be applied to all-payer data for patients 40 and older.

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)*:

- For public reporting of this measure, CMS characterizes the uncertainty associated with the RSRR by estimating the 95% interval estimate.
- 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."
- The developer reports that for the performance period of July 2011-June 2014, the mean hospital RSRR was 20.2%, with a range of 15.5% to 26.6%. The interquartile range was 19.6-20.8%.
- Of 4,663 hospitals in the study cohort, 83 performed "better than the U.S. national rate," 823 performed "no different from the U.S. national rate," 133performed "worse than the U.S. national rate," and 779 were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing.
  - The developer's interpretation of this data is that the variation in rates and number of performance outliers suggests there remain differences in the quality of care received across hospitals for COPD that support measurement

Question for the Committee:

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

2b6. Comparability of data sources/methods:

- While the developer did not decide to include SDS variables in their final model, they did compare measure results with and without SDS adjustment.

2b7. Missing Data

- N/A

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)

2. Scientific Acceptability of Measure Properties

2a1. & 2b1. Specifications

Comments: **Specifications consistent with evidence.

2a2. Reliability Testing

Comments: **The agreement between the two RSRRs for each hospital (as measured by an intra-class correlation coefficient (ICC)) was 0.48; the developer states that according to the conventional interpretation, this is considered a "moderate" level of agreement.

 The developer notes that this analysis was limited to hospitals with 12 or more cases in each split sample, and that splitting the total population into two samples resulted in a sample equivalent of only 1.5 years of data, whereas the measure is reported with the full three years of data. [Note: It is unclear whether the measure itself is limited to hospitals with 12 or more cases; if it is not, then testing was not consistent with the measure as specified

2b2. Validity Testing

Comments: **Additional testing to convert the measure from ICD-9 to ICD-10 was done. Validity testing performed in measure score using face validity.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **Exclusions consistent with evidence.

Effect of SDS variables in a multivariable model is small.

For the current measure cohort C-stat =0.64

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 based on administrative claims data (e.g., DRG, ICD-9/10), which the developers note are routinely generated and collected as part of hospitals' billing processes.
- The developer indicates that all data elements are in defined fields in electronic claims.

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

Committee pre-evaluation comments
Criteria 3: Feasibility

3. Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **Measure is based on administrative claims which are routinely generated.

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

Publicly reported? ☒ **Yes** ☐ **No**

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

OR

Planned use in an accountability program? ☐ **Yes** ☐ **No**

Accountability program details

- Hospital Inpatient Quality Reporting (IQR) Program http://cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html and Hospital Readmission Reduction (HRRP) Program http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program.html

Improvement results

- The developer notes that there has been significant progress in the 30-day RSRR for COPD. The "median 30-day RSRR decreased by 1.4 absolute percentage points from July 2011-June 2012 (median RSRR: 20.9%) to July 2013-June 2014 (median RSRR: 19.5%). The median hospital RSRR from July 2011-June 2014 was 20.2% (IQR 19.6% - 20.8%)."

Feedback :

- During the 2012-2013 MAP review, MAP supported this measure for inclusion in the IQR and HRRP programs. The group noted that the measure addresses a high-impact condition not adequately addressed in the program measure set.

***Questions for the Committee**:*

- *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

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences
Comments: **Publicly reported and used in accountability program.

Criterion 5: Related and Competing Measures

Related or competing measures

- 0275: Chronic Obstructive Pulmonary Disease (COPD) or Asthma in Older Adults Admission Rate (PQI 5)

Harmonization

- Neither of these measures listed above are competing, so harmonization is not necessary.

Pre-meeting public and member comments

Comment by: Dr. Kathy Gans-Brangs, PhD
Organization: AstraZeneca
Comment May 02, 2016: NOTE: this comment is not related to a harmonization or competint measure issue.

- Suggest changing "Age-65 (years, continuous) for patients aged 65 or over cohorts; or Age (years, continuous) for patients aged 18 and over cohorts" to "Age-65 (years, continuous) for patients aged 65 or over cohorts; or Age (years, continuous) for patients aged 40 and over cohorts.".

Comment by: Ms. Elizabeth Godsey
Organization: Vizient, Inc.
Comment May 05, 2016: Vizient, Inc., the largest member-owned health care company in the country, is dedicated to serving members & customers through innovative data-driven solutions, expertise & collaborative opportunities that lead to improved patient outcomes & lower costs. For the readmission measures considered, CMS presented patient-level & hospital specific SES factor beta coefficients & p-values, yet overall model performance were not presented. We request the actual model performance results for model evaluation. For the AHRQ SES Index variable, we request further information on how the binary classification for a measure that ranges between 0-100 was determined & the impact of transforming into a binary representation vs. actual value had on the model performance. This detail along with the overall model performance information would provide the public with the necessary information to truly assess CMS's comment 'Given these findings & 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.' Regarding the complex pathways associated with 30-day readmissions as stated by CMS, we strongly ask CMS to entirely re-evaluate the utility of the 30-day measures. As stated by CMS, factors influencing readmissions are blurred between providers & patients 30-days post discharge resulting in a limited insights in how providers can improve care. We believe CMS's efforts to remove the planned readmissions PR4 logic is a strong step in true opportunity identification; however, more refinement is needed. We recommend a shorter, more actionable 7 day post-discharge readmission timeframe to pinpoint opportunities providers truly can influence & thus, mitigate many of SES confounding factors. The 7-day window provides clearer opportunities for patient stabilization & post-acute discharge planning which the 30-day window doesn't reflect. We recommend CMS provide a 7-day readmission risk adjustment for review. Also, the hospital wide readmission measure evaluates all readmissions within the 30-day window post inpatient discharge & considers readmit cases to also be eligible as the index admission; however, the condition specific measures evaluate only 1 readmit within the 30-day window & cannot be eligible as an index. We ask CMS for the rationale why the different approaches for the same measure as this adds unnecessary complexity which are impractical to manage. We recommend a consistent approach across all readmission measure calculations & recommend evaluating & counting all readmits that occur within the 30-day window so providers have a clear understanding of the # readmits are truly occurring. We support considering a readmit as an index for the next 30-day cycle to again, assist organizations in tracking & improving complete patient care.

Comment by: Ms. Elizabeth Godsey
Organization: Vizient, Inc.

Comment May 05, 2016: Vizient, Inc., the largest member-owned health care company in the country, is dedicated to serving members & customers through innovative data-driven solutions, expertise & collaborative opportunities that lead to improved patient outcomes & lower costs. Vizient reviewed the I-10 translations for respiratory failure that are currently included in the COPD definition and recommend expanding to include J9601, J9602, J9691 & J9692 for the I-9 code 51881 and J9621 and J9622 for the I-9 code 51884 respiratory failure (acute & chronic) with hypoxia and hypercapnia as per the GEMS 2015 mapping. In reviewing the algorithm for AHRQ CCS potentially planned procedure list, AHRQ CCS 169 is listed as exclusion criteria, but within ICD-10 CCS 169 does not exist. Vizient recommends CMS and NQF reviewing this criterion and provide the appropriate ICD-10 translations to address the debridement of wound; infection or burn procedure codes.

Comment by: Ms. Elizabeth Godsey
Organization: Vizient, Inc.
Comment May 05, 2016: Vizient, Inc., the largest member-owned health care company in the country, is dedicated to serving members & customers through innovative data-driven solutions, expertise & collaborative opportunities that lead to improved patient outcomes & lower costs. Vizient requests CMS to review & provide follow-up analysis on more applied/practical alternate modeling approaches to account for within & across hospital variation besides hierarchical modeling. While hierarchical modeling is a valid technique controlling for within & across hospital variation, the approach lacks a tangible, practical framework of an observed to expected ratio that hospitals need to drive patient care. The predicted to expected approach complicates the public's & provider's understanding of how the actual observed values impacts hospital performance. Through numerous member discussions, we heard repeatedly, Oh, you mean that number does really reflect my actual readmissions? How can I improve that number? Even more concerning is the focus the current measure places on improving documentation & coding rather than patient care. Currently, providers see the only direct way to improve the measure is through documentation & coding capture of co-morbidities which count toward the predicted & expected value calculations. We hope this was not the original intention of the measure & this misguided focus is simply an unintended artifact of an overly complicated modeling technique. We recommend analyzing & provide results comparing a model that uses hospital characteristics, such as teaching status or bed size to account for structural differences across hospitals & provide an observed to expected ratio which is much more meaningful for the public & providers. While in the past, CMS has commented they would not incorporate these features due to NQF restrictions; it is important to point out NQF has endorsed other risk adjustment models that incorporate these characteristics (NHSN) & consider these factors in the 30-day risk adjustment as well. Also, we would ask CMS & NQF to institute discrimination performance thresholds for the models given the importance these models bare on CMS's performance programs & public reporting. Currently, no model performs > 0.70, a standard considered fair-good practical performance threshold & while the c-stat does not fully evaluate the model, it certainly should require basic performance standards. Additionally, we ask CMS to provide performance statistics, like AIC, BIC & the Somers' D, Gamma & Tau-a association of predicted probabilities & observed counts for a more comprehensive assessment. Using these standards & model diagnostics, NQF can provide CMS with recommendations for improvement. Until minimum discrimination thresholds are instituted, we recommend NQF remove endorsement of the readmission measures.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 1891

Measure Title: Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: N/A

Date of Submission: 1/29/2016

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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: Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

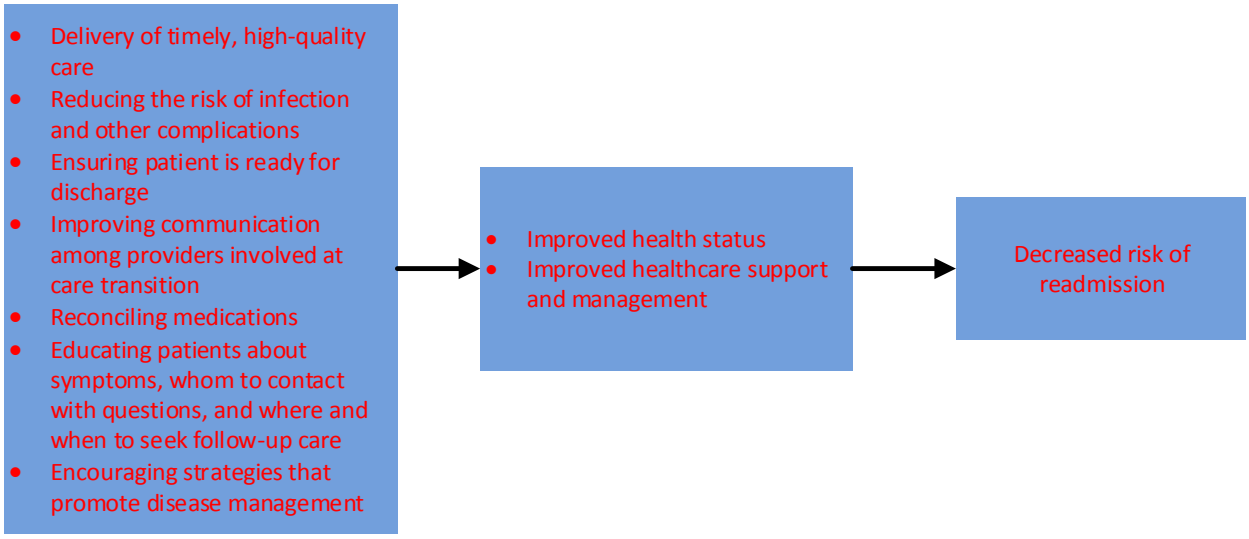
☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.



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 hospitalization for COPD. 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.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

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. Among these seven, COPD ranked fourth (MedPAC, 2007). COPD is a leading cause of readmissions to the hospital

(Jencks et al., 2009). The 30-day readmission rate among patients hospitalized for COPD, from 2003-2004, is 22.6%, accounting for 4% of all 30-day readmissions (Jencks et al., 2009).

The Agency for Health Research and Quality (AHRQ) has also identified COPD as an ambulatory-care-sensitive condition (ACSC). ACSCs are conditions for which good outpatient care can potentially prevent the need for hospitalization or for which early intervention can prevent complications or more severe disease (AHRQ, 2007). COPD is an ASCS that is associated with high readmission rates and high costs to Medicare (MedPAC, 2007). These facts underscore the need for developing strategies to reduce readmissions and subsequent costs associated with COPD admissions. COPD patients require ongoing care and treatment after discharge and are therefore at increased risk for readmission.

Although many current hospital interventions are known to decrease the risk of readmission within 30 days of hospital discharge (Leppin et al., 2014; Benbassat et al., 2000; Naylor et al., 1999; Coleman et al., 2006), current process-based performance measures, cannot capture all the ways that care within the hospital might influence outcomes. Measurement of patient outcomes allows for a comprehensive view of quality of care that reflects complex aspects of care, such as communication between providers and coordinated transitions to the outpatient environment. These aspects are critical to patient outcomes, and are broader than what can be captured by individual process-of-care measures.

The COPD hospital-specific, risk-standardized readmission rate (RSRR) measure is thus intended to inform quality-of-care improvement efforts, as individual process-based performance measures cannot encompass all the complex and critical aspects of care within a hospital that contribute to patient outcomes. As a result, many stakeholders, including patient organizations, are interested in outcomes measures that allow patients and providers to assess relative outcomes performance for hospitals (Krumholz et al., 2007). Improvement in inpatient care and care transitions for this common, costly condition are likely to reduce costly readmissions.

The diagram above indicates some of the many care processes that can influence readmission risk by improving health status or improving healthcare management and support. Early experience with care bundles suggests that that appropriate (guideline recommended care), high-quality, and timely treatment for COPD patients can reduce the risk of readmission within 30 days of hospital discharge (Hopkinson et al., 2012). Studies of integrated care management after hospitals discharge have suggested clinical benefit (Casas et al., 2006; Prieto-Centurion et al., 2014). Recent evidence of declining readmission rates provides further support for the concept that efforts to improve transitional care can affect a patient's risk of readmission.

References:

Agency for Healthcare Research and Quality (AHRQ) Quality Indicators. Guide to Prevention Quality Indicators. 2007. http://www.qualityindicators.ahrq.gov/Downloads/Modules/PQI/V31/pqi_guide_v31.pdf. Accessed November 16, 2015.

Benbassat J, Taragin M. Hospital readmissions as a measure of quality of health care: advantages and limitations. Arch Intern Med. 2000 Apr 24; 160(8):1074-81.

Casas A, Troosters T, Garcia-Aymerich J, et al. Integrated care prevents hospitalisations for exacerbations in COPD patients. Eur Respir J. 2006 Jul; 28(1):123-30. Epub 2006 Apr 12.

Coleman EA, Parry C, Chalmers S, Min SJ. The care transitions intervention: results of a randomized controlled trial. Arch Intern Med. 2006 Sep 25; 166(17):1822-8.

Hopkinson NS, Englebretsen C, Cooley N, et al. Designing and implementing a COPD discharge care bundle. Thorax. 2012; 67:90-92 doi: 10.1136/thoraxjnl-2011-200233.

Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare Fee-for-Service Program. N Engl J Med. 2009; 360(14):1418-28.

Krumholz HM, Normand SL, Spertus JA, et al. Measuring performance for treating heart attacks and heart failure: the case for outcomes measurement. Health Aff (Millwood). 2007 Jan-Feb; 26(1):75-85.

Leppin AL, Gionfriddo MR, Kessler M, et al. Preventing 30-day hospital readmissions: a systematic review and meta-analysis of randomized trials. JAMA Internal Med. 2014; 174(7):1095-107.

Medicare Payment Advisory Committee. Report to the Congress: Promoting Greater Efficiency in Medicare. Washington DC: Medicare Payment Advisory Commission (MedPAC); 2007 Jun 15. Accessed November 16, 2015.

Naylor MD, Brooten D, Campbell R, Jacobsen BS, Mezey MD, Pauly MV, Schwartz JS. Comprehensive discharge planning and home follow-up of hospitalized elders: a randomized clinical trial. JAMA. 1999 Feb 17; 281(7):613-20.

Prieto-Centurion V, Markos MA, Ramey NI, et al. Interventions to reduce rehospitalizations after chronic obstructive pulmonary disease exacerbations. A systematic review. Ann Am Thorac Soc. 2014 Mar; 11(3):417-24. doi: 10.1513/AnnalsATS.201308-254OC.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

N/A. This measure is not an intermediate outcome, process, or structure performance measure.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

N/A. This measure is not an intermediate outcome, process, or structure performance measure.

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

N/A

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

N/A

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

N/A

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

N/A

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

N/A

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

N/A

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

N/A

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

N/A

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

N/A

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

N/A

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

N/A

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

N/A

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

N/A

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

N/A

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

N/A

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

N/A

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

N/A

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

N/A

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

N/A

1a.8 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.8.1 What process was used to identify the evidence?

N/A

1a.8.2. Provide the citation and summary for each piece of evidence.

N/A

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

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all 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_1891_COPD_Readmission_NQF_Evidence_Attachment_02-15-16_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 30-day, all-cause, risk-standardized readmission rates following hospitalization for COPD. 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 their patient case mix, and therefore promote hospital quality improvement and better inform consumers about care quality.

COPD readmission is a priority area for outcomes 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.*

Distribution of Hospital COPD RSRRs over Different Time Periods
Results for each data year
Characteristic//07/2011-06/2012//07/2012-06/2013//07/2013-06/2014//07-2011-06/2014
Number of Hospitals// 4,541 // 4,515 // 4,495 // 4,663
Number of Admissions// 318,209 // 326,867 // 280,239 // 925,315
Mean (SD)// 21.0 (0.8) // 20.1 (0.9) // 19.5 (0.7) // 20.3 (1.2)
Range (min. – max.)// 17.9-26.2 // 16.6-25.3 // 16.9-24.5 // 15.5-26.6
Minimum// 17.9 // 16.6 // 16.9 // 15.5
10th percentile// 20.1 // 19.1 // 18.7 // 18.9
20th percentile// 20.4 // 19.5 // 19.0 // 19.4
30th percentile// 20.6 // 19.8 // 19.2 // 19.7
40th percentile// 20.8 // 19.9 // 19.4 // 20.0
50th percentile// 20.9 // 20.1 // 19.5 // 20.2
60th percentile// 21.1 // 20.2 // 19.6 // 20.4
70th percentile// 21.2 // 20.4 // 19.8 // 20.6
80th percentile// 21.5 // 20.7 // 20.0 // 21.0
90th percentile// 22.0 // 21.2 // 20.4 // 21.7
Maximum// 26.2 // 25.3 // 24.5 // 26.6

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the**

literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.
N/A

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for 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.*
Distribution of COPD RSRRs by Proportion of Dual Eligible Patients:
Dates of Data: July 2011 through June 2014
Data Source: Medicare FFS claims

Characteristic//Hospitals with a low proportion (=13.5%) Dual Eligible patients//Hospitals with a high proportion (=28.2%) Dual Eligible patients
Number of Measured Hospitals// 960 // 958
Number of Patients// 241,848 patients in low-proportion hospitals/ 173,032 in high-proportion hospitals
Maximum// 25.3// 26.6
90th percentile// 21.5// 22.0
75th percentile// 20.8 // 21.1
Median (50th percentile)// 20.0// 20.4
25th percentile// 19.4// 19.6
10th percentile// 18.7// 19.1
Minimum // 15.5 // 17.3

Distribution of COPD RSRRs by Proportion of African-American Patients:
Dates of Data: July 2011 through June 2014
Data Source: Medicare FFS claims

Characteristic// Hospitals with a low proportion (=0.0%) African-American patients//Hospitals with a high proportion (=9.2%) African-American patients
Number of Measured Hospitals// 1,182 // 960
Number of Patients// 119,954 patients in low-proportion hospitals/ 269,532in high-proportion hospitals
Maximum// 25.1// 25.8
90th percentile// 21.3// 22.2
75th percentile// 20.7// 21.2
Median (50th percentile)// 20.0// 20.4
25th percentile// 19.4// 19.6
10th percentile// 18.8// 19.0
Minimum // 15.5// 16.7

Distribution of COPD RSRRs by Proportion of Patients with AHRQ SES Index Scores Equal to or Below 45.0:
Dates of Data: July 2011 through June 2014
Data Source: Medicare FFS claims and the American Community Survey (2008-2012) data

Characteristic//Hospitals with low proportion of patients with AHRQ SES index score equal to or below 45.0 (=2.8%)// Hospitals with a high proportion of patients with AHRQ SES index score equal to or below 45.0 (=43.0%)
Number of Measures Hospitals// 959 // 960
Number of Patients// 210,243 patients in hospitals with low proportion of patients with AHRQ SES index score equal to or below 45.0 //189,314 patients in hospitals with high proportion of patients with AHRQ SES index score equal to or below 45.0
Maximum// 25.3// 26.6
90th percentile// 21.9// 22.1
75th percentile// 20.9// 21.1
Median (50th percentile)// 20.1// 20.4
25th percentile// 19.4// 19.7
10th percentile// 18.8// 19.1

Minimum // 16.9// 16.8

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, 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.**
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. Among these seven, COPD ranked fourth (MedPAC, 2007). COPD is a leading cause of readmissions to the hospital (Jencks et al., 2009).

The 30-day readmission rate among patients hospitalized for COPD, from 2003-2004, is 22.6%, accounting for 4% of all 30-day readmissions (Jencks et al., 2009).

The Agency for Health Research and Quality (AHRQ) has also identified COPD as an ambulatory-care-sensitive condition (ACSC). ACSCs are conditions for which good outpatient care can potentially prevent the need for hospitalization or for which early intervention can prevent complications or more severe disease (AHRQ, 2007). COPD is an ACSC that is associated with high readmission rates and high costs to Medicare (MedPAC, 2007). These facts underscore the need for developing strategies to reduce readmissions and subsequent costs associated with COPD admissions. COPD patients require ongoing care and treatment after discharge and are therefore at increased risk for readmission.

A hospital-level 30-day, all-cause, readmission measure will inform healthcare providers about opportunities to improve care and strengthen incentives for quality improvement, particularly for care at the time of transitions (e.g., discharge to home or a skilled nursing facility). Improvements to inpatient care and care transitions for this common condition are likely to reduce costly readmissions.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
Agency for Healthcare Research and Quality (AHRQ) Quality Indicators. Guide to Prevention Quality Indicators. 2007. http://www.qualityindicators.ahrq.gov/Downloads/Modules/PQI/V31/pqi_guide_v31.pdf. Accessed November 16, 2015.

Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. N Engl J Med. 2009; 360(14):1418-28.

Medicare Payment Advisory Committee. Report to the Congress: Promoting Greater Efficiency in Medicare. Washington DC: Medicare Payment Advisory Commission (MedPAC); 2007 Jun 15. Accessed November 16, 2015.

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):*
Pulmonary/Critical Care, Pulmonary/Critical Care : Chronic Obstructive Pulmonary Disease (COPD), Pulmonary/Critical Care : Dyspnea

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Care Coordination : Readmissions, Safety, Safety : Complications

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:** NQF_1891_COPD_Readmission_S2b_Readmission_Data_Dictionary_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.
Annual Updates
1. Updated CC map.

Rationale: The ICD-9-CM CC map is updated annually to capture all relevant comorbidities coded in patient administrative claims data.

No other updates or changes have been made since the last endorsement except for use of new years of data for public reporting.

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 readmission. We define readmission as an inpatient admission for any cause, with the exception of certain planned readmissions, within 30 days from the date of discharge from the index admission for patients discharged from the hospital with a principal discharge diagnosis of COPD or principal discharge diagnosis of respiratory failure with a secondary discharge diagnosis of acute exacerbation of COPD. If a patient has more than one unplanned admission (for any reason) within 30 days after discharge from the index admission, only the first one is counted as a readmission. The measure looks for a dichotomous yes or no outcome of whether each admitted patient has an unplanned readmission within 30 days. However, if the first readmission after discharge is considered planned, any subsequent unplanned readmission is not counted as an outcome for that index admission because the unplanned readmission could be related to care provided during the intervening planned readmission rather than during the index admission.

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 COPD hospitalization.

Denominator Time Window: This measure was developed with 12 months of data. 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.
The measure counts readmissions to any acute care hospital for any cause within 30 days of the date of discharge of the index COPD admission, excluding planned readmissions as defined below.

Planned Readmission Algorithm (Version 3.0)
The Planned Readmission Algorithm is a set of criteria for classifying readmissions as planned among the general Medicare population using Medicare administrative claims data. The algorithm identifies admissions that are typically planned and may occur within 30 days of discharge from the hospital.

The Planned Readmission Algorithm has three fundamental principles:

1. A few specific, limited types of care are always considered planned (obstetric delivery, transplant surgery, maintenance chemotherapy/ immunotherapy, rehabilitation);
2. Otherwise, a planned readmission is defined as a non-acute readmission for a scheduled procedure; and
3. Admissions for acute illness or for complications of care are never planned.

The algorithm was developed in 2011 as part of the Hospital-Wide Readmission measure. In 2013, CMS applied the algorithm to its other readmission measures. In applying the algorithm to condition- and procedure-specific measures, teams of clinical experts reviewed the algorithm in the context of each measure-specific patient cohort and, where clinically indicated, adapted the content of the algorithm to better reflect the likely clinical experience of each measure's patient cohort. For the COPD readmission measure, CMS used the Planned Readmission Algorithm without making any changes.

The Planned Readmission Algorithm and associated code tables are attached in data field S.2b (Data Dictionary or Code Table).

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 40 years or older. We have explicitly tested the measure in both age groups.

The cohort includes admissions for patients discharged from the hospital with either a principal discharge diagnosis of COPD (see codes below) OR a principal discharge diagnosis of respiratory failure (see codes below) with a secondary discharge diagnosis of acute exacerbation of COPD (see codes below) and with a complete claims history for the 12 months prior to admission. The measure is currently publicly reported by CMS for those patients 65 years and older who are Medicare FFS beneficiaries admitted to non-federal hospitals.

Additional details are provided in S.9 Denominator Details.

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

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)*
To be included in the measure cohort used in public reporting, patients must meet the following inclusion criteria:

1. Principal discharge diagnosis of COPD or principal discharge diagnosis of respiratory failure with a secondary discharge diagnosis of COPD with exacerbation
2. Enrolled in Medicare fee-for-service (FFS)
3. Aged 65 or over
4. Discharged alive from a non-federal acute care hospital
5. Not transferred from another acute care facility

6. Enrolled in Part A and Part B Medicare for the 12 months prior to the date of admission, and enrolled in Part A during the index admission.

This measure can also be used for an all-payer population aged 40 years and older. We have explicitly tested the measure in both patients aged 40 years and older and those aged 65 years or older (see Testing Attachment for details).

International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) codes used to define the cohort for each measure are:

ICD-9-CM codes used to define COPD:
491.21 Obstructive chronic bronchitis with (acute) exacerbation
491.22 Obstructive chronic bronchitis with acute bronchitis
491.8 Other chronic bronchitis
491.9 Unspecified chronic bronchitis
492.8 Other emphysema
493.20 Chronic obstructive asthma, unspecified
493.21 Chronic obstructive asthma with status asthmaticus
493.22 Chronic obstructive asthma with (acute) exacerbation
496 Chronic airway obstruction, not elsewhere classified
518.81 Acute respiratory failure (Principal diagnosis when combined with a secondary diagnosis of COPD with exacerbation [491.21, 491.22, 493.21, or 493.22])
518.82 Other pulmonary insufficiency, not elsewhere classified (Principal diagnosis when combined with a secondary diagnosis of COPD with exacerbation [491.21, 491.22, 493.21, or 493.22])
518.84 Acute and chronic respiratory failure (Principal diagnosis when combined with a secondary diagnosis of COPD with exacerbation [491.21, 491.22, 493.21, or 493.22])
799.1 Respiratory arrest (Principal diagnosis when combined with a secondary diagnosis of COPD with exacerbation [491.21, 491.22, 493.21, or 493.22])

ICD-9-CM codes used to define acute exacerbation of COPD:
491.21 Obstructive chronic bronchitis with (acute) exacerbation
491.22 Obstructive chronic bronchitis with acute bronchitis
493.21 Chronic obstructive asthma with status asthmaticus
493.22 Chronic obstructive asthma with (acute) exacerbation

ICD-10-CM codes used to define COPD:
J44.1 Chronic obstructive pulmonary disease with (acute) exacerbation
J44.0 Chronic obstructive pulmonary disease with acute lower respiratory infection
J41.8 Mixed simple and mucopurulent chronic bronchitis
J42 Unspecified chronic bronchitis
J43.9 Emphysema, unspecified
J44.9 Chronic obstructive pulmonary disease, unspecified
J96.00 Acute respiratory failure, unspecified whether with hypoxia or hypercapnia
J96.90 Respiratory failure, unspecified, unspecified whether with hypoxia or hypercapnia
J80 Acute respiratory distress syndrome
J96.20 Acute and chronic respiratory failure, unspecified whether with hypoxia or hypercapnia
R09.2 Respiratory arrest

ICD-10-CM codes used to define acute exacerbation of COPD:
J44.1 Chronic obstructive pulmonary disease with (acute) exacerbation
J44.0 Chronic obstructive pulmonary disease with acute low respiratory infection

An ICD-9 to ICD-10 crosswalk is attached in field S.2b. (Data Dictionary or Code Table).

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
The readmission measures exclude index admissions for patients:

1. Without at least 30 days post-discharge enrollment in FFS Medicare.
2. Discharged against medical advice (AMA);
3. Admitted within 30 days of a prior index admission.

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)*
1. Admissions without at least 30 days post-discharge enrollment in FFS Medicare are determined by examining the Medicare Enrollment Database (EDB).

2. Discharges against medical advice (AMA) are identified using the discharge disposition indicator in claims data.

3. COPD admissions within 30 days of discharge from a qualifying COPD index admission are identified by comparing the discharge date from the index admission with subsequent admission dates.

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 employs a hierarchical logistic regression model to create a hospital-level 30-day, all-cause, RSRR. In brief, the approach simultaneously models data at the patient and hospital levels to account for the variance in patient outcomes within and between hospitals (Normand & Shahian, 2007). At the patient level, the model adjusts the log-odds of readmission within 30 days of discharge for age and selected clinical covariates. At the hospital level, the approach models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of readmission at the hospital, after accounting for patient risk. If there were no differences among hospitals, then after adjusting for patient risk, the hospital intercepts should be identical across all hospitals.

Candidate and Final Risk-adjustment Variables: Candidate variables were patient-level risk-adjustors that were 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 claims records extending 12 months prior to and including the index admission. For the measure currently implemented by CMS, these risk-adjusters are identified using both inpatient and outpatient Medicare FFS claims data. However, in the all-payer hospital discharge database measure, the risk-adjustment variables can be obtained only from inpatient claims in the prior 12 months and the index admission.

The model adjusts for case-mix differences based on the clinical status of patients 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 (Pope et al., 2000). A file that contains a list of the ICD-9-CM codes and their groupings into CCs is attached in data field S.2b (Data Dictionary or Code Table). In addition, only comorbidities that convey information about the patient at admission or in the 12 months prior, and not complications that arise during the course of the index hospitalization, are included in the risk adjustment. Hence, we do not risk adjust for CCs that may represent adverse events of care when they are only recorded in the index admission.

The final set of risk adjustment variables is:

Demographics
Age-65 (years, continuous) for patients aged 65 or over cohorts; or Age (years, continuous) for patients aged 18 and over cohorts.

Comorbidities
History of mechanical ventilation (ICD-9 procedure codes: 93.90, 96.70, 96.71, 96.72)
Sleep apnea (ICD-9 diagnosis codes: 327.20, 327.21, 327.23, 327.27, 327.29, 780.51, 780.53, 780.57)
Respirator dependence/respiratory failure (CC 77-78)
Cardio-respiratory failure and shock (CC 79)
Congestive heart failure (CC 80)
Acute coronary syndrome (CC 81-82)
Chronic atherosclerosis or angina (CC 83-84)
Specified arrhythmias and other heart rhythm disorders (CC 92-93)
Other and unspecified heart disease (CC 94)
Vascular or circulatory disease (CC 104-106)
Fibrosis of lung and other chronic lung disorder (CC 109)
Pneumonia (CC 111-113)
History of infection (CC 1, 3-6)
Metastatic cancer and acute leukemia (CC 7)
Lung, upper digestive tract, and other severe cancers (CC 8)
Lymphatic, head and neck, brain, and other major cancers; breast, colorectal and other cancers and tumors; other respiratory and heart neoplasms (CC 9-11)
Other digestive and urinary neoplasms (CC 12)
Diabetes mellitus (DM) or DM complications (CC 15-20, 119-120)
Protein-calorie malnutrition (CC 21)
Disorders of fluid/electrolyte/acid-base (CC 22-23)
Other endocrine/metabolic/nutritional disorders (CC 24)
Pancreatic disease (CC 32)
Peptic ulcer, hemorrhage, other specified gastrointestinal disorders (CC 34)
Other gastrointestinal disorders (CC 36)
Severe hematological disorders (CC 44)
Iron deficiency and other/unspecified anemia and blood disease (CC 47)
Dementia or other specified brain disorders (CC 49-50)
Drug/alcohol psychosis or dependence (CC 51-52)
Major psychiatric disorders (CC 54-56)
Depression (CC 58)
Anxiety disorders (CC 59)
Other psychiatric disorders (CC 60)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)
Polyneuropathy (CC 71)
Stroke (CC 95-96)
Renal failure (CC 131)
Decubitus ulcer or chronic skin ulcer (CC 148-149)
Cellulitis, local skin infection (CC 152)
Vertebral fractures (CC 157)

References:
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.

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

Pope GC, et al. 2000. Principal Inpatient Diagnostic Cost Group Models for Medicare Risk Adjustment. Health Care Financing Review 21(3): 93-118.

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*)

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.)*
The measure estimates hospital-level 30-day, all-cause, RSRRs following hospitalization for COPD using hierarchical logistic regression models. In brief, the approach simultaneously models data at the patient and hospital levels to account for variance in patient outcomes within and between hospitals (Normand and Shahian, 2007). At the patient level, it models the log-odds of readmission within 30 days of discharge from the index admission using age, selected clinical covariates, and a hospital-specific intercept. At the hospital level, it models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of a readmission at the hospital, after accounting for patient risk. The hospital-specific intercepts are given a distribution 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.

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

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

This calculation transforms the ratio of predicted over expected into a rate that is compared to the national observed readmission rate. The hierarchical logistic regression models are described fully in the original methodology report (Grosso et al., 2011).

Reference:
Grosso L, Lindenauer P, Wang C, et al. Hospital-level 30-day Readmission Following Admission for an Acute Exacerbation of Chronic Obstructive Pulmonary Disease. 2011.

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

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.

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 survey or patient-reported data.

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.
Administrative claims

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:

1. 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.

2. 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).

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

4. 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 COPD 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/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_1891_COPD_Readmission_NQF_Testing_Attachment_01-29-16_v1.0.docx

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

Measure Number (*if previously endorsed*)**:** 1891
Measure Title: Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization
Date of Submission: 1/29/2016
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☒ Outcome (*including PRO-PM*)
☐ 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 as well as the Medicare Enrollment Database (EDB). Additionally, census as well as claims data were used to assess socioeconomic factors and race (dual eligible 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?

The dates used vary by testing type; see Section 1.7 for details.

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

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

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

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

For this measure, hospitals are the measured entities. All non-federal, acute care inpatient US hospitals (including territories) with Medicare fee-for-service (FFS) beneficiaries aged 65 years or over are included. The number of measured entities (hospitals) varies by testing type; see Section 1.7 for details.

1.6. How many and which patients 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 hospitals, and number of admissions used in each type of testing are as follows:

For reliability testing (Section 2a2)
The reliability of the model was tested by randomly selecting 50% of the Medicare patients aged 65 years or over from the most recent 3-year cohort and developing a risk-adjusted model for this group. We then developed a second model for the remaining 50% of patients and compared the two. Thus, for reliability testing, we randomly split **Dataset 1** into two samples. In each year of measure reevaluation, we also re-fit the model and compared the frequencies and model coefficients of risk variables (condition categories for patient comorbidities) and model fit across 3 years (**Dataset 1** below).

Dataset 1 (2015 public reporting cohort, version 4.0): Medicare Part A Inpatient and Outpatient and Part B Outpatient claims
Dates of Data: July 1, 2011 – June 30, 2014
Number of Admissions: 925,315
Patient Descriptive Characteristics: average age=76.8, %male=41.2
Number of Measured Hospitals: 4,663

For validity testing (Section 2b2): Medicare Part A Inpatient and Outpatient and Part B Outpatient claims
To create the model development and validation samples (**Dataset 2**), we applied the inclusion and exclusion criteria to all 2008 admissions. We randomly selected half of all COPD admissions in 2008 that met the inclusion and exclusions criteria to create a model development sample and used the remaining admissions as our model validation sample.

Dataset 2 (original measure development and validation samples)
Date of Data: 2008

First half of split sample (development sample)
-Number of Admissions: 176,480
-Number of Measured Hospitals: 4,546

Second half of split sample (validation sample)
-Number of Admissions: 176,151
-Number of Measured Hospitals: 4,553

For testing of measure exclusions (Section 2b3)
Dataset 1 (current public reporting cohort)

For testing of measure risk adjustment (Section 2b4)
Dataset 1 (current public reporting cohort)

Dataset 2 (development dataset): Medicare Part A Inpatient and Outpatient and Part B Outpatient claims

Dataset 3 (all payer dataset): California Patient Discharge Data

Dates of Data: January 1, 2006 – December 31, 2006
Number of Admissions: 45,480 (all 18+ total); 18,647 (FFS 65+); 11,014 (non-FFS 65+); 15,819 (all 18-64)
Patient Descriptive Characteristics: mean age=69 (all 18+ total); mean age=77 (FFS 65+); mean age=77 (non-FFS 65+); mean age=55 (all 18-64)
Number of Measured Hospitals: >450 non-Federal acute care hospitals

The measure was applied to California Patient Discharge Data, a large, linked all-payer database of patient hospital admissions. Records are linked by a unique patient identification number, allowing us to determine patient history from previous hospitalizations.

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

For testing of socioeconomic status (SES) factors and race in risk models (Section 2b4.3)
Dataset 1 and **Dataset 4**: 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 SES.

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 1**)
- Validated AHRQ SES index score is a composite of 7 different variables found in the census data (**Dataset 4**)

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, 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 a COPD readmission is much more limited (Elixhauser et al., 2008; Feemster and Au, 2014;

Prieto-Centurion et al., 2013; Sharma et al., 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 1**)
- African-American race (**Dataset 1**)
- 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 4**)

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 5-digit zip code and census block levels only. The data are not currently available at the 9-digit zip code level. In this submission, we present analysis using the 5-digit level. However, we are currently performing analysis at the census block level, the most granular level possible. We hope to present the results of the census block-level analysis to the committee.

References:

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.

Eapen ZJ, McCoy LA, Fonarow GC, et al. Utility of socioeconomic status in predicting 30-day outcomes after heart failure hospitalization. Circ Heart Fail. May 2015; 8(3):473-80.

Elixhauser A, Au DH, Podulka J. Readmissions for chronic obstructive pulmonary disease, 2008 [Internet]. Rockville, MD: Agency for Healthcare Research and Quality. HCUP Statistical Brief #121. Accessed: July 12 2013. Available at: http://www.hcup-us.ahrq.gov/reports/statbriefs/sb121.jsp.

Feemster LC, Au DH. Penalizing hospitals for chronic obstructive pulmonary disease readmissions. American journal of respiratory and critical care medicine. Mar 15 2014; 189(6):634-639.

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.

Prieto-Centurion V, Gussin HA, Rolle AJ, et al. Chronic obstructive pulmonary disease readmissions at minority-serving institutions. Annals of the American Thoracic Society. Dec 2013; 10(6):680-684.

Sharma G, Kuo YF, Freeman JL, et al. Outpatient follow-up visit and 30-day emergency department visit and readmission in patients hospitalized for chronic obstructive pulmonary disease. Archives of internal medicine. Oct 11 2010; 170(18):1664-1670.

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**)

The frequency of some model variables increased and others decreased between 2011 and 2014, which may reflect an increase or decrease rate of specific comorbidities in the FFS population. For example, there was a notable increase in percent frequency for "other psychiatric disorders (CC 60)" (27.8% to 33.3%), "cardio-respiratory failure or shock (CC 79)" (36.3% to 39.5%), "sleep apnea (ICD-9 codes 327.20, 327.21, 327.23, 327.27, 327.29, 780.51, 780.53, and 780.57)" (17.7% to 19.9%), and "other endocrine/metabolic/nutritional disorders (CC 24)" (80.9% to 83.1%). There was a notable decrease in percent frequency for "fibrosis of lung or other chronic lung disorders (CC 109)" (18.1% to 15.5%), "coronary atherosclerosis or angina (CC 83-84)" (54.6% to 52.8%), and "congestive heart failure (CC 80)" (45.7% to 44.6%). Examination of the odds ratios for

each risk variable in the model shows that, overall, the odds ratios for individual risk variables remained relatively constant across three years.

For the model variable frequencies and risk variable odds ratios, see the *2015 Measure Updates and Specifications Report* (Dorsey et al. 2015).

Measure Score Reliability Results (**Dataset 1**)
There were 925,315 admissions in the combined 3-year sample, with 461,505 in one sample and 463,810 in the other randomly selected sample. The agreement between the two RSRRs for each hospital was 0.48, which according to the conventional interpretation is "moderate" (Landis and Koch, 1977).

Note that this analysis was limited to hospitals with 12 or more cases in each split sample. The intra-class correlation coefficient is based on a split sample of three years of data, resulting in a volume of patients in each sample equivalent to only 1.5 years of data, whereas the measure is reported with the full three years of data.

Reference:
Dorsey K, Grady J, Desai N, et al. 2015 Condition-Specific Measures Updates and Specifications Report Hospital-Level 30-Day Risk-Standardized Readmission Measures Acute Myocardial Infarction – Version 8.0 Heart Failure – Version 8.0 Pneumonia – Version 8.0 Chronic Obstructive Pulmonary Disease – Version 8.0 Stroke – Version 8.0. 2014; https://www.qualitynet.org/dcs/BlobServer?blobkey=id&blobnocache=true&blobwhere=1228890435217&blobheader=multipart%2Foctet-stream&blobheadername1=Content-Disposition&blobheadervalue1=attachment%3Bfilename%3DRdmn_AMIHFPNCOPDSTK_Msr_UpdtRpt.pdf&blobcol=urldata&blobtable=MungoBlobs. Accessed October 30, 2015.

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 over time of the risk factor frequencies and odds ratios suggests that the underlying data elements are reliable. Additionally, the ICC score demonstrates moderate agreement of measure scores across samples using a conservative approach to assessment.

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 claims-based measures, through use of established measure development guidelines, and by systematic assessment of measure face validity by a Technical Expert Panel (TEP) of national experts and stakeholder organizations.

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 seven NQF-endorsed measures currently in public reporting (AMI, heart failure, and pneumonia mortality and readmission and coronary artery bypass graft surgery or CABG 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) (Krumholz et al., 2006 [3]; Keenan et al., 2008), AMI patients (Cooperative Cardiovascular Project data) (Krumholz et al., 2006 [2]), pneumonia patients (National Pneumonia Project dataset) (Bratzler et al., 2011), and CABG patients (Shahian et al., 2014; Suter et al., 2014). 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.

We have also completed two national, multi-site validation efforts for two procedure-based complications measures (for primary elective hip/knee arthroplasty and implantable cardioverter defibrillator [ICD]). Both projects demonstrated strong agreement between complications coded in claims and abstracted medical chart data.

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 (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" (Krumholz et al., 2006 [1]).

Validity as Assessed by External Groups:

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

The working group was assembled, and regular meetings were held throughout the development phase. The working group was tailored for development of this measure and consisted of three physicians who are board-certified in pulmonary and critical care medicine and a pharmacoepidemiologist with expertise in COPD. All members have expertise in quality measure development. The working group meetings addressed key issues related to measure development including weighing the pros and cons of and finalizing key decisions (e.g., defining the measure cohort and outcome) to ensure the measure is meaningful, useful, and well-designed. The working group provided a forum for focused expert review and discussion of technical issues during measure development prior to consideration by the broader TEP.

In addition to the working group, and in alignment with the CMS MMS, 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 physicians, 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.

Following completion of the preliminary model, we solicited public comment on the measure through the CMS site: https://www.cms.gov/MMS/17_CallforPublicComment.asp. The public comments were then posted publicly for 30 days. The resulting input was taken into consideration during the final stages of measure development and contributed to minor modifications to the measure.

Face Validity as Determined by TEP:
One means of confirming the validity of this measure was face validity assessed by our Technical Expert Panel (TEP), which included 11 members including individuals with diverse perspectives and backgrounds, including clinicians, consumers, hospitals, purchasers, and experts in quality improvement.

List of TEP Members
-Darlene Bainbridge, MS, NHA, CPHQ, CPHRM (President/CEO, Darlene D. Bainbridge & Associates, Inc.)
-Robert A. Balk, MD (Director of Pulmonary and Critical Care Medicine, Rush University Medical Center)
-Dale Bratzler, DO, MPH (President and CEO, Oklahoma Foundation for Medical Quality)
-Scott Cerreta, RRT (Director of Education, COPD Foundation)
-Gerard J. Criner, MD (Director of Temple Lung Center and Divisions of Pulmonary and Critical Care Medicine, Temple University)
-Guy D'Andrea, MBA (President, Discern Consulting)
-Jonathan Fine, MD (Director of Pulmonary Fellowship, Research and Medical Education, Norwalk Hospital)
-David Hopkins, MS, PhD (Senior Advisor, Pacific Business Group on Health)
-Fred Martin Jacobs, MD, JD, FACP, FCCP, FCLM (Executive Vice President and Director, Saint Barnabas Quality Institute)
-Natalie Napolitano, MPH, RRT□NPS (Respiratory Therapist, Inova Fairfax Hospital)
-Russell Robbins, MD, MBA (Principal and Senior Clinical Consultant, Mercer)

We systematically assessed the face validity of the measure score as an indicator of quality by soliciting the TEP members' agreement with the following statement: "The risk-standardized readmission rates obtained from the COPD readmission measure as specified will provide an accurate reflection of quality."

On a six-point scale (1=Strongly disagree, 2=Moderately disagree, 3=Somewhat disagree, 4=Somewhat agree, 5=Moderately agree, 6=Strongly agree), 10 of 11 TEP members responded to the survey question as follows: one TEP member Strongly Disagreed, two Somewhat Agreed, four Moderately Agreed, and three Strongly Agreed. Of the TEP members who responded, 90% agreed (70% moderately or strongly agreed) that the measure will provide an accurate reflection of quality. We therefore gave the measure a moderate rating for face validity. In summary, these results demonstrated TEP agreement with the overall face validity of the measure as specified.

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

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

Krumholz HM, Brindis RG, Brush JE, et al. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. Circulation. January 24, 2006; 113(3):456-462. [1]

Krumholz HM, Wang Y, Mattera JA, et al. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. Circulation 2006; 113(13):1683-92. [2]

Krumholz HM, Wang Y, Mattera JA, et al. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with heart failure. Circulation 2006; 113:1693-1701. [3]

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.

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

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

ICD-9 to ICD-10 Conversion

Statement of Intent

[X] Goal was to convert this measure to a new code set, fully consistent with the intent of the original measure.
[ ] Goal was to take advantage of the more specific code set to form a new version of the measure, but fully consistent with the original intent.
[ ] The intent of the measure has changed.

Process of Conversion

ICD-10 codes were identified using 2015 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. An ICD-9 to ICD-10 crosswalk is attached in field S.2b. (Data Dictionary or Code Table).

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

The performance of the first half of the split sample (development sample) and second half of the split sample (validation sample) from **Dataset 2** was similar. The areas under the receiver operating characteristic (ROC) curve for the two models are 0.627 and 0.629, respectively.

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?*)

The results between the first half of the split sample and second half of the split sample from **Dataset 2** proved to be similar for each of the model testing that was performed. The ROC results were nearly identical and in line with other readmission models.

Validity as Assessed by External Groups:

The face validity testing results demonstrated TEP agreement with overall face validity of the measure as specified.

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*)

All exclusions were determined by careful clinical review and have been made based on clinically relevant decisions to ensure accurate calculation of the measure. To ascertain impact of exclusions on the cohort, we examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion (**Dataset 1**). These exclusions are consistent with similar NQF-endorsed outcome measures. Rationales for the exclusions are detailed in data field S.10 (Denominator Exclusions).

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*)

In **Dataset 1**:

Exclusion	N	%	Distribution across hospitals (N=3,840): Min, 25th, 50th, 75th percentile, max
1. Without at least 30 days of post-discharge enrollment in FFS Medicare	5,173	0.51%	(0.00, 0.00, 0.00, 0.67, 50.00)
2. Discharged against medical advice (AMA)	5,966	0.59%	(0.00, 0.00, 0.00, 0.70, 25.00)
3. COPD admission within 30 days of a prior COPD index admission	75,166	7.43%	(0.00, 4.17, 6.49, 8.43, 100.00)

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*)

Exclusion 1 (patients without at least 30 days of post-discharge enrollment in FFS Medicare for index admissions) accounts for 0.51% of all index admissions excluded from the initial cohort. This exclusion is

needed since the 30-day readmission outcome cannot be assessed in this group since claims data are used to determine whether a patient was readmitted.

Exclusion 2 (patients who are discharged AMA) accounts for 0.59% of all index admissions excluded from the initial index cohort. This exclusion is needed for acceptability of the measure to hospitals, who do not have the opportunity to adequately deliver full care and prepare the patient for discharge. Given that a very small percent of patients are being excluded, it is unlikely that is exclusion affects the measure score.

Exclusion 3 (patients with admission within 30 days of a prior index admission) if a patient has an admission within 30 days of discharge from the index admission, that admission is not included in the cohort so that admission can be both an index admission and readmission. This exclusion accounts for 7.43% of all index admissions excluded from the initial index cohort.

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 41 risk factors**

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

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

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (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 discharge for age, 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 final set of risk-adjustment variables is:

• Age-65 (years, continuous) for patients aged 65 or over cohorts; or Age (years, continuous) for patients aged 18 and over cohorts
• History of mechanical ventilation
• Sleep apnea
• History of infection
• Metastatic cancer or acute leukemia
• Lung, upper digestive tract, and other severe cancers
• Lymphatic, head and neck, brain, and other major cancers; breast, colorectal and other cancers and tumors; other respiratory and heart neoplasms
• Other digestive and urinary neoplasms
• Diabetes mellitus (DM) or DM complications
• Protein-calorie malnutrition
• Disorders of fluid/electrolyte/acid-base
• Other endocrine/metabolic/nutritional disorders
• Pancreatic disease
• Peptic ulcer, hemorrhage, other specified gastrointestinal disorders
• Other gastrointestinal disorders
• Severe hematological disorders
• Iron deficiency or other unspecified anemias and blood disease
• Dementia or other specified brain disorders
• Drug/alcohol psychosis or dependence
• Major psychiatric disorders
• Depression
• Anxiety disorders
• Other psychiatric disorders
• Hemiplegia, paraplegia, paralysis, functional disability
• Polyneuropathy
• Respirator dependence/respiratory failure
• Cardio-respiratory failure and shock
• Congestive heart failure
• Acute coronary syndrome
• Coronary atherosclerosis or angina
• Hypertensive heart and renal disease or encephalopathy
• Specified arrhythmias and other heart rhythm disorders
• Other or unspecified heart disease
• Stroke

• Vascular or circulatory disease
• Fibrosis of lung or other chronic lung disorders
• Pneumonia
• Renal failure
• Decubitus ulcer or chronic skin ulcer
• Cellulitis, local skin infection
• Vertebral fractures

Socioeconomic Status (SES) Factors and Race

We selected variables representing 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 COPD Readmission

To examine the relationship between SES and race variables and hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following COPD hospitalization, 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 COPD readmission. Eleven studies were reviewed by title and abstract, and ten studies were excluded from full-text review. To this, we added an additional article recommended by an expert consultant. While limited data were identified meeting these criteria, the studies reviewed found that health disparities indicators were associated with increased risk of COPD readmission (Elixhauser et al., 2008; Sharma et al., 2010).

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 measure 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 1.8, the following SES and race variables were considered:

- African American race (as compared to all others)
- Dual eligible status
- AHRQ SES index score

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:

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 2014; 7:391-7.

Eapen ZJ, McCoy LA, Fonarow GC, et al. Utility of socioeconomic status in predicting 30-day outcomes after heart failure hospitalization. Circ Heart Fail. May 2015; 8(3):473-80.

Elixhauser A, Au DH, Podulka J. Readmissions for chronic obstructive pulmonary disease, 2008. Rockville, MD: Agency for Healthcare Research and Quality. HCUP Statistical Brief #121. Accessed: July 12 2013. Available at: http://www.hcup-us.ahrq.gov/reports/statbriefs/sb121.jsp.

Gilman M, Adams EK, Hockenberry JM, et al. 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.

Krumholz HM, Brindis RG, Brush JE, et al. Standards for statistical models used for public reporting of health outcomes. Circulation. 2006; 113: 456-462. Available at: http://circ.ahajournals.org/content/113/3/456.full.pdf+html . Accessed January 14, 2016.

Normand S-LT, Shahian DM. Statistical and Clinical Aspects of Hospital Outcomes Profiling. 2007/05 2007:206-226.

Reames BN, Birkmeyer NJ, Dimick JB, et al. Socioeconomic disparities in mortality after cancer surgery: failure to rescue. JAMA surgery 2014; 149:475-81.

Sharma G, Kuo YF, Freeman JL, et al. Outpatient follow-up visit and 30-day emergency department visit and readmission in patients hospitalized for chronic obstructive pulmonary disease. Archives of internal medicine. Oct 11 2010; 170(18):1664-1670.

Skinner J, Chandra A, Staiger D, et al. 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.

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

Below is a table showing the final variables in the model with associated odds ratios.

Final Model Variables (variables meeting criteria in field 2b4.3) (**Dataset 1**)

Variable	07/2011-06/2014 OR (95% CI)
Age minus 65 (years above 65, continuous)	1.00 (1.00 - 1.00)
History of mechanical ventilation (ICD-9 codes 93.90, 96.70, 96.71, 96.72)	1.16 (1.14 - 1.18)
Sleep apnea (ICD-9 codes 327.20, 327.21, 327.23, 327.27, 327.29, 780.51, 780.53, 780.57)	0.99 (0.97 - 1.00)
Respirator dependence/respiratory failure (CC 77-78)	1.05 (1.01 - 1.09)
Cardio-respiratory failure or shock (CC 79)	1.22 (1.20 - 1.23)
Congestive heart failure (CC 80)	1.21 (1.20 - 1.23)
Acute coronary syndrome (CC 81-82)	1.08 (1.06 - 1.10)
Coronary atherosclerosis or angina (CC 83-84)	1.09 (1.08 - 1.10)
Specified arrhythmias and other heart rhythm disorders (CC 92-93)	1.15 (1.14 - 1.17)
Other and unspecified heart disease (CC 94)	1.07 (1.05 - 1.08)
Vascular or circulatory disease (CC 104-106)	1.08 (1.06 - 1.09)
Fibrosis of lung and other chronic lung disorders (CC 109)	1.10 (1.09 - 1.12)
Pneumonia (CC 111-113)	1.10 (1.08 - 1.11)
History of infection (CC 1, 3-6)	1.07 (1.05 - 1.08)
Metastatic cancer or acute leukemia (CC 7)	1.24 (1.20 - 1.28)
Lung, upper digestive tract, and other severe cancers (CC 8)	1.22 (1.20 - 1.25)
Lymphatic, head and neck, brain, and other major cancers; breast, colorectal and other cancers and tumors; other respiratory and heart neoplasms (CC 9-11)	1.02 (1.00 - 1.03)
Other digestive and urinary neoplasms (CC 12)	0.96 (0.94 - 0.98)
Diabetes mellitus (DM) or DM complications (CC 15-20, 119-120)	1.05 (1.04 - 1.06)
Protein-calorie malnutrition (CC 21)	1.15 (1.13 - 1.17)
Disorders of fluid/electrolyte/acid-base (CC 22-23)	1.16 (1.15 - 1.18)

Variable	07/2011-06/2014 OR (95% CI)
Other endocrine/metabolic/nutritional disorders (CC 24)	0.93 (0.92 - 0.95)
Pancreatic disease (CC 32)	1.05 (1.01 - 1.08)
Peptic ulcer, hemorrhage, other specified gastrointestinal disorders (CC 34)	1.07 (1.06 - 1.09)
Other gastrointestinal disorders (CC 36)	1.07 (1.06 - 1.09)
Severe hematological disorders (CC 44)	1.19 (1.14 - 1.24)
Iron deficiency or other unspecified anemias and blood disease (CC 47)	1.17 (1.16 - 1.18)
Dementia or other specified brain disorders (CC 49-50)	1.00 (0.98 - 1.01)
Drug/alcohol psychosis or dependence (CC 51-52)	1.16 (1.13 - 1.18)
Major psychiatric disorders (CC 54-56)	1.04 (1.02 - 1.06)
Depression (CC 58)	1.02 (1.01 - 1.04)
Anxiety disorders (CC 59)	1.08 (1.06 - 1.10)
Other psychiatric disorders (CC 60)	1.11 (1.09 - 1.12)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)	1.06 (1.04 - 1.08)
Polyneuropathy (CC 71)	1.07 (1.06 - 1.09)
Hypertensive heart and renal disease or encephalopathy (CC 89)	1.12 (1.10 - 1.14)
Stroke (CC 95-96)	1.01 (0.99 - 1.03)
Renal failure (CC 131)	1.06 (1.05 - 1.08)
Decubitus ulcer or chronic skin ulcer (CC 148-149)	1.08 (1.06 - 1.10)
Cellulitis, local skin infection (CC 152)	1.05 (1.04 - 1.07)
Vertebral fractures (CC 157)	1.15 (1.13 - 1.18)

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 COPD cohort varies across measured entities. The median percentage of dual eligible patients is 20.0% (interquartile range [IQR] 13.5%-28.2%). The median percentage of African-American patients is 2.4% (IQR 0.0%-9.2%). The median percentage of patients with an ARHQ SES index score equal to or below 45.0 is 14.6% (IQR 2.8%-43.0%).

Empirical association with the outcome (univariate)

The patient-level observed COPD readmission rate is higher for dual eligible patients, 22.8%, compared with 19.6% for all other patients. The readmission rate for African-American patients was also higher at 22.1% compared with 20.1% for patients of all other races. Similarly the readmission rate for patients with an AHRQ SES index score equal to or below 45.0 was 20.9% compared with 20.0% for patients with an AHRQ SES index score above 45.0.

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. 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 small. We also find that the c-statistic is unchanged with the addition of any of these variables into the model. Furthermore we find that 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 -0.005% (interquartile range [IQR] -0.027% – 0.032%, minimum -0.348% – maximum 0.213%) with a correlation coefficient between RSRRs for each hospital with and without dual eligibility added of 0.99888. The median absolute change in hospitals' RSRRs when adding a race indicator is 0.007% (IQR -0.005% – 0.016%, minimum -0.305% – maximum 0.044%) with a correlation coefficient between RSRRs for each hospital with and without race added of 0.99973. The median absolute change in

hospitals' RSRRs when adding an indicator for a low AHRQ SES index score is 0.017% (IQR -0.054% – 0.068%, minimum -1.209% – maximum 0.941%) with a correlation coefficient between RSRRs for each hospital with and without an indicator for a low AHRQ SES index score added of 0.99295.

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

Both the patient-level and hospital-level dual eligible and low AHRQ SES Index effects were significantly associated with COPD readmission in the decomposition analysis. The patient-level race effect was not appreciably different from zero, though the hospital-level race effect was significant. If the dual eligible or low AHRQ SES Index variables are 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. If race is used as a risk adjustment variable, it will primarily capture an effect of the hospital on the outcome, not the effect of an intrinsic characteristic 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.

COPD Readmission Decomposition Analysis

Parameter	Estimate (Standard Error)	P-value
Dual Eligible – Patient-Level	0.0843 (0.00672)	<.0001
Dual Eligible – Hospital-Level	0.2077 (0.0256)	<.0001
African American – Patient-Level	0.00575 (0.0105)	0.5824
African American – Hospital-Level	0.1961 (0.0195)	<.0001
AHRQ SES Index – Patient-Level	0.0226 (0.00769)	0.0033
AHRQ SES Index – Hospital-Level	0.0766 (0.0102)	<.0001

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 (**Dataset 1 & Dataset 2**)
We computed three summary statistics for assessing model performance (Harrell and Shih, 2001) for the development and validation cohort:

Discrimination Statistics
(1) Area under the receiver operating characteristic (ROC) curve (the c-statistic) is the probability that predicting the outcome is better than chance, which is a measure of how accurately a statistical model is able to distinguish between a patient with and without an outcome.
(2) Predictive ability (discrimination in predictive ability measures the ability to distinguish high-risk subjects from low-risk subjects; therefore, we would hope to see a wide range between the lowest decile and highest decile.)

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

We tested the performance of the model for **Dataset 1** and **Dataset 2** described in section 1.7.

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

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 original measure development cohort (**Dataset 2**), the results are summarized below:

- 1st half of randomly split sample (development sample): C-statistic = 0.627; Dataset Predictive ability (lowest decile %, highest decile %) = (11.57, 38.08)
- 2nd half of randomly split sample (validation sample): C-statistic = 0.629; Dataset Predictive ability (lowest decile %, highest decile %) = (11.73, 39.19)

For the current measure cohort (version 4.0) (**Dataset 1**) the results are summarized below:

- C-statistic = 0.64
- Predictive ability (lowest decile %, highest decile %) = (10.1, 36.5)

For comparison of model with and without inclusion of SES factors, see above section.

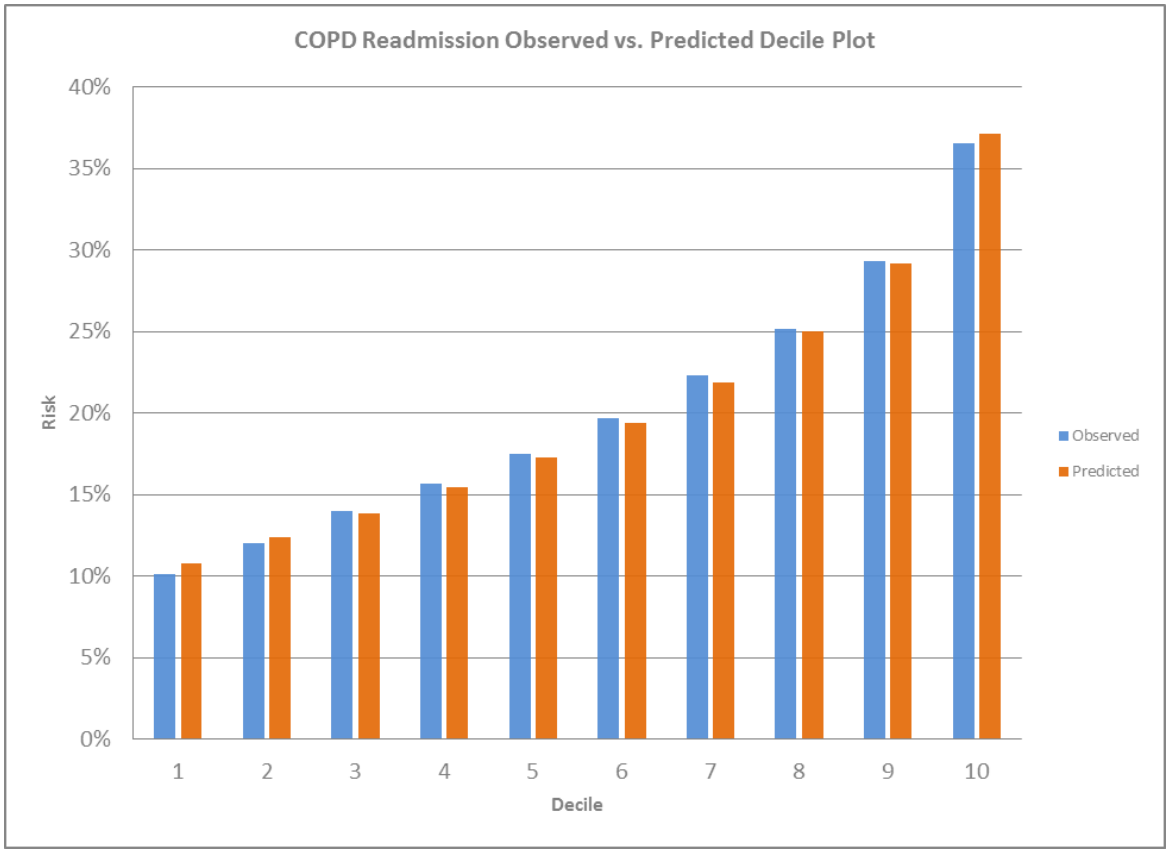
2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

For the original measure development cohort (**Dataset 2**) the results are summarized below:

- 1st half of split sample (development sample): Calibration: (-0.034, 0.970)
- 2nd half of split sample (validation sample): Calibration: (0.004, 0.994)

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. Below, we present the risk decile plot showing the distributions for Medicare FFS data from July 2011 to June 2014 (**Dataset 1**).



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-statistics of 0.64 indicate fair model discrimination (**Dataset 1**). 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 one, there is potential evidence of over-fitting. The calibration value of close to 0 at one end and close to 1 to 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 good 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*)

This measure is fully risk-adjusted using a hierarchical logistic regression model to calculate hospital RSRRs accounting for differences in hospital case-mix.

Application to Patients Aged 18 and Older (**Dataset 3**)
We applied the model to all-payer data from California. The analytic sample included 45,480 cases aged 18 and older 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.

To help determine whether the measure could be applied to a population of patients aged 18+, we examined the interaction terms between age (18-64 vs. 65+) and each of the other risk factors. Specifically, we fit the model in all patients 18+ with and without interaction terms and (a) conducted a reclassification analysis to compare risk prediction at the patient level; (b) compared the c-statistic; and (c) compared hospital-level risk-standardized rates (scatterplot, correlation coefficient, and R2) to assess whether the model with interactions is different from the current model in profiling hospital rates.

When the model was applied to all patients 18 and over (18+), overall discrimination was good (c-statistic=0.669). 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 (a) the reclassification analysis demonstrated that nearly all patients were found to be in a similar risk category; (b) the c-statistic was nearly identical (0.673 vs. 0.669); and (c) hospital-level risk-standardized rates were highly correlated (r=0.991). Thus, the inclusion of the interactions did not substantively affect either patient-level model performance or hospital-level results.

We conducted this testing prior to specifying the measure for patients age 40 and over. Restricting the patient cohort to age 40 and over, however, is not likely to affect the results given that only 1.5% of patients were between the ages of 18 and 39. Therefore, the measure can be applied to all-payer data for patients 40 and older.

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 (because it 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*)

Analyses of Medicare FFS data show substantial variation in RSRRs among hospitals. Using data from July 2011-June 2014 (**Dataset 1**), the median hospital RSRR was 20.2%, with a range of 15.5% to 26.6%. The interquartile range was 19.6%-20.8%.

Out of 4,663 hospitals in the U.S., 27 performed "better than the U.S. national rate," 3,730 performed "no different from the U.S. national rate," and 83 performed "worse than the U.S. national rate." 823 were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing.

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 the quality of care received across hospitals for COPD that support measurement to reduce the variation.

Note: Over the three years of the measure reporting period, the COPD readmission rate has decreased from 21.0% (July 2011 to June 2012) to 19.5% (July 2013 to June 2014). Despite recent decreases in readmission rates nationally, the readmission rate for the 2015 public reporting period (July 2011 to June 2014) for COPD Medicare FFS patients is at 20.3%.

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 criterion is directed 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).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of 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 demonstrating comparability of 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.**

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 (Hospital IQR) Program, Centers for Medicare and Medicaid Services (CMS)

Purpose: The 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 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 reporting 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 Hospital 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 included 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 Inpatient Prospective Payment System (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 COPD. The median 30-day RSRR decreased by 1.4 absolute percentage points from July 2011-June 2012 (median RSRR: 20.9%) to July 2013-June 2014 (median RSRR: 19.5%). The median hospital RSRR from July 2011-June 2014 was 20.2% (IQR 19.6% - 20.8%).

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 ongoing monitoring of 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)**
0070 : Coronary Artery Disease (CAD): Beta-Blocker Therapy-Prior Myocardial Infarction (MI) or Left Ventricular Systolic Dysfunction (LVEF <40%)

0275 : Chronic Obstructive Pulmonary Disease (COPD) or Asthma in Older Adults Admission Rate (PQI 05)
0701 : Functional Capacity in COPD patients before and after Pulmonary Rehabilitation
0709 : Proportion of patients with a chronic condition that have a potentially avoidable complication during a calendar year.
1561 : Relative Resource Use for People with COPD
1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)
1893 : Hospital 30-Day, all-cause, risk-standardized mortality rate (RSMR) following chronic obstructive pulmonary disease (COPD) hospitalization

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?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
We did not include in our list of related measures any non-outcome (e.g., process) measures with the same target population as our measure. Because this is an outcome measure, clinical coherence of the cohort takes precedence over alignment with related non-outcome measures. Furthermore, non-outcome measures are limited due to broader patient exclusions. This is because they typically only include a specific subset of patients who are eligible for that measure (for example, patients who receive a specific medication or undergo a specific procedure).

5b. Competing Measures

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

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
N/A

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** 2015_Measures_Reevaluation_Condition-Specific_Readmission_AUS_Report_FINAL_508_Compliant.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services (CMS)
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.

The working group involved in the initial measure development is detailed in the original technical report available at www.qualitynet.org.

Our measure development team consisted of the following members:
Laura M. Grosso, PhD, MPH
Peter Lindenauer, MD, MSc
Changqin Wang, MD, MS
Shantal Savage, BA
Jaymie Potteiger, MPH
Zameer Abedin, BA
Lori L. Geary, MPH
Yun Wang, PhD
Elizabeth E. Drye, MD, SM
Harlan M. Krumholz, MD, SM

Technical Expert Panel Members:
Darlene Bainbridge, MS, NHA, CPHQ, CPHRM, President/CEO, Darlene D. Bainbridge & Associates, Inc.

Robert A. Balk, MD, Director of Pulmonary and Critical Care Medicine, Rush University Medical Center

Dale Bratzler, DO, MPH, President and CEO, Oklahoma Foundation for Medical Quality

Scott Cerreta, RRT, Director of Education, COPD Foundation

Gerard J. Criner, MD, Director of Temple Lung Center and Divisions of Pulmonary and Critical Care Medicine, Temple University

Guy D' Andrea, MBA, Discern Consulting

Jonathan Fine, MD, Director of Pulmonary Fellowship, Research and Medical Education, Norwalk Hospital

David Hopkins, MS, PHD, Senior Advisor, Pacific Business Group on Health

Fred Martin Jacobs, MD, JD, FACP, FCCP, FCLM, Executive Vice President and Director, Saint Barnabas Quality Institute

Natalie Napolitano, MPH, RRT-NPS, Respiratory Therapist, Inova Fairfax Hospital

Russell Robbins, MD, MBA, Principal and Senior Clinical Consultant, Mercer

Working Group Panel Members:
David Au, MD, MS, Investigator, VA Puget Sound Healthcare System, Northwest HSR&D Center of Excellence;
Associate Professor of Medicine, Department of Medicine, University of Washington

Jerry Krishnan, MD, PhD, Associate Professor, Departments of Medicine and Health Studies, University of Chicago;
Director, Asthma Center and Refractory Obstructive Lung Disorders Clinic, University of Chicago

Todd Lee, PharmD, PhD, Associate Professor, Departments of Pharmacy Practice and Pharmacy Administration, University of Illinois at Chicago; Senior Investigator, Center for Management of Complex Chronic Care (CMC3), Hines VA Hospital

Richard Mularski, MD, MCR, MSHS, Clinical Investigator, Center for Health Research, Kaiser Permanente; Clinical Assistant Professor of Medicine, Oregon Health & Science University

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2014 Ad.3 **Month and Year of most recent revision:** 07, 2015 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?** 01, 2017
Ad.6 **Copyright statement:** N/A Ad.7 **Disclaimers:** N/A
Ad.8 **Additional Information/Comments:** N/A

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 #: 0330
Measure Title: Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization
Measure Steward: Centers for Medicare & Medicaid Services (CMS)
Brief Description of Measure: The measure estimates a hospital-level risk-standardized readmission rate (RSRR) for patients discharged from the hospital with a principal diagnosis of heart failure (HF). The outcome (readmission) is defined as unplanned readmission for any cause within 30 days of the discharge date for the index admission (the admission included in the measure cohort). A specified set of planned readmissions do not count in the readmission outcome. The target population is patients 18 and over. CMS annually reports the measure for patients who are 65 years or older, are enrolled in fee-for-service (FFS) Medicare, and hospitalized in non-federal hospitals or Veterans Health Administration (VA) hospitals.
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 HF. 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.

HF 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.

Numerator Statement: The outcome for this measure is 30-day readmission. We define readmission as an inpatient admission for any cause, with the exception of certain planned readmissions, within 30 days from the date of discharge from the index HF admission. If a patient has more than one unplanned admissions (for any reason) within 30 days after discharge from the index admission, only one is counted as a readmission. The measure looks for a dichotomous yes or no outcome of whether each admitted patient has an unplanned readmission within 30 days. However, if the first readmission after discharge is considered planned, any subsequent unplanned readmission is not counted as an outcome for that index admission, because the unplanned readmission could be related to care provided during the intervening planned readmission rather than during the index admission.
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 explicitly tested the measure in both age groups.

The cohort includes admissions for patients aged 18 years and older discharged from the hospital with either a principal discharge diagnosis of HF (see codes below) and with a complete claims history for the 12 months prior to admission. The measure is currently publicly reported by CMS for those patients 65 years and older who are Medicare FFS beneficiaries admitted to non-federal hospitals or Veterans Health Administration (VA) hospitals.

Additional details are provided in S.9 Denominator Details.
Denominator Exclusions: The readmission measures excludes admissions:

1. Ending in discharges against medical advice
Rationale: Providers did not have the opportunity to deliver full care and prepare the patient for discharge.
2. Without at least 30 days of 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.
3. Occurring within 30 days of discharge from an index admission
Rationale: This exclusion ensures that no hospitalization will be considered as both a readmission and an index admission within the same measure.
4. With a procedure code for LVAD implantation or heart transplantation either during the index admission or in the 12 months prior to the index admission
Rationale: Patients with these procedures are a highly-selected group of patients with a different risk of the readmission outcome.

Measure Type: Outcome
Data Source: Administrative claims
Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: May 15, 2008 **Most Recent Endorsement Date:** Jan 18, 2012

Maintenance of Endorsement -- 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. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

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.

Summary of evidence:

- As a rationale for measuring this health outcome, the developer suggests that hospitals are able to influence readmission rates through a broad range of clinical activities including communication between providers, prevention of, and response to, complications, patient safety and coordinated transitions to the outpatient environment.

- The developer states that there are no updates to the evidence since the last submission.

Question for the Committee:

- *Since there are no updates to the evidence, does the Committee agree that there is no need for repeat discussion and vote on Evidence?*
- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

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.

- The developer provides performance data from four measurement periods, covering a total of 1,210,454 admissions.
- The data show that during the measurement period of 07/2011-06/2014, heart failure readmission rates ranged from a minimum of 16% to a maximum of 32.1%, with the 10th percentile at 20.7%, the 50th percentile at 22.3%, and the 90th percentile at 24.3%.

Disparities

- To help in assessment of potential disparities, the developers also provide performance scores (using July 2013-June 2014 data) for hospitals serving a low proportion of dual eligible patients vs. those serving a high proportion of dual eligible patients, performance scores for hospitals serving a low proportion of African-American patients vs. those serving a high proportion of African-American patients, and performance scores for hospitals serving a low proportion of patients with AHRQ SES Index Score index score equal to or below 42.7 vs. those serving a high proportion of patients with an AHRQ SES index score equal to or below 42.7.

- Hospitals serving a low proportion (=8.7%) Dual Eligible patients had a slightly lower median readmission rates (-0.7%) compared to hospitals serving a high proportion (=22.0%) Dual Eligible patients. Hospitals serving a low proportion (=0.0%) African-American patients had a slightly lower median readmissions rates (-0.9%) compared to hospitals serving a high proportion (=13.2%) African-American patients. Finally, hospitals serving a low proportion of patients below AHRQ SES index score of 42.7had slightly lower median readmissions rates (-0.9%) compared to hospitals serving a high proportion of patients below AHRQ SRS index score of 42.7.

- By proportion of **Dual Eligible Patients**:

  Characteristic//Hospitals with a low proportion (≤8.7%) Dual Eligible patients//Hospitals with a high proportion (≥22.0%) Dual Eligible patients
  Number of Measured Hospitals// 997 // 1,003
  Number of Patients// 333,931 patients in low-proportion hospitals/ 195,234 in high-proportion hospitals
  Maximum// 28.1// 32.1
  90th percentile// 24.1// 24.9
  75th percentile// 23.1 // 23.9
  Median (50th percentile)// 22.1// 22.8
  25th percentile// 21.2// 22.0

10th percentile// 20.3// 21.2
Minimum // 16.0 // 18.0

- By proportion of **African-American Patients**:

  Characteristic// Hospitals with a low proportion (≤0.0%) African-American patients//Hospitals with a high proportion (≥13.2%) African-American patients
  Number of Measured Hospitals// 1,067 // 999
  Number of Patients// 99,898 patients in low-proportion hospitals/ 373,385 in high-proportion hospitals
  Maximum// 26.6// 28.6
  90th percentile// 23.7// 25.1
  75th percentile// 22.8// 23.9
  Median (50th percentile)// 22.0// 22.9
  25th percentile// 21.3// 21.9
  10th percentile// 20.6// 20.9
  Minimum // 18.0// 18.4

- By Proportion of Patients with AHRQ SES Index Scores Equal or Below 42.7:

  Characteristic//Hospitals with a low proportion of patients below AHRQ SES index score of 42.7 (≤9.2%)// Hospitals with a high proportion of patients below AHRQ SES index score of 42.7 (≥38.3%)
  Number of Measures Hospitals// 999 // 999
  Number of Patients// 257,667 patients in hospitals with low proportion of patients below AHRQ SES index score of 42.7 //218,581 patients in hospitals with high proportion of patients below AHRQ SES index score of 42.7
  Maximum// 27.7// 32.1
  90th percentile// 23.7// 25.1
  75th percentile// 22.8// 24.0
  Median (50th percentile)// 21.9// 22.8
  25th percentile// 21.2// 22.0
  10th percentile// 20.3// 21.2
  Minimum // 16.0// 18.4

Questions for the Committee:

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

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

Committee pre-evaluation comments

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

1a. Evidence to Support Measure Focus

Comments: **Agree with developer statement that "hospitals are able to influence readmission rates through a broad range of clinical activities including communication between providers, prevention of, and response to, complications, patient safety and coordinated transitions to the outpatient environment."

PASS

**The evidence that 30-day all-cause readmission (as opposed to later readmission) can be reduced by specific interventions is very weak. While there is a theoretical rationale/ conceptual model for reducing readmissions the data from actual trials is disappointing.

1b. Performance Gap

Comments: **There is a significant gap in performance across high and low-performance hospitals:
"heart failure readmission rates ranged from a minimum of 16% to a maximum of 32.1%, with the 10th percentile at 20.7%, the 50th percentile at 22.3%, and the 90th percentile at 24.3%."
**It is difficult to know what the goal performance is. There is likely a readmission rate that is too low and will be associated with adverse outcomes. However, if one just examines the variation across hospitals it suggests there remains a performance gap.

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

Comments: **NA
**Not applicable

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.

- This measure calculates 30-day readmissions for patients hospitalized with heart failure (HF).
- The measure produces a risk-standardized readmission rate (RSRR), which is calculated as the ratio of the number of "predicted" to the number of "expected" readmission at a given hospital, multiplied by the national observed readmission rate.
- The denominator includes patients aged 18 years and older discharged from the hospital with a principal discharge diagnosis of HF and with a complete claims history for the 12 months prior to admission. The measure can also be calculated for patients aged 65 and older only.
- The numerator includes patients who were readmitted for any cause, with the exception of certain planned readmissions, within 30 days from the date of discharge from the index HF admission.
- The denominator population is defined using ICD-9 and ICD-10 codes; a list of applicable codes is included in the submission.
- The data sources for this measure may include Medicare Part A and B claims, Veterans Health Administration claims, the Medicare Enrollment Database (EDB), and all-payer data sources such as the California Patient Discharge Database.
- The measure's time window can be specified from one to three years.
- The measure is risk-adjusted using a statistical risk model (see details below).

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.

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

- Datasets used for testing included Medicare Parts A and B claims, Veterans' Health Administration claims, as well as the Medicare Enrollment Database (EDB). Additionally, census data were used to assess socio-demographic factors.
- **Data element reliability**:
  - With regard to data element reliability, the developer notes that the measure has been developed to avoid the use of claims data elements that are thought to be coded inconsistently across hospitals or providers, instead using fields that are consequential for payment and which are audited by CMS.
  - In addition, the developer compared frequencies and odds ratios of variables from their risk model across three years of data in order to assess the consistency of those variables over time.
- **Performance score reliability:**
  - The developer defines performance score reliability as the degree to which repeated measurements of the same entity agree with each other.
  - In line with this thinking, the developer's approach to assessing score-level reliability was to consider the extent to which assessments of a hospital using different but randomly-selected subsets of patients produce similar measures of hospital performance. The developers refer to this as a "test-retest" approach; it may also be called a "split-half" method. This is generally considered an approrpate method of testing reliability.

Results of reliability testing

- **Data element reliability**:
  - Summarizing the results of this analysis, the developer notes that the frequency of some model variables increased between 2011 and 2014, which may reflect an increased or decreased rate of specific comorbidities in the FFS population.
  - The developer states that examination of the odds ratios for each risk variable in the model shows that, overall, the odds ratios for individual risk variables remained relatively constant across the three years.
  - The developer interprets the stability of the risk factor odds ratios over time as suggesting that the underlying data elements are reliable.
- **Performance score reliability:**
  - A total of 1,210,454 admissions over a 3-year period were examined, with 604,022 in one sample and 606,432 in the other randomly-selected sample. Two risk-standardized readmission rates (RSRR) were calculated for each hospital: one from each of the two separate samples.
    - The agreement between the two RSRRs for each hospital (as measured by an intra-class correlation coefficient (ICC)) was 0.58; the developer states that according to the conventional interpretation, this is considered a "moderate" level of agreement.
    - The developer notes that this analysis was limited to hospitals with 12 or more cases in each split sample, and that splitting the total population into two samples resulted in a sample equivalent of only 1.5 years of data, whereas the measure is reported with the full three years of data. [Note: It is unclear whether the measure itself is limited to hospitals with 12 or more cases; if it is not, then testing was not conducted with the measure as specified.]

Guidance from the Reliability Algorithm

- Question 1. Submitted specifications are precise, unambiguous, and complete. Measure can be consistently implemented.

- Question 2. Empirical reliability testing was conducted using statistical tests with the measure as specified.
- Question 3. Empirical validity testing of patient-level data was conducted.
- Question 4. Reliability testing was conducted with computed performance measure scores for each measured entity.
- Question 5. Random split-half correlation was used to assess the proportion of variability due to real differences among the measured entities.
- Question 6. The ICC was 0.58 which is considered a moderate level of agreement.

Questions for the Committee:

- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*
- *Does the measure testing match the measure specifications?*

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.

- This measure estimates 30-day all-cause readmission rates for patients discharged from an acute care hospital with a diagnosis of heart failure using a risk-standardized readmission rate (RSRR), which is calculated as the ratio of the number of "predicted" to the number of "expected" readmission at a given hospital, multiplied by the national observed readmission rate
- As a rationale for measuring this health outcome, the developers suggest that hospitals are able to influence readmission rates through a broad range of clinical activities, including prevention of complications, improving communication among providers involved at care transition, discharge planning, management of care transitions, patient education, and encouraging strategies that promote disease management.
- During the previous ad-hoc review of the updated planned readmissions algorithm used in the measure, the Expert Panel requested that CMS issue an advisory to hospitals and the public explaining that the new rates while lower are not a result of improvement in care, but rather an artifact due to the change in methodology. Overall, they agreed that the indicated changes sufficed in supporting the exclusion of planned readmissions and said exclusions adequately improved the validity of this measure.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**
Specification not completely consistent with evidence

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 developer demonstrated measure validity through medical record validation.
    - The HF readmission administrative model (original model specification prior to completion of the planned readmission algorithm) was validated against a medical record model with the same cohort of patients for which hospital-level HF readmission medical record data are available.
    - A measure cohort was developed with medical record data using the inclusion/exclusion criteria and risk-adjustment strategy.
    - A sample of 64,329 patients was matched for comparison.
- The developer assessed the areas under the receiver operating characteristic (ROC) curve for the two models, the predictive ability comparing readmission rates in the lowest predicted decile and the highest predicted decile.
- The RSRRs were estimated using the corresponding hierarchical logistic regression administrative and medical record models for the linked patient sample and the linear relationship between the two sets of estimates was examined using regression techniques and weighting by the total number of cases in each hospital.
- The developer notes the performance of the administrative and medical records models were similar and areas under the receiver operating characteristic (ROC) curve were 0.61 and 0.58, respectively, for the two models.
- The developer also found the models to be similar in terms of predictive ability.
- The correlation coefficient of the standardized rates from the administrative and medical record models was 0.97. The developer notes that this shows the resulting measure from the administrative claims model is as good as that from the medical record model.

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**

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:

- Patients in the following categories are excluded from the measure:
    - Discharged against medical advice (AMA);
    - Without at least 30 days post-discharge enrollment in FFS Medicare;
    - Heart failure admission within 30 days of a prior heart failure index admission;
    - Left ventricular assist device (LVAD) or transplant in index admission or prior year
- To determine the impact of exclusions, the developer examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion.
- The number and percentage of patients excluded for each criterion are as follows:
    1. Discharged against medical advice (AMA): **5,868 (.44%)**
    2. Without at least 30 days post-discharge enrollment in FFS Medicare for index admissions: **6,681 (.5%)**

3. Heart failure admission within 30 days of a prior heart failure index admission: **2,371 (.18%)**

- The developer also provides the distribution across hospitals for each exclusion criterion.

Questions for the Committee:

- *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: **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 measure employs a hierarchical logistic regression model (a form of hierarchical generalized linear model [HGLM]) to create a hospital-level 30-day risk-standardized readmission rate (RSRR).
- The developer suggests that 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.
- The developer notes that this approach simultaneously models data at the patient and hospital levels to account for the variance in patient outcomes both within and between hospitals.
- Variables considered for inclusion in the model 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 covariates are defined using condition categories (CCs), which are clinically-meaningful groupings of more than 15,000 ICD-9-CM diagnosis codes.
- The measure does not adjust for CCs that were possible adverse events of care and that were only recorded in the index admission.
- The final set of 37 risk-adjustment variables is included in the testing attachment; the odds ratio associated with each variable is also provided.
- The developers also considered a number of variables related to sociodemographic status (SDS) for potential inclusion in the risk-adjustment model. Candidate SDS variables were selected for examination based on a review of literature and national data sources.
- **Conceptual analysis of the need for SDS adjustment:**
    - The developers note there is a large body of literature linking various SES factors and African-American race to worse health status and higher readmission risk with income, education, and occupational level being the most commonly examined variables. The developers state that the literature directly examining how SES factors or race might influence the likelihood of older, insured, Medicare patient of being readmitted within 30 days of an admission for heart failure is more limited.
    - The developers state that few studies directly address causal pathways for SDS factors to affect 30-day

readmission rates or examine the role of the hospital in these pathways.

- There are at least four potential pathways for SDS factors to affect 30-day readmission rates:
  - One potential pathway is the relationship to health status at the time of admission. SDS factors 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. The developers note that this pathway should be largely accounted for by their clinical risk-adjustment model.
  - The next potential path way is that patients with low income and African-American patient are more likely to be seen in lower quality hospitals, which can contribute to increased risk of readmission.
  - The third major pathway is that a patient's race or SDS status cause them to experience differential, lower quality care or may not receive the differentiated care they require.
  - Finally, some SES risk factors may affect the likelihood of readmission without directly affecting health status at admission or the quality of care received during the hospitalization. Patients may have worse outcomes due to competing economic priorities or a lack of access to care outside the hospital.

- **Empirical analysis of SDS factors:**
  - The developers considered African-American race, dual-eligible status-i.e. enrolled in both Medicare and Medicaid, and AHRQ-validated SES index score (summarizing the information from the following variables: percentage of people in the labor force who are unemployed, percentage of people living below poverty level, median household income, median value of owner-occupied dwellings, percentage of people ≥25 years of age with less than a 12th-grade education, percentage of people ≥25 years of age completing ≥4 years of college, and percentage of households that average ≥1 people per room). The developers assessed the relationship between the SES variables and race with the outcome and examined the incremental effect in a multivariable mode.
  - The developer stated that they examined all patient-level indicators of both SES and race/ethnicity that are reliably available for all Medicare beneficiaries and linkable to claims data and selected those that are most valid.
  - The developer assessed the relationship between the SDS variables and the 30-day heart failure readmission rate and examined the incremental effect of SDS in a multivariable model, evaluating the extent to which the addition of any one of these variables improved model performance or changed hospital results.
  - The developer notes that 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. Therefore, the developers performed a decomposition analysis to assess the independent effects of the SES and race variables at the patient level and the hospital level.
  - The developers' analysis found that the prevalence of SDS factors in the heart failure cohort does vary across measured entities.
  - With regard to the empirical association of each SDS variable with the outcome (univariate), the analysis found that patient-level observed heart failure readmission rate for dual-eligible patients was higher, at 25.5% compared with 21.9% for all other patients. The readmission rate for African-American patients was also higher at 24.8% compared with 22.1% for patients of all other races. Similarly the readmission

rate for patients with an AHRQ SES index score equal to or below 42.7 was 24.3% compared with 21.8% for patients with an AHRQ SES index score above 42.7.

- With regard to the strength and significance of the SDS variables in the context of a multivariable model, the developers' analysis found that the effect size of each of these variables is small, the c-statistic (i.e., predictive value) is unchanged with the addition of any of these variables into the model, and the addition of any of these variables into the model has little to no effect on hospital performance.
  - The median absolute change in hospitals' RSRRs when adding a dual eligibility indicator is 0.0094% (interquartile range [IQR] -0.029% – 0.0386%, minimum -0.4499% – maximum 0.1559%) with a correlation coefficient between RSRRs for each hospital with and without dual eligibility added of 0.9993.
  - The median absolute change in hospitals' RSRRs when adding a race indicator is 0.0197% (IQR -0.0284% – 0.0538%, minimum -0.7499% – maximum 0.1576%) with a correlation coefficient between RSRRs for each hospital with and without race added of 0.9987.
  - The median absolute change in hospitals' RSRRs when adding an indicator for a low AHRQ SES index score adjusted for cost of living at the census block group level is 0.0377% (IQR -0.0502% – 0.1096%, minimum -0.9712% – maximum 0.2990%) with a correlation coefficient between RSRRs for each hospital with and without an indicator for a low AHRQ SES Index score adjusted for cost of living at the census block group level added of 0.9974.
- The developers state that the patient-level and hospital-level dual eligible, race, and low AHRQ SES Index effects were significantly associated with heart failure readmission in the decomposition analysis. The developers note that if the dual eligible, race, or low AHRQ SES Index variables are 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.
- To assess the relative contributions of the patient- and hospital-level effects, the developers calculated a range of predicted probabilities of readmission for the SES or race variables and clinical covariates (comorbidities).
- For SES and race variables, the hospital-level effect is greater than the patient-level effect (delta). For clinical variables, the patient-level effect (delta) is greater than the hospital-level effect for renal failure and metastatic cancer and equal to the hospital-level effect for COPD. The developers state that this consistent pattern demonstrates that SES and race variables have a much greater hospital-level effect than patient-level effect. The clinical variables consistently had the opposite pattern, with a greater effect at the patient level than at the hospital level. The developers concluded that including SES and race variables into the model would predominantly adjust for a hospital-level effect, which is an important signal of hospital quality.
- The developers state that given these findings and complex pathways that could explain any relationship between SDS and readmission, they did not incorporate SDS variables into the measure.

- **Risk Model Diagnostics:**
  - To assess the overall performance of their risk-adjustment model, the developers computed three summary statistics, including:
    - *Area under the receiver operating characteristic (ROC) curve* (also known as a c-statistic, which measures the probability that the model's prediction of the outcome is better than chance)

        - *Predictive ability* (the model's ability to distinguish high-risk subjects from low-risk subjects)
        - *Over-fitting indices (model calibration)* (to ensure that the model is not only describing the relationship between predictive variables and outcome in the development dataset but also providing valid predictions in new patients)
    - For the current measure cohort, the findings from this analysis are as follows:
        - C-statistic: **0.63**
            - A c-statistic of 0.608 means that for 60.8% of all possible pairs of patients—one who was readmitted and one who was not—the model correctly assigned a higher probability to those who were readmitted. Generally, a c-statistic of at least 0.70 is considered acceptable.
            - The developers interpret this as 'fair' model discrimination.
        - Predictive ability (lowest decile %, highest decile %): (**12.89%, 35.418%**)
            - The developers state that this indicates a wide range between the lowest decile and highest decile, indicating the ability to distinguish high-risk subjects from low-risk subjects.
        - Overfitting indices (model calibration) [presented as (γ0, γ1)]:
            - The developer states that if the γ0 in the validation samples are substantially far from zero and the γ1 is substantially far from one, there is potential evidence of over-fitting. The calibration value of close to 0 at one end and close to 1 to the other end indicates good calibration of the model.
                - 1st half of split sample: Calibration: (**0,1**)
                - 2nd half of split sample: Calibration: **(-0.02, 1.01)**
    - The developer's overall interpretation of the results of their analysis is that the findings demonstrate the risk-adjustment model adequately controls for differences in patient characteristics (case mix).
    - The developer also conducted additional analyses to determine whether the measure could be applied to a population of patients aged 18+ using all-payer data.
    - The developers report that their results indicate their model had good discrimination and predictive ability in this group.

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)*:

- For public reporting of this measure, CMS characterizes the uncertainty associated with the RSRR by estimating the 95% interval estimate.
- 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."
- The developer reports that for the performance period of July 2011-June 2014, the mean hospital RSRR was 22.34%,

with a range of 15.98% to 32.08%. The interquartile range was 21.58%-23.22%.

- Of 4,778 hospitals in the study cohort, 1 performed "better than the U.S. national rate," 3,766 performed "no different from the U.S. national rate," 133performed "worse than the U.S. national rate," and 779 were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing.
- The developer's interpretation of this data is that the variation in rates and number of performance outliers suggests there remain differences in the quality of care received across hospitals for HF that support measurement to reduce the variation.

Question for the Committee:

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

2b6. Comparability of data sources/methods:

- While the developer did not decide to include SDS variables in their final model, they did compare measure results with and without SDS adjustment.

2b7. Missing Data

- N/A

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

Comments: **Specifications are not inconsistent with evidence.

**The specifications are consistent with the title of the measure.

2a2. Reliability Testing

Comments: **see summary of validity testing

**The metric is a reliable measures of readmission given the completeness of medicare fee for service reporting.

2b2. Validity Testing

Comments: **see summary of validity testing

**The measure is already in use.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **no

**Meaningful differences is an important threat to validity. We do not know if a change in readmission indicates a change in quality of care.

Comparability to other performance measures is another threat. Dr. Krumholz has shown that those Medicare hospitals that do better on readmission rates do worse on mortality rates. The same was shown for VA vs. Medicare hospitals.

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 based on administrative claims data (e.g., DRG, ICD-9/10), which the developers note are routinely generated and collected as part of hospitals' billing processes.
- The developer indicates that all data elements are in defined fields in electronic claims.

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

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **This measure is based on administrative data. This measure is highly feasible.

**Highly feasible, already in use.

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

Publicly reported? ☒ **Yes** ☐ **No**

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

OR

Planned use in an accountability program? ☐ **Yes** ☐ **No**

Accountability program details

- Hospital Inpatient Quality Reporting (IQR) Program http://cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html and Hospital Readmission Reduction (HRRP) Program http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program.html

Improvement results

- The developer reports: "There has been significant progress in 30-day RSRR for HF. The median 30-day RSRR decreased by 1.6 absolute percentage points from July 2011-June 2012 (median RSRR: 23.1%) to July 2013-June 2014 (median RSRR: 21.5%). The median hospital RSRR from July 2011-June 2014 was 22.3% (IQR 21.6% - 23.2%)."

Feedback :

- During the 2012-2013 MAP review, MAP supported this measure for inclusion in the IQR and HRRP programs. The group agreed that the new specifications are an improvement over the existing finalized measure.

***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

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **Use of this performance measure for public reporting was associated with reduction in hospital RSRR by 1.6% between 2011-2012 and 2013-2014.

**Measure is being used to penalize hospitals by CMS.

Criterion 5: Related and Competing Measures

Related or competing measures

- None

Harmonization

- There are no related or competing measures, so harmonization is not necessary.

Pre-meeting public and member comments

Comment by: Ms. Elizabeth Godsey
Organization: Vizient, Inc.
Comment May 05, 2016: Vizient, Inc., the largest member-owned health care company in the country, is dedicated to serving members & customers through innovative data-driven solutions, expertise & collaborative opportunities that lead to improved patient outcomes & lower costs. Vizient requests CMS to review & provide follow-up analysis on more applied/practical alternate modeling approaches to account for within & across hospital variation besides hierarchical modeling. While hierarchical modeling is a valid technique controlling for within & across hospital variation, the approach lacks a tangible, practical framework of an observed to expected ratio that hospitals need to drive patient care. The predicted to expected approach complicates the public's & provider's understanding of how the actual observed values impacts hospital performance. Through numerous member discussions, we heard repeatedly, Oh, you mean that number does really reflect my actual readmissions? How can I improve that number? Even more concerning is the focus the current measure places on improving documentation & coding rather than patient care. Currently, providers see the only direct way to

improve the measure is through documentation & coding capture of co-morbidities which count toward the predicted & expected value calculations. We hope this was not the original intention of the measure & this misguided focus is simply an unintended artifact of an overly complicated modeling technique. We recommend analyzing & provide results comparing a model that uses hospital characteristics, such as teaching status or bed size to account for structural differences across hospitals & provide an observed to expected ratio which is much more meaningful for the public & providers. While in the past, CMS has commented they would not incorporate these features due to NQF restrictions; it is important to point out NQF has endorsed other risk adjustment models that incorporate these characteristics (NHSN) & consider these factors in the 30-day risk adjustment as well. Also, we would ask CMS & NQF to institute discrimination performance thresholds for the models given the importance these models bare on CMS's performance programs & public reporting. Currently, no model performs > 0.70, a standard considered fair-good practical performance threshold & while the c-stat does not fully evaluate the model, it certainly should require basic performance standards. Additionally, we ask CMS to provide performance statistics, like AIC, BIC & the Somers' D, Gamma & Tau-a association of predicted probabilities & observed counts for a more comprehensive assessment. Using these standards & model diagnostics, NQF can provide CMS with recommendations for improvement. Until minimum discrimination thresholds are instituted, we recommend NQF remove endorsement of the readmission measures.

Comment by: Ms. Elizabeth Godsey
Organization: Vizient, Inc.
Comment May 05, 2016: Vizient, Inc., the largest member-owned health care company in the country, is dedicated to serving members & customers through innovative data-driven solutions, expertise & collaborative opportunities that lead to improved patient outcomes & lower costs. For the readmission measures considered, CMS presented patient-level & hospital specific SES factor beta coefficients & p-values, yet overall model performance were not presented. We request the actual model performance results for model evaluation. For the AHRQ SES Index variable, we request further information on how the binary classification for a measure that ranges between 0-100 was determined & the impact of transforming into a binary representation vs. actual value had on the model performance. This detail along with the overall model performance information would provide the public with the necessary information to truly assess CMS's comment 'Given these findings & 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.' Regarding the complex pathways associated with 30-day readmissions as stated by CMS, we strongly ask CMS to entirely re-evaluate the utility of the 30-day measures. As stated by CMS, factors influencing readmissions are blurred between providers & patients 30-days post discharge resulting in a limited insights in how providers can improve care. We believe CMS's efforts to remove the planned readmissions PR4 logic is a strong step in true opportunity identification; however, more refinement is needed. We recommend a shorter, more actionable 7 day post-discharge readmission timeframe to pinpoint opportunities providers truly can influence & thus, mitigate many of SES confounding factors. The 7-day window provides clearer opportunities for patient stabilization & post-acute discharge planning which the 30-day window doesn't reflect. We recommend CMS provide a 7-day readmission risk adjustment for review. Also, the hospital wide readmission measure evaluates all readmissions within the 30-day window post inpatient discharge & considers readmit cases to also be eligible as the index admission; however, the condition specific measures evaluate only 1 readmit within the 30-day window & cannot be eligible as an index. We ask CMS for the rationale why the different approaches for the same measure as this adds unnecessary complexity which are impractical to manage. We recommend a consistent approach across all readmission measure calculations & recommend evaluating & counting all readmits that occur within the 30-day window so providers have a clear understanding of the # readmits are truly occurring. We support considering a readmit as an index for

the next 30-day cycle to again, assist organizations in tracking & improving complete patient care.

Comment by: Ms. Elizabeth Godsey
Organization: Vizient, Inc.
Comment May 05, 2016: Vizient, Inc., the largest member-owned health care company in the country, is dedicated to serving members & customers through innovative data-driven solutions, expertise & collaborative opportunities that lead to improved patient outcomes & lower costs. Vizient's coding expert reviewed the I-10 translations for the additional exclusions for the HF 30-day readmission measure & recommend not including I-10 code 5A02216, 5A02116 as these codes can be used as separate codes that extend beyond the LVAD patient population. For instance, these two codes can be used to capture intra-op & intra-procedure cardiac output during such procedures as valvuloplasty, angioplasty, intra-cardiac procedures & even during cardiac catheterization. These two codes can be coded when the cardiac output support is performed regardless of whether or not a ventricular device is inserted. Additionally, Vizient recommends removing the following I-10 lung transplant codes, 0BYM0Z0, 0BYMOZ1,& 0BYM0Z2 from the heart transplant I-9 code translation as these codes are specific to lung transplant in I-10 & do not involve the same I-9 combination code translations needed adequately capture heart transplantation. We reviewed the I-9 procedure codes used for the HF readmission criteria & recommend excluding 3762 as this inserted/removal occurs within the same encounter); thus, not reflective of a true bridge to heart transplant encounter which Vizient believes is the goal of this exclusion criteria. In reviewing the algorithm for AHRQ CCS potentially planned procedure list, AHRQ CCS 169 is listed as exclusion criteria, but within ICD-10 CCS 169 does not exist. Vizient recommends CMS and NQF reviewing this criterion and provide the appropriate ICD-10 translations to address the debridement of wound; infection or burn procedure codes.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0330

Measure Title: Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: N/A

Date of Submission: 1/29/2016

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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 heart failure (HF) hospitalization

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

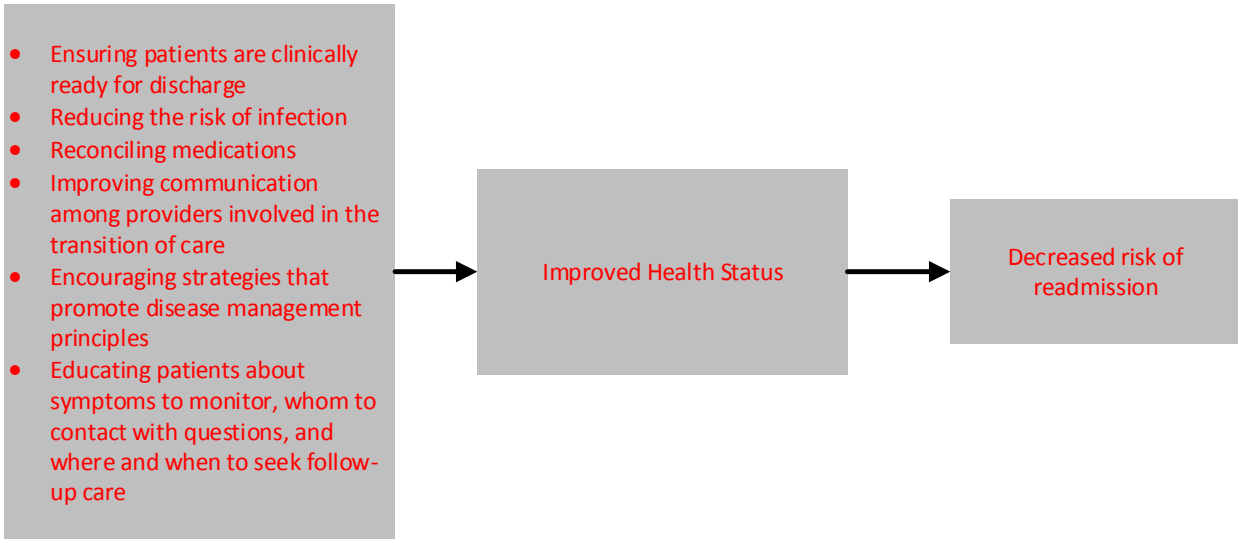
☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.



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 hospitalization for heart failure. 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.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

The diagram above indicates some of the many care processes that can influence readmission risk. In general, randomized controlled trials have shown that improvement in the following areas can directly reduce readmission rates: quality of care during the initial admission; improvement in communication with patients, their caregivers, and their clinicians; patient education; predischarge assessment; and coordination of care after discharge. Evidence that hospitals have been able to reduce readmission rates through these quality-of-care initiatives illustrates the degree to which hospital practices can affect readmission rates. Successful randomized trials have reduced 30-day readmission rates by 20-40% [1-11]. Since 2008, 14 Medicare Quality Improvement Organizations have been funded to focus on care transitions, applying lessons learned from clinical trials. Several have been notably successful in reducing

readmissions. The strongest evidence supporting the efficacy of improved discharge processes and enhanced care at transitions is a randomized controlled trial by the Project RED (Re-Engineered Discharge) intervention, in which a nurse was assigned to each patient as a discharge advocate, responsible for patient education, follow-up, medication reconciliation, and preparing individualized discharge instructions sent to the patient's primary care provider [1]. There was also a follow-up phone call from a pharmacist within 4 days of discharge. This intervention demonstrated a 30% reduction in 30-day readmissions [1]. Hospital processes that reflect the quality of inpatient and outpatient care such as discharge planning, medication reconciliation, and coordination of outpatient care have been shown to reduce readmission rates [12]. Although readmission rates are also influenced by hospital system characteristics, such as the bed capacity of the local health care system, these hospital characteristics should not influence quality of care [13]. Therefore, this measure does not risk adjust for such hospital characteristics.

The Medicare Payment Advisory Commission (MedPAC) has called for hospital-specific public reporting of readmission rates, identifying HF as a priority condition [14]. MedPAC finds that readmissions are common, costly, and often preventable. Based on 2005 Medicare data, MedPAC estimates that about 12.5% of Medicare heart failure (HF) admissions were followed by a readmission within 15 days, accounting for more than 90,000 admissions at a cost of $590 million. Between July 2005 and June 2008, the median 30-day readmission rate for heart failure was 24.4%, with a range of 15.9% to 34.4% [15].

HF incidence approaches 10 per 1000 of the population after 65 years of age [16]; prevalence of HF in the U.S. is estimated at nearly 6 million [17-18]. HF is the most common principal discharge diagnosis among older adults and the third highest for hospital reimbursements in 2005 [19-20], and the leading cause of readmission among Medicare beneficiaries, with nearly half of HF patients expected to return to the hospital within 6 months of discharge [21-22]. All-cause 30-day readmission rates per 1,000 patients discharged with HF increased by 11% between 1992 and 2001 [23]. HF readmission is a costly event and represents an undesirable outcome of care from the patient's perspective, and highly disparate HF readmission rates among hospitals suggest there is room for improvement [14, 24]. Moreover, there is substantial inter-hospital variation in the risk of readmission that is not clearly explained by differences in case mix.

The HF risk-standardized readmission rate (RSRR) measure is thus intended to inform quality-of-care improvement efforts, as individual process-based performance measures cannot encompass all the complex and critical aspects of care within a hospital that contribute to patient outcomes. Many stakeholders, including patient organizations, are interested in outcomes measures that allow patients and providers to assess relative outcomes performance for hospitals.

References:

1. Jack BW, Chetty VK, Anthony D, Greenwald JL, Sanchez GM, Johnson AE, et al. A reengineered hospital discharge program to decrease rehospitalization: a randomized trial. Ann Intern Med 2009;150(3):178-87.

2. Coleman EA, Smith JD, Frank JC, Min SJ, Parry C, Kramer AM. Preparing patients and caregivers to participate in care delivered across settings: the Care Transitions Intervention. J Am Geriatr Soc 2004;52(11):1817-25.

3. 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 2009;57(3):395-402.

4. 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.

5. Koehler BE, Richter KM, Youngblood L, Cohen BA, Prengler ID, Cheng D, 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. J Hosp Med 2009;4(4):211-218.

6. Mistiaen P, Francke AL, Poot E. Interventions aimed at reducing problems in adult patients discharged from hospital to home: a systematic metareview. BMC Health Serv Res 2007;7:47.

7. 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 1994;120(12):999-1006.

8. Naylor MD, Brooten D, Campbell R, Jacobsen BS, Mezey MD, Pauly MV, et al. Comprehensive discharge planning and home follow-up of hospitalized elders: a randomized clinical trial. Jama 1999;281(7):613-20.

9. 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 2002;17(3):186-92.

10. Weiss M, Yakusheva O, Bobay K. Nurse and patient perceptions of discharge readiness in relation to postdischarge utilization. Med Care 2010;48(5):482-6.

11. 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.

12. Nelson EA, Maruish ME, Axler JL. Effects of Discharge Planning and Compliance With Outpatient Appointments on Readmission Rates. Psychiatr Serv. July 1 2000;51(7):885-889.

13. Fisher ES, Wennberg JE, Stukel TA, Sharp SM. Hospital Readmission Rates for Cohorts of Medicare Beneficiaries in Boston and New Haven. New England Journal of Medicine. 1994;331(15):989-995.

14. MedPAC. Report to the Congress: Promoting Greater Efficiency in Medicare. Washington, DC: Medicare Payment Advisory Commission, 2007.

15. Krumholz HM, Merrill AR, Schone EM, Schreiner GC, Chen J, Bradley EH, Wang Y, Wang Y, Lin Z, Straube BM, Rapp MT, Normand SL, Drye EE. 2009. Patterns of hospital performance in acute myocardial infarction and heart failure 30-day mortality and readmission. Circ Cardiovasc Qual Outcomes (2):407-413.

16. National Heart, Lung, and Blood Institute. Unpublished tabulation of NHANES, 1971-1975, 1976-1980, 1988-1994, 1999-2002, 2003-2006, and extrapolation to the U.S. population, 2007.

17. Mozaffarian D, Benjamin EJ, Go AS, et al. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131(4):e29-322.

18. Lloyd-Jones D et al,American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics--2010 update: a report from the American Heart Association. Circulation. 2010 Feb 23;121(7):e46-e215. Epub 2009 Dec 17.

19. Jessup M, Brozena S. Medical progress: heart failure. N Engl J Med 2003;348:2007–18.

20. Centers for Medicare & Medicaid Services, Office of Information Services (OIS). Available at: http://www.cms.hhs.gov/MedicareFeeforSvcPartsAB/Downloads/SSDischarges0405.pdf, accessed October 21, 2006.

21. Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. N Engl J Med. 2009;360(14):1418-28.

22. Krumholz HM, Parent EM, Tu N, Vaccarino V, Wang Y, Radford MJ, Hennen J. Readmission after hospitalization for congestive heart failure among Medicare beneficiaries. Arch Intern Med. 1997;157:99-104.

23. Centers for Medicare & Medicaid Services, Mathematica Policy Research, Medicare Quality Monitoring System (MQMS) Report: Heart Failure, 1992-2001. Available at: http://www.mathematica-mpr.com/publications/PDFs/mqmsheart.pdf, accessed December 06, 2010.

24. Bernheim SM, Grady JN, Lin Z, Wang Y, Wang Y, Savage SV, Bhat KR, Ross JS, Desai MM, Merrill AR, Han LF, Rapp MT, Drye EE, Normand SL, Krumholz HM. National patterns of risk-standardized mortality and readmission for acute myocardial infarction and heart failure. Update on publicly reported outcomes measures based on the 2010 release. Circ Cardiovasc Qual Outcomes. 2010;3(5):459-67.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

N/A. This measure is not an intermediate outcome, process, or structure performance measure.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

N/A

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

N/A

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

N/A

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

N/A

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

N/A

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

N/A

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

N/A

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

N/A

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

N/A

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

N/A

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

N/A

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

N/A

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade: .

N/A

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

N/A

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: Click here to enter date range

N/A

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

N/A

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

N/A

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

N/A

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

N/A

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

N/A

1a.8 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.8.1 What process was used to identify the evidence?

N/A

1a.8.2. Provide the citation and summary for each piece of evidence.

N/A

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

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

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

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 HF. 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.

HF 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.*

Distribution of Hospital HF RSRRs over Different Time Periods
Results for each data year
Characteristic//07/2011-06/2012//07/2012-06/2013//07/2013-06/2014//07-2011-06/2014
Number of Hospitals/4,680/ /4,655/ /4,604/ /4,778/
Number of Admissions/413,033/ /405,188/ /392,233/ /1,210,454/
Mean (SD)/23.2 (1.2)/22.5 (1.0)/ /21.6 (0.9)/ /22.4 (1.5)/
Range (min. – max.)/18.4 – 29.6/ /17.5 – 28.7/ /17.1 – 26.8/ /16.0 – 32.1/

Minimum/18.4/ /17.5/ /17.1/ /16.0/
10th percentile/21.8/ /21.3/ /20.5/ /20.7/
20th percentile/22.4/ /21.8/ /21.0/ /21.4/
30th percentile/22.7/ /22.0/ /21.2/ /21.8/
40th percentile/22.9/ /22.2/ /21.4/ /22.1/
50th percentile/23.1/ /22.4/ /21.5/ /22.3/
60th percentile/23.4/ /22.6/ /21.7/ /22.6/
70th percentile/23.6/ /22.8/ /21.9/ /23.0/
80th percentile/24.0/ /23.1/ /22.2/ /23.5/
90th percentile/24.7/ /23.7/ /22.7/ /24.3/
Maximum/29.6/ /28.7/ /26.8/ /32.1/

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

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for 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.*
Distribution of HF RSRRs by Proportion of Dual Eligible Patients:
Dates of Data: July 2013 through June 2014
Data Source: Medicare FFS claims

Characteristic//Hospitals with a low proportion (≤8.7%) Dual Eligible patients//Hospitals with a high proportion (≥ 22.0%) Dual Eligible patients
Number of Measured Hospitals// 997 // 1,003
Number of Patients// 333,931 patients in low-proportion hospitals/ 195,234 in high-proportion hospitals
Maximum// 28.1// 32.1
90th percentile// 24.1// 24.9
75th percentile// 23.1 // 23.9
Median (50th percentile)// 22.1// 22.8
25th percentile// 21.2// 22.0
10th percentile// 20.3// 21.2
Minimum // 16.0 // 18.0

Distribution of HF RSRRs by Proportion of African-American Patients:
Dates of Data: July 2013 through June 2014
Data Source: Medicare FFS claims

Characteristic// Hospitals with a low proportion (≤0.0%) African-American patients//Hospitals with a high proportion (≥13.2%) African-American patients
Number of Measured Hospitals// 1,067 // 999
Number of Patients// 99,898 patients in low-proportion hospitals/ 373,385 in high-proportion hospitals
Maximum// 26.6// 28.6
90th percentile// 23.7// 25.1
75th percentile// 22.8// 23.9
Median (50th percentile)// 22.0// 22.9

25th percentile// 21.3// 21.9
10th percentile// 20.6// 20.9
Minimum // 18.0// 18.4

Distribution of HF RSRRs by Proportion of Patients with AHRQ SES Index Scores Below 42.7:
Dates of Data: July 2013 through June 2014
Data Source: Medicare FFS claims and the American Community Survey (2008-2012) data

Characteristic//Hospitals with a low proportion of patients below AHRQ SES index score of 42.7 (≤9.2%)// Hospitals with a high proportion of patients below AHRQ SES index score of 42.7 (≥38.3%)
Number of Measures Hospitals// 999 // 999
Number of Patients// 257,667 patients in hospitals with low proportion of patients below AHRQ SES index score of 42.7 //218,581 patients in hospitals with high proportion of patients below AHRQ SES index score of 42.7
Maximum// 27.7// 32.1
90th percentile// 23.7// 25.1
75th percentile// 22.8// 24.0
Median (50th percentile)// 21.9// 22.8
25th percentile// 21.2// 22.0
10th percentile// 20.3// 21.2
Minimum // 16.0// 18.4

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, High resource use, Severity of illness, Patient/societal consequences of poor quality
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.**
The Medicare Payment Advisory Commission (MedPAC) has called for hospital-specific public reporting of readmission rates, identifying HF as a priority condition (MedPAC, 2007). MedPAC finds that readmissions are common, costly, and often preventable. Based on 2005 Medicare data, MedPAC estimates that about 12.5% of Medicare HF admissions were followed by a readmission within 15 days, accounting for more than 90,000 admissions at a cost of $590 million. Between July 2005 and June 2008, the median 30-day readmission rate for heart failure was 24.4%, with a range of 15.9% to 34.4% (Krumholz et al., 2009).

HF is the most common principal discharge diagnosis among older adults and the third highest for hospital reimbursements in 2005 (CMS/OIS, 2006), and the leading cause of readmission among Medicare beneficiaries, with nearly half of HF patients expected to return to the hospital within 6 months of discharge (Jencks et al., 2009; Krumholz et al., 1997). All-cause 30-day readmission rates per 1,000 patients discharged with HF increased by 11% between 1992 and 2001 (CMS/MPR/MQMS, 2003). HF readmission is a costly event and represents an undesirable outcome of care from the patient's perspective, and highly disparate HF readmission rates among hospitals suggest there is room for improvement (MedPAC, 2007; Bernheim et al., 2010).

Readmission rates are influenced by the quality of inpatient and outpatient care, the availability and use of effective disease management programs, and the bed capacity of the local healthcare system. Some of the variation in readmissions may be attributable to delivery system characteristics (Fisher et al., 1994). Also, interventions during and after a hospitalization can be effective in reducing readmission rates in geriatric populations generally (Benbassat and Taragin, 2000; Naylor et al., 1999; Coleman et al., 2006; Courtney et al., 2009; Jack et al., 2009; Voss et al., 2011) and for heart failure patients specifically (Gonseth et al., 2004; Phillips et al., 2004; Koelling et al., 2005; Jovicic et al., 2006). Moreover, such interventions can be cost saving (Coleman et al., 2006; Naylor et al., 1999;). Tracking readmissions also emphasizes improvement in care transitions and care coordination. Although discharge planning is required by Medicare as a condition of participation for hospitals, transitional care focuses more broadly on "hand-offs" of care from one setting to another and may have implications for quality and costs (Coleman, 2005).

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

MedPAC. Report to the Congress: Promoting Greater Efficiency in Medicare. Washington, DC: Medicare Payment Advisory Commission, 2007.

Krumholz HM, Merrill AR, Schone EM, Schreiner GC, Chen J, Bradley EH, Wang Y, Wang Y, Lin Z, Straube BM, Rapp MT, Normand SL, Drye EE. 2009. Patterns of hospital performance in acute myocardial infarction and heart failure 30-day mortality and readmission. Circ Cardiovasc Qual Outcomes (2):407-413.

Centers for Medicare & Medicaid Services, Office of Information Services (OIS). Available at: http://www.cms.hhs.gov/MedicareFeeforSvcPartsAB/Downloads/SSDischarges0405.pdf, accessed October 21, 2006.

Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. N Engl J Med. 2009;360(14):1418-28.

Krumholz HM, Parent EM, Tu N, Vaccarino V, Wang Y, Radford MJ, Hennen J. Readmission after hospitalization for congestive heart failure among Medicare beneficiaries. Arch Intern Med. 1997;157:99-104.

Centers for Medicare & Medicaid Services, Mathematica Policy Research, Medicare Quality Monitoring System (MQMS) Report: Heart Failure, 1992-2001. Available at: http://www.mathematica-mpr.com/publications/PDFs/mqmsheart.pdf, accessed December 06, 2010.

Bernheim SM, Grady JN, Lin Z, Wang Y, Wang Y, Savage SV, Bhat KR, Ross JS, Desai MM, Merrill AR, Han LF, Rapp MT, Drye EE, Normand SL, Krumholz HM. National patterns of risk-standardized mortality and readmission for acute myocardial infarction and heart failure. Update on publicly reported outcomes measures based on the 2010 release. Circ Cardiovasc Qual Outcomes.
2010;3(5):459-67.

Fisher ES, Wennberg JE, Stukel TA, et al. 1994. Hospital readmission rates for cohorts of Medicare beneficiaries in Boston and New Haven. N Engl J Med 331(15):989-995.

Benbassat J, Taragin M. 2000. Hospital readmissions as a measure of quality of health care: advantages and

limitations. Arch Intern Med 160(8):1074-1081.

Naylor MD, Brooten D, Campbell R, et al. 1999. Comprehensive discharge planning and home follow-up of hospitalized elders: a randomized clinical trial. JAMA 281(7):613-620.

Coleman EA, Parry C, Chalmers S, et al. 2006. The care transitions intervention: results of a randomized controlled trial. Arch Intern Med 166:1822-1828.

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 2009;57(3):395-402.

Jack BW, Chetty VK, Anthony D, Greenwald JL, Sanchez GM, Johnson AE, et al. A reengineered hospital discharge program to decrease rehospitalization: a randomized trial. Ann Intern Med 2009;150(3):178-87.

Voss R, Gardner R, Baier R, Butterfield K, Lehrman S, Gravenstein S. The care transitions intervention: translating from efficacy to effectiveness. Arch Intern Med Jul 25 2011;171(14):1232-1237.

Gonseth J, Guallar-Castillon P, Banegas JR, Rodriguez-Artalejo F. The effectiveness of disease management programmes in reducing hospital re-admission in older patients with heart failure: a systematic review and meta-analysis of published reports. Eur Heart J. 2004;25:1570-95.

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. 2004;291:1358-67.

Koelling T, Johnson M, Cody R, Aaronson K. Discharge education improves clinical outcomes in patients with chronic heart failure. Circulation. 2005;111:179-85.

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.

Coleman EA. 2005. Background Paper on Transitional Care Performance Measurement. Appendix I. In: Institute of Medicine, Performance Measurement: Accelerating Improvement. Washington, DC: National Academy Press.

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 : Congestive Heart Failure

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Care Coordination : Readmissions, Safety, Safety : Readmissions

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:** NQF_0330_HF_Readmission_S2b_Data_Dictionary_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.
Annual Updates
1. Each year we updated to the most current version of the Agency for Healthcare Research and Quality Clinical Classifications Software (AHRQ CCS) software by identifying any changes from the previous version that might affect the measure.
2. In addition, we have updated the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) Hierarchical Condition Categories (HCC) map annually to capture any changes that might affect the measure's risk model. The version of the HCC map used for this measure has not been changed since 2013.

Updates by Year
2015
1. Respecified the measure by updated to CMS Planned Readmission Algorithm (Version 4.0).
Rationale: Version 4.0 incorporates improvements made following a validation study of the algorithm using data from a medical record review and input from clinical experts. These changes improve the accuracy of the algorithm by decreasing the number of readmissions that the algorithm mistakenly designates as planned/unplanned by removing five procedure categories and adding one procedure category.

2. Updated cohort to exclude patients with an LVAD implantation or heart transplantation either during the index admission or in the 12 months prior to the index admission.
Rationale: The use of LVADs, in particular, has increased dramatically since the time of measure development. These patients represent a clinically distinct, highly-selected group of patients.

2014
1. Updated to CMS Planned Readmission Algorithm (Version 3.0).
Rationale: Version 3.0 incorporates improvements made following a validation study of the algorithm using data from a medical record review. These changes improve the accuracy of the algorithm by decreasing the number of readmissions that the algorithm mistakenly designated as planned by removing two procedure categories and adding several acute diagnoses.

2013
1. Updated to CMS Planned Readmission Algorithm (Version 2.1).
Rationale: Version 2.1 incorporated improvements to the original algorithm made following an extensive review by clinical experts and stakeholder feedback submitted during the hospital wide-readmission measure's public comment period and 2012 dry run.

2. Modified the planned readmission algorithm handling of admissions to psychiatric and rehabilitation hospitals.
Rationale: Psych and rehab hospitals in Maryland have the same provider ID number as acute care hospitals. Therefore, readmissions are not counted if the patient has a principal diagnosis code beginning with a "V57" (indication of admission to a rehab unit) or if all three of the following criteria are met: (1) the admission being evaluated as a potential readmission has a psychiatric principal discharge diagnosis code (ICD-9 codes 290-319); (2) the index admission has a discharge disposition code to a psychiatric hospital or psychiatric unit from the index admission; and (3) the admission being evaluated as a potential readmission occurred during the same day as or the day following the index discharge. The criteria for identifying such admissions are available in the 2010 Measures Maintenance Technical Report: Acute Myocardial Infarction, Heart Failure, and Pneumonia 30-Day Risk-Standardized Readmission Measures.

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 readmission. We define readmission as an inpatient admission for any cause, with the exception of certain planned readmissions, within 30 days from the date of discharge from the index HF admission. If a patient has more than one unplanned admissions (for any reason) within 30 days after discharge from the index admission, only one is counted as a readmission. The measure looks for a dichotomous yes or no outcome of whether each admitted patient has an unplanned readmission within 30 days. However, if the first readmission after discharge is considered planned, any subsequent unplanned readmission is not counted as an outcome for that index admission, because the unplanned readmission could be related to care provided during the intervening planned readmission rather than during the index admission.

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 HF hospitalization.

Denominator Time Window: This measure was developed with 12 months of data. 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.
The measure counts readmissions to any acute care hospital for any cause within 30 days of the date of discharge of the index HF admission, excluding planned readmissions as defined below.

Planned Readmission Algorithm (Version 4.0)
The Planned Readmission Algorithm is a set of criteria for classifying readmissions as planned among the general Medicare population using Medicare administrative claims data. The algorithm identifies admissions that are

typically planned and may occur within 30 days of discharge from the hospital.

The Planned Readmission Algorithm has three fundamental principles:

1. A few specific, limited types of care are always considered planned (transplant surgery, maintenance chemotherapy/immunotherapy, rehabilitation);
2. Otherwise, a planned readmission is defined as a non-acute readmission for a scheduled procedure; and
3. Admissions for acute illness or for complications of care are never planned.

The algorithm was developed in 2011 as part of the Hospital-Wide Readmission measure. In 2013, CMS applied the algorithm to its other readmission measures. In applying the algorithm to condition- and procedure-specific measures, teams of clinical experts reviewed the algorithm in the context of each measure-specific patient cohort and, where clinically indicated, adapted the content of the algorithm to better reflect the likely clinical experience of each measure's patient cohort.
For the heart failure readmission measure, CMS used the Planned Readmission Algorithm without making any changes.

The Planned Readmission Algorithm and associated code tables are attached in data field S.2b (Data Dictionary or Code Table). For more details on the Planned Readmission Algorithm, please see the report titled "2015 Condition-Specific Measures Updates and Specifications Report Hospital-Level 30-Day Risk-Standardized Readmission Measures for HF, version 4.0" posted in data field A.1 or at https://www.qualitynet.org/dcs/BlobServer?blobkey=id&blobnocache=true&blobwhere=1228890435217&blobheader=multipart%2Foctet-stream&blobheadername1=Content-Disposition&blobheadervalue1=attachment%3Bfilename%3DRdmn_AMIHFPNCOPDSTK_Msr_UpdtRpt.pdf&blobcol=urldata&blobtable=MungoBlobs.

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 explicitly tested the measure in both age groups.

The cohort includes admissions for patients aged 18 years and older discharged from the hospital with either a principal discharge diagnosis of HF (see codes below) and with a complete claims history for the 12 months prior to admission. The measure is currently publicly reported by CMS for those patients 65 years and older who are Medicare FFS beneficiaries admitted to non-federal hospitals or Veterans Health Administration (VA) hospitals.

Additional details are provided in S.9 Denominator Details.

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

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)*
To be included in the measure cohort used in public reporting, patients must meet the following additional inclusion criteria:

1. Having a principal discharge diagnosis of heart failure;
2. Enrolled in Medicare FFS Part A and Part B for the 12 months prior to the date of the admission, and enrolled in Part A during the index admission;
3. Aged 65 or over;
4. Discharged alive from a non-federal short-term acute care hospital; and

5. Not transferred to another acute care facility.

This measure can also be used for an all-payer population aged 18 years and older. We have explicitly tested the measure in both patients aged 18 years and older and those aged 65 years or older (see Testing Attachment for details).

International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) codes used to define the cohort for each measure are:
ICD-9-CM codes used to define HF:
402.01 Malignant hypertensive heart disease with heart failure
402.11 Benign hypertensive heart disease with heart failure
402.91 Unspecified hypertensive heart disease with heart failure
404.01 Hypertensive heart and chronic kidney disease, malignant, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified
404.03 Hypertensive heart and chronic kidney disease, malignant, with heart failure and with chronic kidney disease stage V or end stage renal disease
404.11 Hypertensive heart and chronic kidney disease, benign, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified
404.13 Hypertensive heart and chronic kidney disease, benign, with heart failure and chronic kidney disease stage V or end stage renal disease
404.91 Hypertensive heart and chronic kidney disease, unspecified, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified
404.93 Hypertensive heart and chronic kidney disease, unspecified, with heart failure and chronic kidney disease stage V or end stage renal disease
428.0 Congestive heart failure, unspecified
428.1 Left heart failure
428.20 Systolic heart failure, unspecified
428.21 Acute systolic heart failure
428.22 Chronic systolic heart failure
428.23 Acute on chronic systolic heart failure
428.30 Diastolic heart failure, unspecified
428.31 Acute diastolic heart failure
428.32 Chronic diastolic heart failure
428.33 Acute on chronic diastolic heart failure
428.40 Combined systolic and diastolic heart failure, unspecified
428.41 Acute combined systolic and diastolic heart failure
428.42 Chronic combined systolic and diastolic heart failure
428.43 Acute on chronic combined systolic and diastolic heart failure
428.9 Heart failure, unspecified

ICD-10 Codes that define the patient cohort:
I110 Hypertensive heart disease with heart failure
I130 Hypertensive heart and chronic kidney disease with heart failure and stage 1 through stage 4 chronic kidney disease, or unspecified chronic kidney disease
I132 Hypertensive heart and chronic kidney disease with heart failure and with stage 5 chronic kidney disease, or end stage renal disease
I509 Heart failure, unspecified
I501 Left ventricular failure
I5020 Unspecified systolic (congestive) heart failure
I5021 Acute systolic (congestive) heart failure
I5022 Chronic systolic (congestive) heart failure
I5023 Acute on chronic systolic (congestive) heart failure

I5030 Unspecified diastolic (congestive) heart failure
I5031 Acute diastolic (congestive) heart failure
I5032 Chronic diastolic (congestive) heart failure
I5033 Acute on chronic diastolic (congestive) heart failure
I5040 Unspecified combined systolic (congestive) and diastolic (congestive) heart failure
I5041 Acute combined systolic (congestive) and diastolic (congestive) heart failure
I5042 Chronic combined systolic (congestive) and diastolic (congestive) heart failure
I5043 Acute on chronic combined systolic (congestive) and diastolic (congestive) heart failure

An ICD-9 to ICD-10 crosswalk is attached in field S.2b. (Data Dictionary or Code Table).

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
The readmission measures excludes admissions:

1. Ending in discharges against medical advice
Rationale: Providers did not have the opportunity to deliver full care and prepare the patient for discharge.
2. Without at least 30 days of 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.
3. Occurring within 30 days of discharge from an index admission
Rationale: This exclusion ensures that no hospitalization will be considered as both a readmission and an index admission within the same measure.
4. With a procedure code for LVAD implantation or heart transplantation either during the index admission or in the 12 months prior to the index admission
Rationale: Patients with these procedures are a highly-selected group of patients with a different risk of the readmission outcome.

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)*
1. Discharges against medical advice are identified using the discharge disposition indicator in claims data.
2. Admissions without at least 30 days post-discharge enrollment in FFS Medicare are determined by examining the Medicare Enrollment Database (EDB).
3. Admissions within 30 days of discharge from a qualifying index admission are identified by comparing the discharge date from the index admission with subsequent admission dates.
4. Procedure codes for LVAD implantation or heart transplantation are identified by the corresponding codes included in claims data. The list of codes used is attached in field S.2b. (Data Dictionary or Code Table).

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 employs a hierarchical logistic regression model to create a hospital-level 30-day RSRR. In brief, the approach simultaneously models data at the patient and hospital levels to account for the variance in patient outcomes within and between hospitals (Normand & Shahian, 2007). At the patient level, the model adjusts the log-odds of readmission within 30 days of discharge for age and selected clinical covariates. At the hospital level, the approach models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of readmission at the hospital, after accounting for patient risk. If there were no differences among hospitals, then after adjusting for patient risk, the hospital intercepts should be identical across all hospitals.

Candidate and Final Risk-adjustment Variables: Candidate variables were patient-level risk-adjustors that were 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 claims records extending 12 months prior to and including the index admission. For the measure currently implemented by CMS, these risk adjusters are identified using both inpatient and outpatient Medicare FFS claims data. However, in the all-payer hospital discharge database measure, the risk-adjustment variables can be obtained only from inpatient claims in the prior 12 months and the index admission.

The model adjusts for case-mix differences based on the clinical status of patients 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 (Pope et al., 2000). A file that contains a list of the ICD-9-CM codes and their groupings into CCs is attached in data field S.2b (Data Dictionary or Code Table). In addition, only comorbidities that convey information about the patient at admission or in the 12 months prior, and not complications that arise during the course of the index hospitalization, are included in the risk adjustment. Hence, we do not risk adjust for CCs that may represent adverse events of care when they are only recorded in the index admission.

The final set of risk-adjustment variables is:

Demographics
Age-65 (years, continuous) for patients aged 65 or over cohorts; or Age (years, continuous) for patients aged 18 and over cohorts; Male (%)

Comorbidities
History of Coronary Artery Bypass Graft (CABG) surgery (ICD-9 diagnosis code V45.81; ICD-9 procedure codes 36.10-36.16)
Cardio-respiratory failure and shock (CC 79)
Congestive heart failure (CC 80)
Acute coronary syndrome (CC 81-82)
Coronary atherosclerosis or angina (CC 83-84)
Valvular or rheumatic heart disease (CC 86)
Specified arrhythmias and other heart rhythm disorders (CC 92-93)
Other or unspecified heart disease (CC 94)
Vascular or circulatory disease (CC 104-106)
Metastatic cancer or acute leukemia (CC 7)
Cancer (CC 8-12)
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)
Liver or biliary disease (CC 25-30)

Peptic ulcer, hemorrhage, other specified gastrointestinal disorders (CC 34)
Other gastrointestinal disorders (CC 36)
Severe hematological disorders (CC 44)
Iron deficiency or other unspecified anemias and blood disease (CC 47)
Dementia or other specified brain disorders (CC 49-50)
Drug/alcohol abuse/dependence/psychosis (CC 51-53)
Major psychiatric disorders (CC 54-56)
Depression (CC 58)
Other psychiatric disorders (CC 60)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)
Stroke (CC 95-96)
Chronic Obstructive Pulmonary Disease (COPD) (CC 108)
Fibrosis of lung or other chronic lung disorders (CC 109)
Asthma (CC 110)
Pneumonia (CC 111-113)
Dialysis status (CC 130)
Renal failure (CC 131)
Nephritis (CC 132)
Other urinary tract disorders (CC 136)
Decubitus ulcer or chronic skin ulcer (CC 148-149)

References:
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.

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

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*)

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.)*
The measure estimates hospital-level 30-day all-cause RSRRs following hospitalization for HF using hierarchical logistic regression models. In brief, the approach simultaneously models data at the patient and hospital levels to

account for variance in patient outcomes within and between hospitals (Normand and Shahian, 2007). At the patient level, it models the log-odds of readmission within 30 days of discharge from the index admission using age, selected clinical covariates, and a hospital-specific intercept. At the hospital level, it models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of a readmission at the hospital, after accounting for patient risk. The hospital-specific intercepts are given a distribution 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.

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

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

This calculation transforms the ratio of predicted over expected into a rate that is compared to the national observed readmission rate. The hierarchical logistic regression models are described fully in the original methodology report (Grosso et al., 2011).

References:

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

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

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.

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 survey or patient-reported data.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
N/A. This measure is not based on a survey or patient-reported data.

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

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:

1. 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.

2. 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).

3. 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.

4. 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/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
0330_MeasureTesting_MSF5.0_Data-635796500435377019.doc,NQF_0330_HF_Readmission_NQF_Testing_Attachment_v1.0.docx

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

Measure Number (*if previously endorsed*)**:** 0330
Measure Title: Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization
Date of Submission: 1/29/2016
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☒ Outcome (*including PRO-PM*)
☐ 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, Veterans' Health Administration claims, 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?

The dates used vary by testing type; see Section 1.7 for details.

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

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

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

For this measure, hospitals are the measured entities. All non-federal, acute inpatient US hospitals (including territories) with Medicare fee-for-service (FFS) beneficiaries aged 65 years and older are included. The number of measured entities (hospitals) varies by testing type; see Section 1.7 for details.

1.6. How many and which patients 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:

For reliability testing (Section 2a2)

The reliability of the model was tested by randomly selecting 50% of the Medicare patients aged 65 years or over in the most recent 3-year cohort and developing a risk-adjusted model for this group. We then developed a second model for the remaining 50% of patients and compared the two. Thus, for reliability testing, we randomly split **Dataset 1** into two samples. In each year of measure reevaluation, we also re-fit the model and compared the frequencies and model coefficients of risk variables (condition categories for patient comorbidities) and model fit across 3 years (**Dataset 1** below).

Dataset 1 (2015 public reporting cohort): Medicare Part A Inpatient and Outpatient and Part B Outpatient claims
Dates of Data: July 1, 2011 – June 30, 2014
Number of Admissions: 1,210,454
Patient Descriptive Characteristics: average age= 80.8, %male= 46.65
Number of Measured Entities: 4,778

For validity testing (Section 2b2)

Dataset 2 (medical record validation): National Heart Failure (NHF) Dataset for clinical data from HF hospital admissions, linked with the Medicare Part A Inpatient and Outpatient and Part B Outpatient claims and the Medicare Enrollment Database to assess the readmission outcome.
Dates of Data: 1998-2001
Number of Admissions: N=64,329 cases matched to FFS Medicare claims
Patient Descriptive Characteristics: %male=41.73%

For testing of measure exclusions (Section 2b3)

Dataset 1 (2015 public reporting cohort)

For testing of measure risk adjustment (Section 2b4)

Dataset 1 (2015 public reporting cohort)

Dataset 3 (development dataset): Medicare Part A Inpatient and Outpatient and Part B Outpatient claims
Dates of Data: 2004
Number of Admissions: N=283,919 (first half of split sample); N=283,528 (second half of split sample)
Number of Measured Entities: 4,669 hospitals (first half of split sample); 4,680 hospitals (second half of split samples); 4,730 hospitals in full 2004 dataset

To create the model development sample (**Dataset 3**), we applied the inclusion and exclusion criteria to all 2004 admissions. We randomly selected half of all HF admissions in 2004 that met the inclusion and exclusion criteria to create a model development sample and used the remaining admissions as our model validation sample.

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

Dataset 4 (all payer dataset, section 2b4.11): California Patient Discharge Data in addition to CMS Medicare FFS data for patients in California hospitals

Dates of Data: January 1, 2006 – December 31, 2006
Number of Discharges: 76,536 (all 18+ total); 33,784 (FFS 65+); 20,989 (non-FFS 65+); 21,763 (all 18-64)
Patient Descriptive Characteristics: mean age=72, %male=50 (all 18+ total); mean age=80, %male=44 (FFS 65+); mean age=80, %male=47 (non-FFS 65+); mean age=53, %male=61 (all 18-64)
Number of Measured Entities: >450 non-Federal acute care hospitals

The measure was applied to California Patient Discharge Data, a large, linked all-payer database of patient hospital admissions. Records are linked by a unique patient identification number, allowing us to determine patient history from previous hospitalizations.

For testing to identify meaningful differences in performance (Section 2b5)

Dataset 1 (2015 public reporting cohort)

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

Dataset 1 (2015 public reporting cohort); **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 1**)
• Validated AHRQ SES index score is a composite of 7 different variables found in the census data (**Dataset 5**)

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 SES variables.

The literature directly examining how different SES factors or race might influence the likelihood of older, insured, Medicare patients being readmitted within 30 days of the index hospitalization for heart failure is limited. However, several studies have indicated that both SES

variables and race variables are associated with increased risk of readmission among patients admitted for heart failure (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 variables' effects are described below in Section 2b4.3.

The SES and race variables used for analysis were:

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

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 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, 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 variable that describes the average SES of people living in defined geographic areas (Bonito et al., 2008). Its value as a proxy for patient-level information is dependent on having the most granular level data with respect to communities that patients live in. In this submission, we present analysis using the census block level, the most granular level possible using American Community Survey data. We used 2009-2013 American Community Survey data and mapped patients' 9-digit ZIP codes via vendor software to the AHRQ SES Index at the census block group level. Given the variation in cost of living across the country, the median income and median property value components of the AHRQ SES Index were adjusted by regional price parity values published by the Bureau of Economic Analysis (BEA). This provides a better marker of low SES neighborhoods in high expense geographic areas. We then calculated an AHRQ SES Index score for census block groups that can be linked to 9-digit ZIP codes. In the HF measure cohort, we were able to assign an AHRQ SES Index score to 99.4% of patient admissions. 87.4% of patient

admissions had calculated AHRQ SES Index scores linked to their 9-digit ZIP codes. 12.0% of patient admissions had only valid 5-digit ZIP codes; we utilized the data for the median 9-digit ZIP code within that 5-digit ZIP code.

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.

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

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**)

The frequency of some model variables increased while others decreased between 2011 and 2014, which may reflect an increased or decreased rate of specific comorbidities in the FFS population. For example, there was a notable increase in the frequency (≥ 2%) of cardio-respiratory failure or shock (from 27.2% to 29.6%) and of other psychiatric disorders (from 17.4% to 21.4%). There was a notable decrease in the frequency of fibrosis of lung or other chronic lung disorders (from 11.7% to 9.6%) and of other urinary tract disorders (from 34.1% to 32.1%). Examination of the odds ratios for each risk variable in the model shows that, overall, the odds ratios for individual risk variables remained relatively constant across the three years.

For the model variable frequencies, see the *2015 Measure Updates and Specifications Report* (Dorsey et al., 2015) attached to this submission. Note that these frequencies reflect the measure results calculated without applying the new exclusion criterion for patients who received left ventricular assist devices or transplant. That exclusion criterion will first be applied in the 2016 publicly reported measure.

Measure Score Reliability Results

There were 1,210,454 admissions in the 3-year split sample (from **Dataset 1**), with 604,022 admissions from 4,028 hospitals in one sample and 606,432 admissions from 4,060 hospitals in the other randomly selected sample. The agreement between the two RSRRs for each hospital was 0.58, which according to the conventional interpretation is "moderate" (Landis & Koch, 1977).

Note that this analysis was limited to hospitals with 12 or more cases in each split sample. The intra-class correlation coefficient is based on a split sample of three years of data, resulting in a volume of patients in each sample equivalent to only 1.5 years of data, whereas the measure is reported with the full three years of data.

References:

Dorsey K, Grady J, Desai N, et al. 2015 Condition-Specific Measures Updates and Specifications Report Hospital-Level 30-Day Risk-Standardized Readmission Measures Acute Myocardial Infarction – Version 8.0 Heart Failure – Version 8.0 Pneumonia – Version 8.0 Chronic Obstructive Pulmonary Disease – Version 8.0 Stroke – Version 8.0. 2014; https://www.qualitynet.org/dcs/BlobServer?blobkey=id&blobnocache=true&blobwhere=1228890435217&blobheader=multipart%2Foctet-stream&blobheadername1=Content-Disposition&blobheadervalue1=attachment%3Bfilename%3DRdmn_AMIHFPNCOPDSTK_Msr_UpdtRpt.pdf&blobcol=urldata&blobtable=MungoBlobs. Accessed October 30, 2015.

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 odds ratios over time suggests that the underlying data elements are reliable. Additionally, the ICC score demonstrates moderate agreement across samples using a conservative approach to assessment.

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 medical record validation.

Medical Record Validation

During original measure development we validated the HF readmission administrative model (original model specification prior to completion of the planned readmission algorithm) against a medical record model with the same cohort of patients for which hospital-level HF readmission medical record data are available (**Dataset 2**). We developed a medical record measure to compare with the administrative measure. We developed a measure cohort with the medical record data using the inclusion/exclusion criteria and risk-adjustment strategy that was consistent with the claims-based administrative measure but using chart-based risk adjusters, such as blood pressure, not available in the claims data. We then matched a sample of the same patients in the administrative data for comparison. The matched sample included 64,329 patients. We compared the output of the two measures, the state-level performance results, in the same group of patients. Specifically, we assessed the areas under the receiver operating characteristic (ROC) curve for the two models, the predictive ability comparing readmission rates in the lowest predicted decile and the highest predicted decile. We estimated hospital-level RSRRs using the corresponding hierarchical logistic regression administrative and medical record models for the linked patient sample. We then examined the linear relationship between the two sets of estimates using regression techniques and weighting by the total number of cases in each hospital.

References:

Jencks SF, Cuerdon T, Burwen DR et al. Quality of medical care delivered to Medicare beneficiaries: a profile at state and national levels. JAMA. 2000;284:1670-1676.

Jencks SF, Huff ED, Cuerdon T. Change in the quality of care delivered to Medicare beneficiaries, 1998-1999 to 2000-2001. JAMA. 2003;289:305-312.

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

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

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

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

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.

Krumholz HM, Brindis RG,Brush JE, et al. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. *Circulation.* January 24, 2006 2006;113(3):456-462.

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

ICD-9 to ICD-10 Conversion

Statement of Intent

[X] Goal was to convert this measure to a new code set, fully consistent with the intent of the original measure.
[ ] Goal was to take advantage of the more specific code set to form a new version of the measure, but fully consistent with the original intent.
[ ] The intent of the measure has changed.

Process of Conversion

ICD-10 codes were initially identified using 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. Each year we reexamine the codes using the latest version of the GEM mapping software. This was done most recently in 2015. An ICD-9 to ICD-10 crosswalk is attached in field S.2b. (Data Dictionary or Code Table).

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

Medical Record Validation

The performance of the administrative and medical record models was similar. The areas under the receiver operating characteristic (ROC) curve were 0.61 and 0.58, respectively, for the two models. In addition, they were similar with respect to predictive ability. For the administrative model, the predicted readmission rate ranged from 15% in the lowest predicted decile to 38% in the highest predicted decile, a range of 23%. For the medical record model, the corresponding range was 16% to 34%, a range of 18%.

We estimated hospital-level RSRRs using the corresponding hierarchical logistic regression administrative and medical record models for the linked patient sample. We then examined the

linear relationship between the two sets of estimates using regression techniques and weighting by the total number of cases in each hospital. The correlation coefficient of the standardized rates from the administrative and medical record models was 0.97.

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?*)

The results between the administrative and medical record models proved to be similar in each of the model testing that was performed. The ROC results were nearly identical and in line with other readmission models. The correlation between the resulting RSRRs calculated from both models was 0.97 which shows the resulting measure from the administrative claims model is as good as that from the medical record model.

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*)

All exclusions were determined by careful clinical review and have been made based on clinically relevant decisions and to ensure accurate calculation of the measure. To ascertain impact of exclusions on the cohort, we examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion (**Dataset 1**). These exclusions are consistent with similar NQF-endorsed outcome measures. For more details see the attached specifications report.

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*)

In **Dataset 1** (2015 public reporting cohort):

Exclusion	N	%	Distribution across hospitals (N=4,041): Minimum, 25th percentile, 50th percentile, 75th percentile, maximum
1. Discharged against medical advice (AMA)	5,868	0.44 %	(0.0, 0.0, 0.0, 0.6, 8.6)
2. Without at least 30 days post-discharge enrollment in FFS Medicare for index admissions	6,681	0.50 %	(0.0, 0.0, 0.3, 0.8, 11.1)
3. Heart failure admission within 30 days of a prior heart failure index admission	102,005	7.68 %	(0.0, 5.4, 7.1, 8.7, 24.7)

4. Left ventricular assist device (LVAD) or transplant in index admission or prior year	2,371	0.18 %	(0.0, 0.0, 0.0, 0.0, 11.6)

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*)

Exclusion 1 (patients who are discharged AMA) accounts for 0.44% of all index admissions excluded from the initial index cohort. This exclusion is needed for acceptability of the measure to hospitals, who do not have the opportunity to adequately deliver full care and prepare the patient for discharge. Because a very small percent of patients are excluded, this exclusion is unlikely to affect measure score.

Exclusion 2 (patients without at least 30 days of post-discharge enrollment in FFS Medicare for index admissions) accounts for 0.50% of all index admissions excluded from the initial cohort. This exclusion is needed because the 30-day readmission outcome cannot be assessed in this group since claims data are used to determine whether a patient was readmitted. Because a very small percent of patients are excluded, this exclusion is unlikely to affect measure score.

Exclusion 3 (patients with admission within 30 days of a prior index admission) accounts for 7.68% of all index admissions excluded from the initial index cohort. This exclusion is needed to prevent admissions from being counted as both an index admission and a readmission.

Exclusion 4 (patients with LVAD, history of LVAD, transplant, history of transplant) accounts for 0.18% of all index admissions excluded from the initial index cohort. This exclusion is needed to ensure a clinically coherent cohort. Patients undergoing implantation of an LVAD that is designed to offer intermediate to long-term support (weeks to years) as a "bridge" to heart transplant or "destination therapy" represent a clinically distinct, highly-selected group of patients cared for at highly specialized medical centers. Because a very small percent of patients are excluded, this exclusion is unlikely to affect measure score.

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 37 risk factors**

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

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

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient

characteristics (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 final set of risk-adjustment variables is:

Demographic

• Age-65 (years above 65, continuous) for 65 and over cohorts; or Age (years, continuous) for 18 and over cohorts
• Male

Cardiovascular

• History of Coronary Artery Bypass Graft (CABG) (ICD-9 V45.81, 36.10-36.16)
• Cardio-Respiratory Failure or Shock (CC 79)
• Congestive Heart Failure (CC 80)
• Acute Coronary Syndrome (CC 81-82)
• Coronary Atherosclerosis or Angina (CC 83-84)
• Valvular or Rheumatic Heart Disease (CC 86)
• Specified Arrhythmias (CC 92-93)
• Other or Unspecified Heart Disease (CC 94)

Comorbidity

• Vascular or Circulatory Disease (CC 104-106)
• Metastatic Cancer or Acute Leukemia (CC 7)
• Cancer (CC 8-12)
• Diabetes Mellitus (DM) or DM Complications (CC 15-20, 119-120)
• Protein-Calorie Malnutrition (CC 21)
• Disorders of Fluid/Electrolyte/Acid-Base (CC 22-23)
• Liver or Biliary Disease (CC 25-30)
• Peptic Ulcer, Hemorrhage, Other Specified Gastrointestinal Disorders (CC 34)
• Other Gastrointestinal Disorders (CC 36)
• Severe Hematological Disorders (CC 44)
• Iron Deficiency or Other Unspecified Anemias and Blood Disease (CC 47)
• Dementia or Other Specified Brain Disorders (CC 49-50)
• Drug/Alcohol Abuse/Dependence/Psychosis (CC 51-53)
• Major Psychiatric Disorders (CC 54-56)
• Depression (CC 58)
• Other Psychiatric Disorders (CC 60)
• Hemiplegia, Paraplegia, Paralysis, Functional Disability (CC 67-69, 100-102, 177-178)
• Stroke (CC 95-96)
• Chronic obstructive pulmonary disease (COPD) (CC 108)
• Fibrosis of Lung or Other Chronic Lung Disorders (CC 109)
• Asthma (CC 110)
• Pneumonia (CC 111-113)
• End-Stage Renal Disease or Dialysis (CC 129-130)
• Renal Failure (CC 131)
• Nephritis (CC 132)
• Other Urinary Tract Disorders (CC 136)
• Decubitus Ulcer or Chronic Skin Ulcer (CC 148-149)

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 HF Readmission

To examine the relationship between SES and race variables and hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure hospitalization, 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 heart failure readmission. Fifty studies were initially reviewed, and 36 studies were excluded from full-text review based on the above criteria. Studies indicated that SES/race variables were associated with 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; Aseltine et al., 2015; Calvillo-King et al., 2013; McHugh, Carthon, and Kang 2010; Damiani et al., 2015; Berenson and Shih 2012), though there may not be a significant effect on hospital-level profiling (Blum et al., 2014).

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 final variables in the model with associated odds ratios (OR).

Final Model Variables (variables meeting criteria in field 2b4.3) (**Dataset 1**)

Variable	07/2011-06/2014 OR (95% CI)
Age minus 65 (years above 65, continuous)	0.998 (0.997-0.999)
Male	0.996 (0.986-1.005)
History of Coronary Artery Bypass Graft (CABG) (ICD-9 V45.81, 36.10-36.16)	0.994 (0.983-1.005)
Cardio-Respiratory Failure or Shock (CC 79)	1.095 (1.083-1.106)
Congestive Heart Failure (CC 80)	1.134 (1.12-1.148)
Acute Coronary Syndrome (CC 81-82)	1.115 (1.103-1.128)
Coronary Atherosclerosis or Angina (CC 83-84)	1.056 (1.044-1.068)
Valvular or Rheumatic Heart Disease (CC 86)	1.054 (1.045-1.064)
Specified Arrhythmias (CC 92-93)	1.05 (1.039-1.061)
Other or Unspecified Heart Disease (CC 94)	1.042 (1.032-1.052)
Vascular or Circulatory Disease (CC 104-106)	1.073 (1.063-1.084)
Metastatic Cancer or Acute Leukemia (CC 7)	1.148 (1.116-1.182)
Cancer (CC 8-12)	1.021 (1.01-1.032)

Variable	07/2011-06/2014 OR (95% CI)
Diabetes Mellitus (DM) or DM Complications (CC 15-20, 119-120)	1.083 (1.072-1.093)
Protein-Calorie Malnutrition (CC 21)	1.079 (1.063-1.094)
Disorders of Fluid/Electrolyte/Acid-Base (CC 22-23)	1.111 (1.099-1.122)
Liver or Biliary Disease (CC 25-30)	1.069 (1.055-1.084)
Peptic Ulcer, Hemorrhage, Other Specified Gastrointestinal Disorders (CC 34)	1.066 (1.053-1.078)
Other Gastrointestinal Disorders (CC 36)	1.063 (1.052-1.074)
Severe Hematological Disorders (CC 44)	1.179 (1.151-1.208)
Iron Deficiency or Other Unspecified Anemias and Blood Disease (CC 47)	1.139 (1.127-1.151)
Dementia or Other Specified Brain Disorders (CC 49-50)	1.014 (1.004-1.025)
Drug/Alcohol Abuse/Dependence/Psychosis (CC 51-53)	1.096 (1.082-1.11)
Major Psychiatric Disorders (CC 54-56)	1.042 (1.028-1.057)
Depression (CC 58)	1.017 (1.006-1.029)
Other Psychiatric Disorders (CC 60)	1.055 (1.043-1.067)
Hemiplegia, Paraplegia, Paralysis, Functional Disability (CC 67-69, 100-102, 177-178)	1.042 (1.026-1.058)
Stroke (CC 95-96)	1.023 (1.008-1.039)
Chronic obstructive pulmonary disease (COPD) (CC 108)	1.154 (1.143-1.165)
Fibrosis of Lung or Other Chronic Lung Disorders (CC 109)	1.064 (1.05-1.079)
Asthma (CC 110)	1.014 (1-1.029)
Pneumonia (CC 111-113)	1.088 (1.077-1.098)
End-Stage Renal Disease or Dialysis (CC 129-130)	1.121 (1.1-1.144)
Renal Failure (CC 131)	1.182 (1.17-1.194)
Nephritis (CC 132)	1.089 (1.067-1.112)
Other Urinary Tract Disorders (CC 136)	1.067 (1.056-1.077)
Decubitus Ulcer or Chronic Skin Ulcer (CC 148-149)	1.104 (1.091-1.118)

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 heart failure cohort varies across measured entities. The median percentage of dual eligible patients is 13.8% (interquartile range [IQR] 8.7%- 22.0%). The median percentage of African-American patients is 3.3% (IQR 0.0%- 13.2%). The median percentage of patients with an AHRQ SES index score adjusted for cost of living at the census block group level equal to or below 42.7 is 21.2% (IQR 9.2%- 38.3%).

Empirical association with the outcome (univariate)

The patient-level observed heart failure readmission rate is higher for dual eligible patients, 25.5%, compared with 21.9% for all other patients. The readmission rate for African-American patients was also higher at 24.8% compared with 22.1% for patients of all other races. Similarly the readmission rate for patients with an AHRQ SES index score equal to or below 42.7 was 24.3% compared with 21.8% for patients with an AHRQ SES index score above 42.7.

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 multivariable model that includes all of the claims-based clinical variables, the effect size of each of these variables is small. 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 0.0094% (interquartile range [IQR] -0.0290% – 0.0386%, minimum -0.4499% – maximum 0.1559%) with a correlation coefficient between RSRRs for each hospital with and without dual eligibility added of 0.9993. The median absolute change in hospitals' RSRRs when adding a race indicator is 0.0197% (IQR -0.0284% – 0.0538%, minimum -0.7499% – maximum 0.1576%) with a correlation coefficient between RSRRs for each hospital with and without race added of 0.9987. The median absolute change in hospitals' RSRRs when adding an indicator for a low AHRQ SES index score adjusted for cost of living at the census block group level is 0.0377% (IQR -0.0502% – 0.1096%, minimum -0.9712% – maximum 0.2990%) with a correlation coefficient between RSRRs for each hospital with and without an indicator for a low AHRQ SES Index score adjusted for cost of living at the census block group level added of 0.9974.

Contextual Effect Analysis

As described in 2b4.3, we performed a decomposition analysis for each SES and race variable to assess whether there was a corresponding contextual effect. In order to better interpret the magnitude of results, we performed the same analysis for selected clinical risk factors. The results are described in the first table below (the decomposition table).

Both the patient-level and hospital-level dual eligible, race, and low AHRQ SES Index effects were significantly associated with heart failure readmission in the decomposition analysis. **That the hospital level effects were significant indicates that if the dual eligible, race, or low**

AHRQ SES Index variables are 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.

To assess the relative contributions of the patient- and hospital-level effects, we calculated a range of predicted probabilities of readmission for the SES or race variables and clinical covariates (comorbidities), as described in section 2b4.3. The results are presented in the figure and second table below (table of predicted probabilities for SES and race variables).

For SES and race variables, the hospital-level effect (P95-P5) is greater than the patient-level effect (delta) (second table below; the table of predicted probabilities for SES and race variables). For clinical variables, the patient-level effect (delta) is greater than the hospital-level effect (P95-P5) for renal failure and metastatic cancer and equal to the hospital-level effect for COPD (third table below; the table of predicted probabilities for clinical variables). This consistent pattern demonstrates that SES and race variables have a much greater hospital-level effect than patient-level effect. The clinical variables consistently had the opposite pattern, with a greater effect at the patient level than at the hospital level. Therefore, including SES and race variables into the model would predominantly adjust for a hospital-level effect, which is an important signal of hospital quality.

In the context of our conceptual model, we find clear evidence supporting the first two mechanisms by which SES might be related to poor outcomes. First we find that, although unadjusted rates of readmission are higher for patients of low SES or African-American race, the addition of SES to the readmission risk model, which already adjusts for clinical factors, makes very little difference. In particular, there is little-to-no change in model performance or hospital results with the addition of SES. This suggests that the model already largely accounts for the differences in clinical risk factors (degree of illness and comorbidities) among patients of varied SES.

Second, the predominance of the hospital-level effect of SES and race variables in the decomposition analyses suggests the risk associated with low SES is in large part due to lower quality of care at hospitals where more patients with these risk factors are treated; hospitals caring for socially- and economically-disadvantaged patients have higher readmission risk for **all** of their patients. Patients with low SES or African-American race indicators tend to receive care more frequently at lower quality hospitals compared with patients with high SES indicators. Direct adjustment for patient SES would essentially "over adjust" the measure, that is to say, it would be adjusting for an endogenous factor, one that influences the outcome through the site of treatment (hospital), as much as through an attribute of the patient.

In comparison, we did not observe the same predominance of the hospital-level effect among the clinical covariates, reinforcing the sense that SES and race factors have a distinct causal pathway in their impact on readmission risk.

Summary

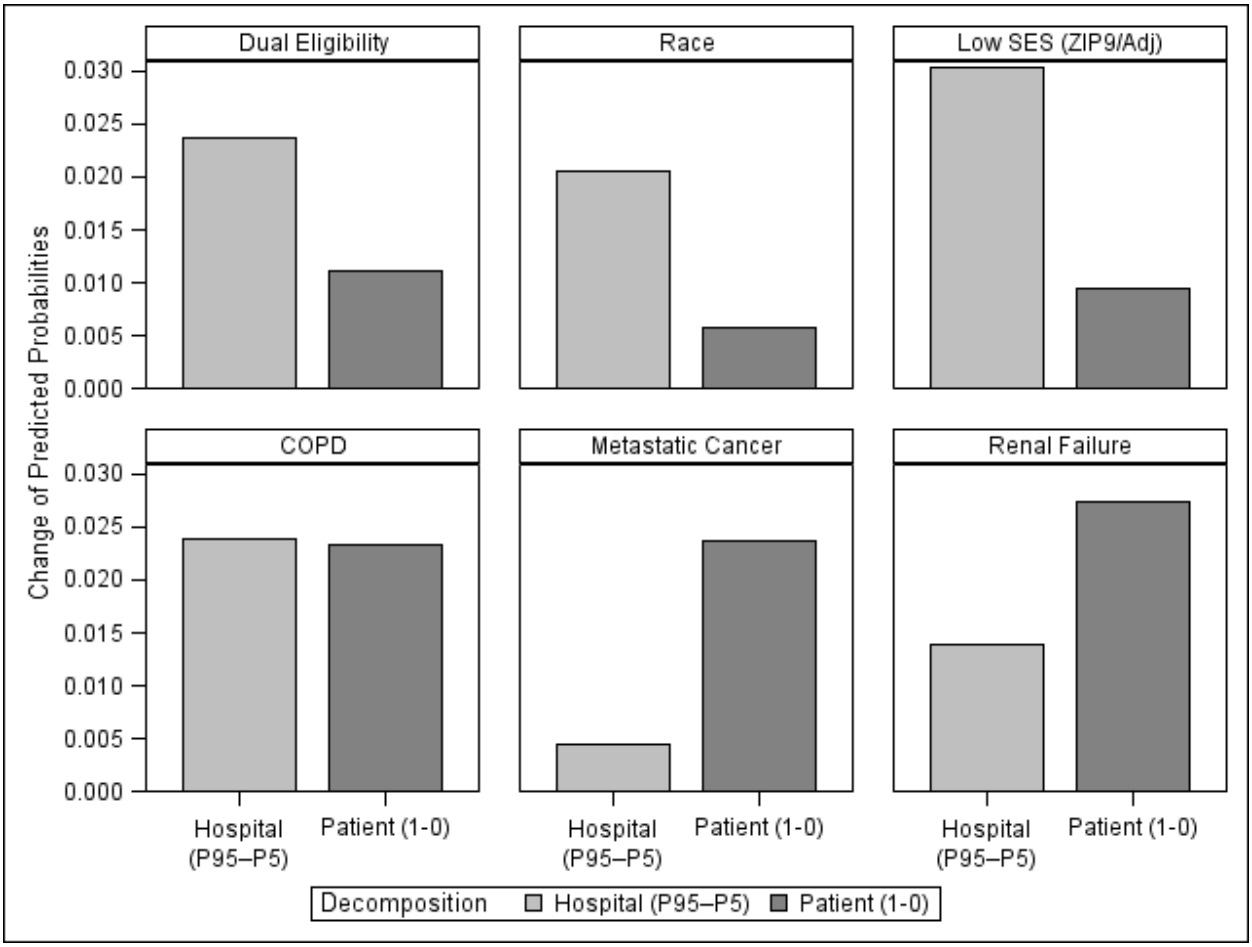
We found wide variation in the distribution of the three SES and race factors we examined, and we found that all three had some association with readmission risk. However, adjustment for these factors did not have an appreciable impact on hospital RSRRs, suggesting that existing clinical risk factors capture much of the risk related to low SES and African-American race. More importantly, we found that for all three factors there was a greater hospital-level effect,

compared with the patient-level effect, indicating that patient-level adjustment alone would adjust for quality differences between hospitals. Therefore, we did not include SES or race factors in our final risk model.

Heart Failure Readmission Decomposition Analysis

Parameter	Estimate (Standard Error)	P-value
Dual Eligible – Patient-Level	0.0645 (0.0064)	<0.0001
Dual Eligible – Hospital-Level	0.3955 (0.0356)	<0.0001
African American – Patient-Level	0.0336 (0.0077)	<0.0001
African American – Hospital-Level	0.2616 (0.0229)	<0.0001
Low SES census block group (AHRQ SES index, linked to 9-digit ZIP – Adjusted for Cost of Living) – Patient-Level	0.0550 (0.0056)	<0.0001
Low SES census block group (AHRQ SES index, linked to 9-digit ZIP – Adjusted for Cost of Living) – Hospital-Level	0.2464 (0.0186)	<0.0001
Renal Failure – Patient-Level	0.1640 (0.0054)	<0.0001
Renal Failure – Hospital-Level	0.2684 (0.0446)	<0.0001
Metastatic Cancer – Patient-Level	0.1356 (0.0147)	<0.0001
Metastatic Cancer – Hospital-Level	0.5435 (0.1822)	0.0029
COPD – Patient-Level	0.1379 (0.0049)	<0.0001
COPD – Hospital-Level	0.3999 (0.0357)	<0.0001

Change of Predicted Probabilities for SES and Race Compared with Clinical Variables in the HF Readmission Measure (July 2011-June 2014)



*Low SES (ZIP9/Adj) measured by linking patients' 9-digit ZIP codes to AHRQ SES Index at the census block group level, adjusted for cost of living

Predicted Probabilities for SES and Race Variables in the HF Readmission Measure (July 2011-June 2014)

Hospital SES/Race Risk Factor Percentile	Dual Eligibility				Race				Low SES census block group (AHRQ SES index, linked to 9-digit ZIP – Adjusted for Cost of Living)			
	Var_.J bar	Var_ij=0 for all patients	Var_ij=1 for all patients	Delta (Patient Effect)	Var_.J bar	Var_ij=0 for all patients	Var_ij=1 for all patients	Delta (Patient Effect)	Var_.J bar	Var_ij=0 for all patients	Var_ij=1 for all patients	Delta (Patient Effect)
0%	0.0000	0.2127	0.2234	0.0107	0.0000	0.2177	0.2233	0.0056	0.0000	0.2117	0.2208	0.0091
5%	0.0364	0.2150	0.2258	0.0108	0.0000	0.2177	0.2233	0.0056	0.0066	0.2120	0.2211	0.0091
10%	0.0530	0.2161	0.2269	0.0108	0.0000	0.2177	0.2233	0.0056	0.0272	0.2128	0.2219	0.0091
25%	0.0870	0.2183	0.2292	0.0109	0.0000	0.2177	0.2233	0.0056	0.0922	0.2154	0.2246	0.0092
50%	0.1377	0.2217	0.2327	0.0110	0.0328	0.2191	0.2248	0.0056	0.2116	0.2203	0.2296	0.0093
75%	0.2195	0.2272	0.2384	0.0112	0.1324	0.2235	0.2292	0.0057	0.3827	0.2274	0.2370	0.0095
90%	0.3137	0.2336	0.2450	0.0114	0.3098	0.2314	0.2373	0.0059	0.5934	0.2364	0.2462	0.0098
95%	0.3817	0.2383	0.2498	0.0116	0.4565	0.2381	0.2441	0.0060	0.7188	0.2419	0.2518	0.0099
100%	0.7333	0.2636	0.2760	0.0123	1.0000	0.2641	0.2704	0.0064	1.0000	0.2544	0.2647	0.0102
P95 – P5 (Hospital Effect)	-	0.0233	0.0240	-	-	0.0204	0.0208	-	-	0.0299	0.0307	-

Predicted Probabilities for Clinical Variables in the HF Readmission Measure (July 2011-June 2014)

Hospital SES/Race Risk Factor Percentile	Renal Failure				Metastatic Cancer				Chronic Obstructive Pulmonary Disease			
	Var_.J bar	Var_ij=0 for all patients	Var_ij=1 for all patients	Delta (Patient Effect)	Var_.J bar	Var_ij=0 for all patients	Var_ij=1 for all patients	Delta (Patient Effect)	Var_.J bar	Var_ij=0 for all patients	Var_ij=1 for all patients	Delta (Patient Effect)
0%	0.0909	0.1903	0.2163	0.0260	0.0000	0.2213	0.2448	0.0235	0.0584	0.1845	0.2056	0.0211
5%	0.3158	0.1996	0.2265	0.0269	0.0000	0.2213	0.2448	0.0235	0.3322	0.2011	0.2236	0.0225
10%	0.3592	0.2014	0.2285	0.0270	0.0000	0.2213	0.2448	0.0235	0.3667	0.2033	0.2260	0.0227
25%	0.4283	0.2044	0.2317	0.0273	0.0099	0.2222	0.2458	0.0236	0.4302	0.2073	0.2303	0.0230
50%	0.4968	0.2073	0.2349	0.0276	0.0191	0.2230	0.2467	0.0237	0.4988	0.2118	0.2351	0.0233
75%	0.5529	0.2097	0.2375	0.0278	0.0283	0.2238	0.2476	0.0237	0.5717	0.2166	0.2403	0.0237
90%	0.5976	0.2117	0.2397	0.0280	0.0385	0.2248	0.2486	0.0238	0.6412	0.2212	0.2452	0.0240
95%	0.6240	0.2128	0.2409	0.0281	0.0467	0.2255	0.2494	0.0238	0.6850	0.2242	0.2484	0.0242
100%	0.8696	0.2239	0.2529	0.0291	0.6129	0.2819	0.3092	0.0273	0.9123	0.2400	0.2653	0.0253
P95 – P5 (Hospital Effect)	-	0.0133	0.0145	-	-	0.0043	0.0046	-	-	0.0230	0.0248	-

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 (**Dataset 1, Dataset 2, and Dataset 3**)

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

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)

We tested the performance of the model for **Dataset 1** and **Dataset 3** described in section 1.7. During initial measure development, we tested the performance of the model developed in a randomly selected half of the hospitalizations for HF in 2004 compared with performance calculated from hospitalizations from the other half (**Dataset 3**). As a part of measure reevaluation, each year we assess temporal trends in model performance in the combined 3-year public reporting data (**Dataset 1**). Below, we report the model performance only for the 3-year combined results. For results for each individual year within the combined 3-year data please see the attached specifications report.

Reference:

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*)**:**

For the development cohort (**Dataset 3**) the results are summarized below:

First half of randomly split development sample: C statistic = 0.60; Predictive ability (lowest decile %, highest decile %) = (15, 37)
Second half of randomly split development sample: C statistic = 0.60; Predictive ability (lowest decile %, highest decile %) = (15, 37)

For the current measure cohort (**Dataset 1**) the results are summarized below:

C statistic = 0.60845; Predictive ability (lowest decile %, highest decile %) = (12.890, 35.418)

For comparison of model with and without inclusion of SDS factors, see Section 2b4.4b.

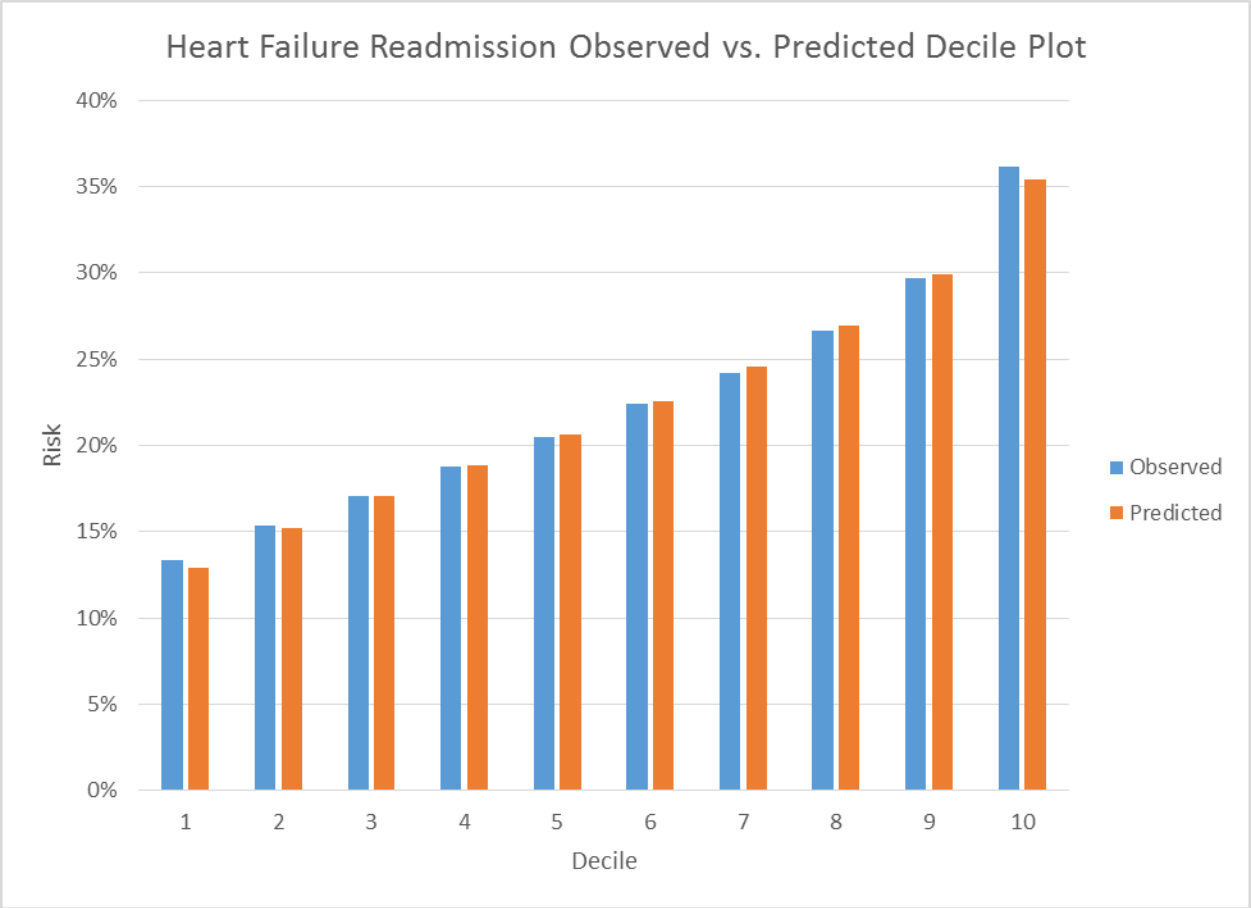
2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

For the original measure development cohort (**Dataset 3**) the results are summarized below:

First half of split sample: Calibration: (0, 1)
Second half of split sample: Calibration: (-0.02, 1.01)

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. Below, we present the risk decile plot showing the distributions for Medicare FFS data from July 2011 to June 2014 (**Dataset 1**).



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-statistics of 0.60, 0.60, and 0.60845 for the model development, validation, and current public reporting data (**Datasets 3, 3,** and **1** respectively) demonstrate consistent and fair model discrimination (**Datasets 1 and 3**). The models also 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 development and validation samples (**Dataset 3**) are substantially far from zero and the γ1 is substantially far from one, there is potential evidence of over-fitting. The calibration value of close to 0 at one end and close to 1 to 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*)

This measure is fully risk-adjusted using a hierarchical logistic regression model to calculate hospital RSRRs accounting for differences in hospital case-mix.

Application to Patients Aged 18 and Older (**Dataset 4**)

We applied the model to all-payer data from California. The analytic sample included 76,536 cases aged 18 and older 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.

To help determine whether the measure could be applied to an population of patients aged 18+, we examined the interaction terms between age (18-64 vs. 65+) and each of the other risk factors. Specifically, we fit the model in all patients 18+ with and without interaction terms and (a) conducted a reclassification analysis to compare risk prediction at the patient level; (b) compared the c-statistic; and (c) compared hospital-level risk-standardized rates (scatterplot, correlation coefficient, and R2) to assess whether the model with interactions is different from the current model in profiling hospital rates.

When the model was applied to all patients 18 and over (18+), overall discrimination was good (c-statistic=0.638). 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, (a) the reclassification analysis demonstrated good patient-level risk prediction (12.0% to 44.1% vs. 13.0% to 43.2%, respectively, from the bottom decile to the top decile of the prediction values); (b) the c-statistic was nearly identical (0.640 vs. 0.638); and (c) hospital-level risk-standardized rates were highly correlated (r=0.998; ICC=0.996). Thus, the inclusion of the 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 and older.

References:

Bernheim SM, Lin Z, Bhat KR, et al. 2010. 2010 Measures Maintenance Technical Report: Acute Myocardial Infarction, Heart Failure, and Pneumonia 30-Day Risk-Standardized Readmission Measures. Report prepared for the Centers for Medicare & Medicaid Services.

Harrell FE, Shih YCT. Using full probability models to compute probabilities of actual interest to decision makers. Int J Technol Assess Health Care. 2001;17:17–26.

Krumholz HM, Normand S-LT, Keenan PS, et al. 2008. Hospital 30-Day Heart Failure Readmission Measure: Methodology. Report prepared for the Centers for Medicare & Medicaid Services.

Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (YNHHSC/CORE) (January 2012). Hospital 30-Day, All-Cause, Risk-Standardized Readmission Rate (RSRR) following Pneumonia Hospitalization. In *Testing Publicly Report 30-Day Acute Myocardial Infarction, Heart Failure, and Pneumonia Risk-Standardized Mortality and Readmission Measures in California All-Payer Data.*

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*)

Analyses of Medicare FFS data show substantial variation in RSRRs among hospitals. Using data from July 2011 – June 2014 (**Dataset 1**), the median hospital RSRR was 22.34%, with a range of 15.98% to 32.08%. The interquartile range was 21.58%-23.22%.

Out of 4,778 hospitals in the U.S., 100 performed "better than the U.S. national rate," 3,766 performed "no different from the U.S. national rate," and 133 performed "worse than the U.S. national rate." 779 were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing.

Note that this analysis included index admissions from July 2011 – June 2014 from the 2015 public reported data (**Dataset 1**). We did not exclude patients who had a left ventricular assist device (LVAD) placed or a transplant in the index admission or the prior year and we used the planned readmission algorithm version 3.0 for measure calculation in these data. The new exclusion criterion and planned readmission algorithm 4.0 will first be applied in the 2016 publically reported measure results.

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 the quality of care received across hospitals for heart failure. This evidence supports continued measurement to reduce the variation.

Note: From the July 2011 to June 2012 reporting year to the July 2013 to June 2014 reporting year, the observed HF readmission rate decreased from 23.2% (July 2011 – June 2012) to 21.6% (July 2013 – June 2014). The observed readmission rate for the 3-year combined public reporting period (July 2011 – June 2014) for HF Medicare FFS patients is 22.4%.

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.**

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)
	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 HF. The median 30-day RSRR decreased by 1.6 absolute percentage points from July 2011-June 2012 (median RSRR: 23.1%) to July 2013-June 2014 (median RSRR: 21.5%). The median hospital RSRR from July 2011-June 2014 was 22.3% (IQR 21.6% - 23.2%).

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)**

0229 : Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following heart failure (HF) hospitalization for patients 18 and older

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 all-cause risk-standardized readmission rate (RSRR) following elective primary total hip arthroplasty (THA) and total knee arthroplasty (TKA)

1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

1891 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization

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?

No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**

We did not include in our list of related measures any non-outcome (e.g., process) measures with the same target population as our measure. Because this is an outcome measure, clinical coherence of the cohort takes precedence over alignment with related non-outcome measures. Furthermore, non-outcome measures are limited due to broader patient exclusions. This is because they typically only include a specific subset of patients who are eligible for that measure (for example, patients who receive a specific medication or undergo a specific procedure).

5b. Competing Measures

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

OR

Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**

Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

N/A

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific

submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.

Attachment **Attachment:** 2015_Measures_Reevaluation_Condition-Specific_Readmission_AUS_Report_FINAL_508_Compliant-635895835180308507.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services (CMS)
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.
The working group involved in the initial measure development is detailed in the original technical report available at www.qualitynet.org.

Our measure development team consisted of the following members:
Harlan Krumholz, M.D., S.M.
Sharon-Lise Normand, Ph.D.*
Patricia Keenan, Ph.D., M.H.S.
Zhenqiu Lin, Ph.D.
Elizabeth Drye, M.D., S.M.
Kanchana Bhat, M.P.H.
Yongfei Wang, M.Sc.
Joseph Ross, M.D., M.H.S.
Jeremiah Schuur, M.D.
Brett Stauffer, M.D. Susannah Bernheim, M.D., M.H.S. Andrew Epstein, Ph.D., M.P.P. Jeph Herrin, Ph.D.
Jessica Federer, B.S. Jennifer Mattera, M.P.H.
Yun Wang, Ph.D. Gregory Mulvey, B.A.
Geoffrey Schreiner, B.S.

*Harvard Medical School, Department of Health Care Policy

Technical Expert Panel Members:
Frederick Masoudi, MD, MSPH, FACC
Martha Radford, MD
John Rumsfeld, MD, PhD, FACC
John Spertus, MD, MPH, FACC
Frank Harrell, PhD

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2009
Ad.3 **Month and Year of most recent revision:** 09, 2012
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?** 01, 2017

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

Ad.8 **Additional Information/Comments:** N/A

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 #: 0506

De.2. **Measure Title:** Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization

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

De.3. **Brief Description of Measure:** The measure estimates a hospital-level 30-day, all-cause, risk-standardized readmission rate (RSRR) for patients discharged from the hospital with either a principal discharge diagnosis of pneumonia, including aspiration pneumonia or a principal discharge diagnosis of sepsis (not severe sepsis) with a secondary diagnosis of pneumonia (including aspiration pneumonia) coded as present on admission (POA). Readmission is defined as unplanned readmission for any cause within 30 days of the discharge date for the index admission. A specified set of planned readmissions do not count as readmissions. CMS annually reports the measure for patients who are 65 years or older and are enrolled in fee-for-service (FFS) Medicare hospitalized in non-federal hospitals.

Please note this measure has been substantially updated since the last submission; as described in S.3., the cohort has been expanded. Throughout this application we refer to this measure as version 8.2.

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 pneumonia. 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 their patient case mix, and therefore promote hospital quality improvement and better inform consumers about care quality.

Pneumonia readmission is a priority area for outcomes 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 readmission. We define readmission as an inpatient admission for any cause, with the exception of certain planned readmissions, within 30 days from the date of discharge from the index admission for patients 18 and older discharged from the hospital with a principal discharge diagnosis of pneumonia, including aspiration pneumonia or a principal discharge diagnosis of sepsis (not

severe sepsis) with a secondary discharge diagnosis of pneumonia (including aspiration pneumonia) coded as POA and no secondary discharge diagnosis of severe sepsis. If a patient has more than one unplanned admission (for any reason) within 30 days after discharge from the index admission, only the first one is counted as a readmission. The measure looks for a dichotomous yes or no outcome of whether each admitted patient has an unplanned readmission within 30 days. However, if the first readmission after discharge is considered planned, any subsequent unplanned readmission is not counted as an outcome for that index admission because the unplanned readmission could be related to care provided during the intervening planned readmission rather than during the index admission.

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

The cohort includes admissions for patients aged 18 years and older discharged from the hospital with principal discharge diagnosis of pneumonia, including aspiration pneumonia or a principal discharge diagnosis of sepsis (not severe sepsis) with a secondary discharge diagnosis of pneumonia (including aspiration pneumonia) coded as POA and no secondary discharge diagnosis of severe sepsis; and with a complete claims history for the 12 months prior to admission. The measure will be publicly reported by CMS for those patients 65 years and older who are Medicare FFS beneficiaries admitted to non-federal hospitals.

Additional details are provided in S.9 Denominator Details.

S.10. **Denominator Exclusions:** The readmission measures exclude index admissions for patients:

1. Discharged against medical advice (AMA);
2. Without at least 30 days post-discharge enrollment in FFS Medicare;
3. Admitted within 30 days of a prior index admission.

Measure Type: Outcome

S.23. **Data Source:** Administrative claims

S.26. **Level of Analysis:** Facility

IF Endorsement Maintenance – Original Endorsement Date: Oct 28, 2008 **Most Recent Endorsement Date:** Mar 06, 2013 (Pulmonary)

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 paired with a measure of hospital-level 30-day, all-cause, risk-standardized mortality (RSMR) following pneumonia hospitalization.

Maintenance of Endorsement -- 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. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence

Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

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.

Summary of evidence provided by the developer:

- The developer suggests that hospitals are able to influence readmission rates through a broad range of clinical activities including communication between providers, prevention of, and response to, complications, patient safety and coordinated transitions to the outpatient environment.
- The developer notes new studies that have demonstrated that appropriate (guideline recommended care), high-quality and timely treatment for pneumonia patients can reduce the risk of readmission within 30 days of hospital discharge (Leppin et al., 2014; Hansen et al., 2011). Additionally, the developer cites recent evidence of declining readmission rates as further support for the concept that care processes during and following hospitalization can affect a patient's risk of readmission. (Lee et al., 2014)
- New evidence is provided since the last endorsement maintenance review. Since its last review, this measure has been updated to include an expanded cohort to include patients with aspiration pneumonia and sepsis.

Question for the Committee:

- *Although the developer provides updated evidence related to aspects of hospitalization for pneumonia, does the Committee agree the underlying rationale for the measure remains reasonable and there is no need for repeat discussion and vote on evidence?*

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.

- The developer provides performance data from four measurement periods, covering a total of 1,469,277 admissions.
- The data show that during the measurement period of 07/2011–06/2014, pneumonia readmission rates ranged from a minimum of 13.1% to a maximum of 24.7%, with the 10th percentile at 16.0%, the 50th percentile at 17.5%, and the 90th percentile at 19.5%.

Disparities

- To help in assessment of potential disparities, the developers also provide performance scores (using 2011-2014 data) for hospitals serving a low proportion of dual eligible patients vs. those serving a high proportion of dual eligible patients, performance scores for hospitals serving a low proportion of African-American patients vs. those serving a high proportion of African-American patients, and performance scores for hospitals serving a low proportion of patients with AHRQ SES Index Score index score equal to or below 45.9 vs those serving a high proportion of patients with an AHRQ SES index score equal to or below 45.9.
- Hospitals serving a low proportion (=6.1%) Dual Eligible patients had a slightly lower median readmission rates (-0.5%) compared to hospitals serving a high proportion (=22.8%) Dual Eligible patients. Hospitals serving a low proportion (=0.0%) African-American patients had a lower median readmissions rates (-.8%) compared to hospitals serving a high proportion (=22.4%) African-American patients. Finally, hospitals serving a low proportion of patients below AHRQ SES index score of 42.7 had slightly lower median readmissions rates (-0.6%) compared to hospitals serving a high proportion of patients below AHRQ SRS index score of 42.7.

- By proportion of **Dual Eligible Patients**:

Characteristic//Hospitals with a low proportion (≤11.2%) Dual Eligible patients//Hospitals with a high proportion (≥25.2%) Dual Eligible patients
Number of Measured Hospitals// 1,088 // 1,085
Number of Patients// 394,137 patients in low-proportion hospitals/266,712 in high-proportion hospitals
Maximum// 24.3 // 24.1
90th percentile// 19.2 // 19.9
75th percentile// 18.2 // 18.8
Median (50th percentile)// 17.3 // 17.8
25th percentile// 16.5 // 17.0
10th percentile// 15.8 // 16.2
Minimum // 13.3 // 14.3

- By proportion of **African-American Patients**:

  Characteristic// Hospitals with a low proportion (≤0.0%) African-American patients//Hospitals with a high proportion (≥8.3%) African-American patients
  Number of Measured Hospitals// 1,275 // 1,085
  Number of Patients//157,004 patients in low-proportion hospitals/428,198 in high-proportion hospitals
  Maximum// 22.7 // 24.7
  90th percentile// 18.5 // 20.2
  75th percentile// 17.9 // 19.1
  Median (50%)// 17.2 // 18.0
  25th percentile// 16.6 // 17.1
  10th percentile// 15.9 // 16.4
  Minimum// 13.4 // 13.7

- By Proportion of Patients with AHRQ SES Index Scores Equal or Below 45.9:

  Characteristic// Hospitals with low proportion of patients with AHRQ SES index score equal to or below 42.7 (≤ 7.8%) // Hospitals with high proportion of patients with AHRQ SES index score equal to or below 42.7 (≥36.8%)
  Number of Measured Hospitals// 1,085 // 1,085
  Number of Patients// 308,131 patients in hospitals with low proportion of patients with AHRQ SES index score equal to or below 42.7 /268,306 patients in hospitals with high proportion of patients with AHRQ SES index score equal to or below 42.7
  Maximum// 22.9 // 24.1
  90th percentile// 18.8 // 20.0
  75th percentile// 18.0 //18.8
  Median (50th percentile)// 17.2 // 17.8
  25th percentile// 16.5 // 17.1
  10th percentile// 15.9 // 16.3
  Minimum // 13.4 // 14.3

Questions for the Committee:

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

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

Committee pre-evaluation comments

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

1a. Evidence to Support Measure Focus

Comments: **The developer's research has identified a facility level component in the analysis of 30-day readmission for pneumonia patients. The updated model includes sepsis as the primary diagnosis with pneumonia as a co-morbid condition and the modes was similar to the original model examined for patients with only pneumonia as a primary diagnosis. Studies from the literature have demonstrated processes in hospitals related to affecting patient capacity for self-care that can reduce the rate of readmission.

1b. Performance Gap

Comments: **There appears to be a fairly wide rang of readmission rates by facility to adequately demonstrate a performance gap.

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

Comments: **NA

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.

- This measure calculates 30-day all-cause readmission rates for patients discharged from an acute care hospital with a diagnosis of pneumonia. This measure was previously endorsed, however the cohort has been expanded to include patients with aspiration pneumonia and sepsis.
- This measure produces a risk-standardized readmission rate (RSRR), calculated as the ratio of the number of "predicted" to the number of "expected" readmission at a given hospital, multiplied by the national observed readmission rate.
- The denominator includes patients aged 18 years and older discharged from the hospital with a principal discharge diagnosis of pneumonia and with a complete claims history for the 12 months prior to admission. The measure can also be calculated for patients aged 65 and older only.
- The numerator includes patients who were readmitted for any cause, with the exception of certain planned readmissions, within 30 days of the date of discharge from the index pneumonia hospitalization.
- The denominator population is defined using ICD-9 and ICD-10 codes; a list of applicable codes is included in the submission.
- The numerator population includes readmissions to any acute care hospital for any cause within 30 days of the date of discharge of the index pneumonia admission, excluding planned readmissions as defined in section s.6.
- The data sources for this measure may include Medicare Part A and B claims, the Medicare Enrollment Database (EDB), and all-payer data sources such as the California Patient Discharge Database.
- The measure's time window can be specified from one to three years. The measure is currently reported with three years of index hospitalizations.
- The measure is risk-adjusted using a statistical risk model (see details below).

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.

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

- The developer has assessed reliability at both the data element and the performance score levels
- Datasets used for testing included Medicare Parts A and B claims as well as the Medicare Enrollment Database (EDB). Additionally, census as well as claims data were used to assess socio-demographic factors.

- **Data element reliability:**
  - With regard to data element reliability, the developer notes that the measure has been developed to avoid the use of claims data elements that are thought to be coded inconsistently across hospitals or providers, instead using fields that are consequential for payment and which are audited by CMS.
  - In addition, the developer compared frequencies and odds ratios of variables from their risk model across three years of data in order to assess the consistency of those variables over time.
- **Performance score reliability:**
  - The developer defines performance score reliability as the degree to which repeated measurements of the same entity agree with each other.
  - In line with this thinking, the developer's approach to assessing score-level reliability was to consider the extent to which assessments of a hospital using different but randomly-selected subsets of patients produce similar measures of hospital performance. The developers refer to this as a "test-retest" approach; it may also be called a "split-half" method. This is generally considered an appropriate method of testing reliability.

Results of reliability testing

- **Data element reliability:**
  - The frequency of some model variables increased and others decreased between 2011 and 2014, which may reflect increased or decreased co-morbidity rates.
  - The developer notes that examination of the odds ratios for each risk variable in the model shows that, overall, the odds ratios for individual risk variables remained relatively constant across three years.
- **Performance score reliability:**
  - A total of 1,469,277 admissions over a 3-year period were examined, with 733,434 in one sample and 735,843 in the other randomly-selected sample. Two risk-standardized mortality rates (RSMR) were calculated for each hospital: one from each of the two separate samples.
  - The agreement between the two RSMRs for each hospital (as measured by an intra-class correlation coefficient (ICC)) was 0.73; the developer states that according to the conventional interpretation, this is considered a "substantial" level of agreement.
  - The developer notes that this analysis was limited to hospitals with 12 or more cases in each split sample, and

that splitting the total population into two samples resulted in a sample equivalent of only 1.5 years of data, whereas the measure is reported with the full three years of data. [*Note: It is unclear whether the measure itself is limited to hospitals with 12 or more cases; if it is not, then testing was not conducted with the measure as specified.]*

- **Guidance from the Reliability Algorithm**

  - Question 1. Submitted specifications are precise, unambiguous, and complete. Measure can be consistently implemented.
  - Question 2. Empirical reliability testing was conducted using statistical tests with the measure as specified.
  - Question 3. Empirical validity testing of patient-level data was conducted.
  - Question 4. Reliability testing was conducted with computed performance measure scores for each measured entity.
  - Question 5. Random split-half correlation was used to assess the proportion of variability due to real differences among the measured entities.
  - Question 6. The ICC was 0.73 which is considered a substantial level of agreement.

Questions for the Committee:

- *Do the testing results presented by the developer demonstrate an adequate level of reliability?*
- *Is the reliability of the data elements sufficiently robust?*
- *In addition to the consistency of measurement results, assessments of performance score reliability often examine the ability of the measure to differentiate between measured entities. Do the reliability testing results reported by the developer demonstrate that meaningful differences in performance can be identified?*
- *Does the testing match the measure specifications?*

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.

- This measure estimates 30-day all-cause readmission rates for patients discharged from an acute care hospital with a diagnosis of pneumonia using a risk-standardized readmission rate (RSRR), which is calculated as the ratio of the number of "predicted" to the number of "expected" readmission at a given hospital, multiplied by the national observed readmission rate
- As a rationale for measuring this health outcome, the developers suggest that hospitals are able to influence readmission rates through a broad range of clinical activities, including prevention of complications, improving communication among providers involved at care transition, discharge planning, management of care transitions, medication reconciliation, patient education, and encouraging strategies that promote disease management.

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

Question for the Committee:

- *The cohort was expanded to capture a broader population of patients admitted for pneumonia and to capture a consistent clinical cohort across hospitals. Do you agree with this expansion?*

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.

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 tested the original version of the measure by comparing the administrative model with a medical-record based model. The results of this testing are included in the citation Krumholz, 2008. The developer notes that the claims-based measure produced results which were highly correlated with those produced through manual chart audit. (Krumholz et al., 2008; Lindenauer et al., 2011) While the developer provided two citations, these data were not provided in the measure submission form.
- The developer also provided justification for the the updated cohort by noting that the cohort expansion is based on changes in clinical and coding practices that have led to greater numbers of patients with pneumonia being coded with sepsis or aspiration pneumonia as a principal discharge diagnosis. Validity testing results to support the expansion were not provided in the measure submission form.

Finally, the measure developer provided spilt-halves testing of the measure which is often considered a test of measure reliability not measure validity. The measure developer provided spilt halves reliability testing results.

Validity testing results:

- The performance of the first half of the split sample and second half of the split sample was similar. The areas under the receiver operating characteristic (ROC) curve for the two models are 0.6419 and 0.6436, respectively. The developer notes ROC results were nearly identical and in line with other readmission models.
- Additional details can be found in the citation Krumholz, 2008.

Questions for the Committee:

- *Do you agree with the developer's approach to assessing validity?*
- *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:

- Patients in the following categories are excluded from the measure:
  - Discharged against medical advice (AMA);
  - Without at least 30 days post-discharge enrollment in FFS Medicare;
  - Admitted within 30 days of a prior index admission.
- To determine the impact of exclusions, the developer examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion.

- The number and percentage of patients excluded for each criterion are as follows:
    1. Discharged against medical advice (AMA): **11,621 (.55%)**
    2. Without at least 30 days post-discharge enrollment in FFS Medicare for index admissions: **169,803 (7.98%)**
    3. Pneumonia admission within 30 days of a prior pneumonia index admission: **64,916 (3.05%)**
- The developer also provides the distribution across hospitals for each exclusion criterion.

Questions for the Committee:

- *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: **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 measure employs a hierarchical logistic regression model (a form of hierarchical generalized linear model [HGLM]) to create a hospital-level 30-day risk-standardized readmission rate (RSRR).
- The developer suggests that 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.
- The developer notes that this approach simultaneously models data at the patient and hospital levels to account for the variance in patient outcomes both within and between hospitals.
- Variables considered for inclusion in the model 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 covariates are defined using condition categories (CCs), which are clinically-meaningful groupings of more than 15,000 ICD-9-CM diagnosis codes.
- The measure does not adjust for CCs that were possible adverse events of care and that were only recorded in the index admission.
- The final set of 41 risk-adjustment variables is included in the testing attachment; the odds ratio associated with each variable is also provided.
- The developers also considered a number of variables related to sociodemographic status (SDS) for potential inclusion in the risk-adjustment model. Candidate SDS variables were selected for examination based on a review of literature and national data sources.
- **Conceptual analysis of the need for SDS adjustment:**
    - The developers note there is a large body of literature linking various SES factors and African-American race to worse health status and higher readmission risk with income, education, and occupational level being the most commonly examined variables. The developers state that the literature directly examining how SES factors or race might influence the likelihood of older, insured, Medicare patient of being readmitted within 30 days of an admission for pneumonia is more limited.

- The developers state that few studies directly address causal pathways for SDS factors to affect 30-day readmission rates or examine the role of the hospital in these pathways.
- There are at least four potential pathways for SDS factors to affect 30-day readmission rates:
  - One potential pathway is the relationship to health status at the time of admission. SDS factors 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. The developers note that this pathway should be largely accounted for by their clinical risk-adjustment model.
  - The next potential path way is that patients with low income and African-American patient are more likely to be seen in lower quality hospitals, which can contribute to increased risk of readmission.
  - The third major pathway is that a patient's race or SDS status cause them to experience differential, lower quality care or may not receive the differentiated care they require.
  - Finally, some SES risk factors may affect the likelihood of readmission without directly affecting health status at admission or the quality of care received during the hospitalization. Patients may have worse outcomes due to competing economic priorities or a lack of access to care outside the hospital.

- **Empirical analysis of SDS factors:**
  - The developers considered African-American race, dual-eligible status-i.e. enrolled in both Medicare and Medicaid, and AHRQ-validated SES index score (summarizing the information from the following variables: percentage of people in the labor force who are unemployed, percentage of people living below poverty level, median household income, median value of owner-occupied dwellings, percentage of people ≥25 years of age with less than a 12th-grade education, percentage of people ≥25 years of age completing ≥4 years of college, and percentage of households that average ≥1 people per room)
  - The developers assessed the relationship between the SES variables and race with the outcome and examined the incremental effect in a multivariable mode.
  - The developer stated that they examined all patient-level indicators of both SES and race/ethnicity that are reliably available for all Medicare beneficiaries and linkable to claims data and selected those that are most valid.
  - The developer assessed the relationship between the SDS variables and the 30-pneumonia readmission rate and examined the incremental effect of SDS in a multivariable model, evaluating the extent to which the addition of any one of these variables improved model performance or changed hospital results.
  - The developer notes that 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. Therefore, the developers performed a decomposition analysis to assess the independent effects of the SES and race variables at the patient level and the hospital level.
  - The developers' analysis found that the prevalence of SDS factors in the pneumonia cohort does vary across measured entities.
  - With regard to the empirical association of each SDS variable with the outcome (univariate), the analysis found that patient-level observed pneumonia readmission rate for dual-eligible patients was higher, at 20.0% compared with 17.1% for all other patients. The readmission rate for African-American patients was also higher at 22.2% compared with 17.2% for patients of all other races. Similarly the readmission

rate for patients with an AHRQ SES index score equal to or below 42.7 was 19.3% compared with 17.1% for patients with an AHRQ SES index score above 42.7.

- With regard to the strength and significance of the SDS variables in the context of a multivariable model, the developers' analysis found that the effect size of each of these variables is small, the c-statistic (i.e., predictive value) is essentially unchanged with the addition of any of these variables into the model, and the addition of any of these variables into the model has little to no effect on hospital performance.
  - The median absolute change in hospitals' RSRRs when adding a dual eligibility indicator is 0.005% (interquartile range [IQR] -0.018% – 0.024%, minimum -0.267% – maximum 0.129%) with a correlation coefficient between RSRRs for each hospital with and without dual eligibility added of 0.99961.
  - The median absolute change in hospitals' RSRRs when adding a race indicator is 0.035% (IQR -0.038% – 0.086%, minimum -1.337% – maximum 0.226%) with a correlation coefficient between RSRRs for each hospital with and without race added of 0.99608.
  - The median absolute change in hospitals' RSRRs when adding a low AHRQ SES Index score indicator to the model is 0.0342% (IQR -0.0254% – 0.0806%, minimum -0.5159% – maximum 0.2296%) with a correlation coefficient between RSRRs for each hospital with and without an indicator for a low AHRQ SES Index score adjusted for cost of living at the census block group level is 0.9981.
- The developers state that the patient-level and hospital-level dual eligible, race, and low AHRQ SES Index effects were significantly associated with pneumonia readmission in the decomposition analysis. The developers note that if the dual eligible, race, or low AHRQ SES Index variables are 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.
- To assess the relative contributions of the patient- and hospital-level effects, the developers calculated a range of predicted probabilities of readmission for the SES or race variables and clinical covariates (comorbidities).
- For SES variables, the hospital-level effect is greater than the patient-level effect (delta). For the race variable, the patient-level effect (delta) is greater than the hospital-level effect. For clinical variables, the patient-level effect (delta) is greater than the hospital-level effect (P95-P5) for lung cancer and COPD. The hospital-level effect (P95-P5) is greater than the patient-level effect (delta) for renal . The developers note there is a consistent pattern demonstrating that SES variables have a much greater hospital-level effect than patient-level effect. Notably, the race variable had a slightly greater patient-level effect. The clinical variables had the opposite pattern, with a greater effect at the patient level than at the hospital level for lung cancer and COPD. However, renal failure had a similar hospital-level and patient-level effect. In sum, the developers feel including SES variables into the model would predominantly adjust for a hospital-level effect, which is an important signal of hospital quality.
- The developers state that given these findings and complex pathways that could explain any relationship between SDS and mortality, which do not all support risk-adjustment, **they did not incorporate SDS variables into the measure.**

- **Risk Model Diagnostics:**
  - To assess the overall performance of their risk-adjustment model, the developers computed three summary statistics, including:

- *Area under the receiver operating characteristic (ROC) curve* (also known as a c-statistic, which measures the probability that the model's prediction of the outcome is better than chance)
- *Predictive ability* (the model's ability to distinguish high-risk subjects from low-risk subjects)
- *Over-fitting indices (model calibration)* (to ensure that the model is not only describing the relationship between predictive variables and outcome in the development dataset but also providing valid predictions in new patients)

- For the current measure cohort, the findings from this analysis are as follows:
  - C-statistic: **0.63**
    - A c-statistic of 0.63 means that for 63% of all possible pairs of patients—one who was readmitted and one who was not—the model correctly assigned a higher probability to those who were readmitted. Generally, a c-statistic of >0.70 is considered acceptable.
    - The developers interpret 0.63 for the c-statistic as 'fair' model discrimination.
  - Predictive ability (lowest decile %, highest decile %): (**9.3%, 32.7%**)
    - The developers state that there is a wide range between the lowest decile and highest decile, indicating the ability to distinguish high-risk subjects from low-risk subjects.
  - Overfitting indices (model calibration) [presented as (γ0, γ1)]:
    - The developer states that if the γ0 in the validation samples are substantially far from zero and the γ1 is substantially far from one, there is potential evidence of over-fitting. The calibration value of close to 0 at one end and close to 1 to the other end indicates good calibration of the model.
      - 1st half of split sample: Calibration: (**0.0230, 0.9911**)
      - 2nd half of split sample: Calibration: (**0.0231, 0.9900**)
- The developer's overall interpretation of the results of their analysis is that the findings demonstrate the risk-adjustment model adequately controls for differences in patient characteristics (case mix).
- The developer also conducted additional analyses to determine whether the measure could be applied to a population of patients aged 18+ using all-payer data.
- The developers report that their results indicate their model had good discrimination and predictive ability in this group.

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)*:

- For public reporting of this measure, CMS characterizes the uncertainty associated with the RSRR by estimating the 95% interval estimate.
- If the RSRR's interval estimate does not include the national observed readmission rate (because it 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."
- The developer reports that for the performance period of July 2011-June 2014, the mean hospital RSMR was

17.5%, with a range of 13.1% to 24.7%. The interquartile range was 16.7%-18.4%.

- Of 4,700 hospitals in the study cohort, 86 performed "better than the U.S. national rate," 4,061 performed "no different from the U.S. national rate," 193 performed "worse than the U.S. national rate," and 360 were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing.
- The developer's interpretation of this data is that the variation in rates and number of performance outliers suggests there remain differences in the quality of care received across hospitals for pneumonia that support measurement to reduce the variation.

Question for the Committee:

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

2b6. Comparability of data sources/methods:

- While the developer did not decide to include SDS variables in their final model, they did compare measure results with and without SDS adjustment.

2b7. Missing Data

- N/A

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

Comments: **Adding the new diagnoses does not appear to alter the relationship between the evidence and the model specifications.

2a2. Reliability Testing

Comments: **The sample size was adequate for the reliability testing and the results supported the measure reliability over time (ICC=0.73).

2b2. Validity Testing

Comments: **Initial testing supported validity using two different types of data: administrative data and medical record data. Separate tests were not done for the updated model using additional dx codes for pneumonia patients.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **Missing data and data excluded does not have a substantial impact on the results of the measure. Standard risk adjustment using claims data was used in the model. Although SDS variables have a significant relationship at the patient level with the readmission measure, their inclusion in the model did not change the results or the strength of the model for the facility effect. In addition, a new risk adjustment variable of respiratory dependence within 12 months prior to the index admission was found to have a strong association with readmission.

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 based on administrative claims data (e.g., DRG, ICD-9/10), which the developers note are routinely generated and collected as part of hospitals' billing processes.
- The developer indicates that all data elements are in defined fields in electronic claims.

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **The data for this measure has been readily available and extensively studied.

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**

OR

Planned use in an accountability program? ☐ **Yes** ☐ **No**

Accountability program details

- Hospital Inpatient Quality Reporting (IQR) Program http://cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html and Hospital Readmission Reduction (HRRP) Program http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program.html

Improvement results

- The developer reports: "median hospital 30-day, all-cause, RSRR for the re-specified pneumonia readmission measure with the expanded cohort (version 8.2) for the 3-year period between July 2011 and June 2014 was 17.5% (IQR 16.7% - 18.3%). The median RSRR decreased by 1.0 absolute percentage points from July 2011-June 2012 (median RSRR: 18.1%) to July 2013-June 2014 (median RSRR: 17.1%)."

Unexpected findings (positive or negative) during implementation

- The developer noted that there are no unexpected findings to report.

Potential harms

- The developer noted that there were no unintended consequences during development, testing or re-specification. They are committed to ongoing monitoring of potential unintended consequences over time.

Feedback :

- During the 2012-2013 MAP review, MAP supported this measure for inclusion in the IQR and HRRP programs. The group agreed that the new specifications are an improvement over the existing finalized measure.

***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

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **There is a potential for misinterpreting the risk-standardized readmission rate because the measure is expressed as a fraction of the national readmission rate; i.e., the ratio of predicted to expected readmissions for a hospital is multiplied by the national observed readmission rate to give the RSRR.

Criterion 5: Related and Competing Measures

Related or competing measures

- 0279: Bacterial Pneumonia Admission Rate (PQI 11)
- 2882: Excess days in acute care (EDAC) after hospitalization for pneumonia

Harmonization

- The developer notes that the measures are not completely harmonized. The developer justifies the difference by noting that for outcome measures clinical coherence of the cohort takes precedence over alignment with related non-outcome measures.

Pre-meeting public and member comments

Comment by: Ms. Elizabeth Godsey
Organization: Vizient, Inc.
Comment May 05, 2016: Vizient, Inc., the largest member-owned health care company in the country, is dedicated to serving members & customers through innovative data-driven solutions, expertise & collaborative opportunities that lead to improved patient outcomes & lower costs. Vizient requests CMS to review & provide follow-up analysis on more applied/practical alternate modeling approaches to account for within & across hospital variation besides hierarchical modeling. While hierarchical modeling is a valid technique controlling for within & across hospital variation, the approach lacks a tangible, practical framework of an observed to expected ratio that hospitals need to drive patient care. The

predicted to expected approach complicates the public's & provider's understanding of how the actual observed values impacts hospital performance. Through numerous member discussions, we heard repeatedly, Oh, you mean that number does really reflect my actual readmissions? How can I improve that number? Even more concerning is the focus the current measure places on improving documentation & coding rather than patient care. Currently, providers see the only direct way to improve the measure is through documentation & coding capture of co-morbidities which count toward the predicted & expected value calculations. We hope this was not the original intention of the measure & this misguided focus is simply an unintended artifact of an overly complicated modeling technique. We recommend analyzing & provide results comparing a model that uses hospital characteristics, such as teaching status or bed size to account for structural differences across hospitals & provide an observed to expected ratio which is much more meaningful for the public & providers. While in the past, CMS has commented they would not incorporate these features due to NQF restrictions; it is important to point out NQF has endorsed other risk adjustment models that incorporate these characteristics (NHSN) & consider these factors in the 30-day risk adjustment as well. Also, we would ask CMS & NQF to institute discrimination performance thresholds for the models given the importance these models bare on CMS's performance programs & public reporting. Currently, no model performs > 0.70, a standard considered fair-good practical performance threshold & while the c-stat does not fully evaluate the model, it certainly should require basic performance standards. Additionally, we ask CMS to provide performance statistics, like AIC, BIC & the Somers' D, Gamma & Tau-a association of predicted probabilities & observed counts for a more comprehensive assessment. Using these standards & model diagnostics, NQF can provide CMS with recommendations for improvement. Until minimum discrimination thresholds are instituted, we recommend NQF remove endorsement of the readmission measures.

Comment by: Ms. Elizabeth Godsey
Organization: Vizient, Inc.
Comment May 05, 2016: Vizient, Inc., the largest member-owned health care company in the country, is dedicated to serving members & customers through innovative data-driven solutions, expertise & collaborative opportunities that lead to improved patient outcomes & lower costs. For the readmission measures considered, CMS presented patient-level & hospital specific SES factor beta coefficients & p-values, yet overall model performance were not presented. We request the actual model performance results for model evaluation. For the AHRQ SES Index variable, we request further information on how the binary classification for a measure that ranges between 0-100 was determined & the impact of transforming into a binary representation vs. actual value had on the model performance. This detail along with the overall model performance information would provide the public with the necessary information to truly assess CMS's comment 'Given these findings & 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.' Regarding the complex pathways associated with 30-day readmissions as stated by CMS, we strongly ask CMS to entirely re-evaluate the utility of the 30-day measures. As stated by CMS, factors influencing readmissions are blurred between providers & patients 30-days post discharge resulting in a limited insights in how providers can improve care. We believe CMS's efforts to remove the planned readmissions PR4 logic is a strong step in true opportunity identification; however, more refinement is needed. We recommend a shorter, more actionable 7 day post-discharge readmission timeframe to pinpoint opportunities providers truly can influence & thus, mitigate many of SES confounding factors. The 7-day window provides clearer opportunities for patient stabilization & post-acute discharge planning which the 30-day window doesn't reflect. We recommend CMS provide a 7-day readmission risk adjustment for review. Also, the hospital wide readmission measure evaluates all readmissions within the 30-day window post inpatient discharge & considers readmit cases to also be eligible as the index admission; however, the condition specific measures evaluate only 1 readmit within

the 30-day window & cannot be eligible as an index. We ask CMS for the rationale why the different approaches for the same measure as this adds unnecessary complexity which are impractical to manage. We recommend a consistent approach across all readmission measure calculations & recommend evaluating & counting all readmits that occur within the 30-day window so providers have a clear understanding of the # readmits are truly occurring. We support considering a readmit as an index for the next 30-day cycle to again, assist organizations in tracking & improving complete patient care.

Comment by: Ms. Elizabeth Godsey
Organization: Vizient, Inc.
Comment May 05, 2016: Vizient, Inc., the largest member-owned health care company in the country, is dedicated to serving members & customers through innovative data-driven solutions, expertise & collaborative opportunities that lead to improved patient outcomes & lower costs. Vizient agrees with CMS's additions to the denominator cohort definition to include aspiration pneumonia & sepsis w secondary dx of pneumonia and recommends CMS and NQF add the following I-10 translation codes per the 2015 GEMS, J1000, J1001, J1008, J1108 which are the I-10 equivalent to the existing ICD-9 code 4870 which exists in the measure definition. Upon review, Vizient noticed no ICD-10 translation was provided for severe sepsis ICD-9 codes 995.92 or 785.52. Vizient recommends including ICD-10 codes R6520 and R6521 as per the GEMS 2015 mapping. Within the PN readmission specifications, the planned readmission exclusion algorithm references V3.0 yet, V4.0 is currently proposed. Vizient recommends CMS provide consistent V4.0 planned readmission algorithm for all the readmission measures. In reviewing the algorithm for AHRQ CCS potentially planned procedure list, AHRQ CCS 169 is listed as exclusion criteria, but within ICD-10 CCS 169 does not exist. Vizient recommends CMS and NQF reviewing this criterion and provide the appropriate ICD-10 translations to address the debridement of wound; infection or burn procedure codes.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0506

Measure Title: Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: N/A

Date of Submission: 1/29/2016

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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: Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

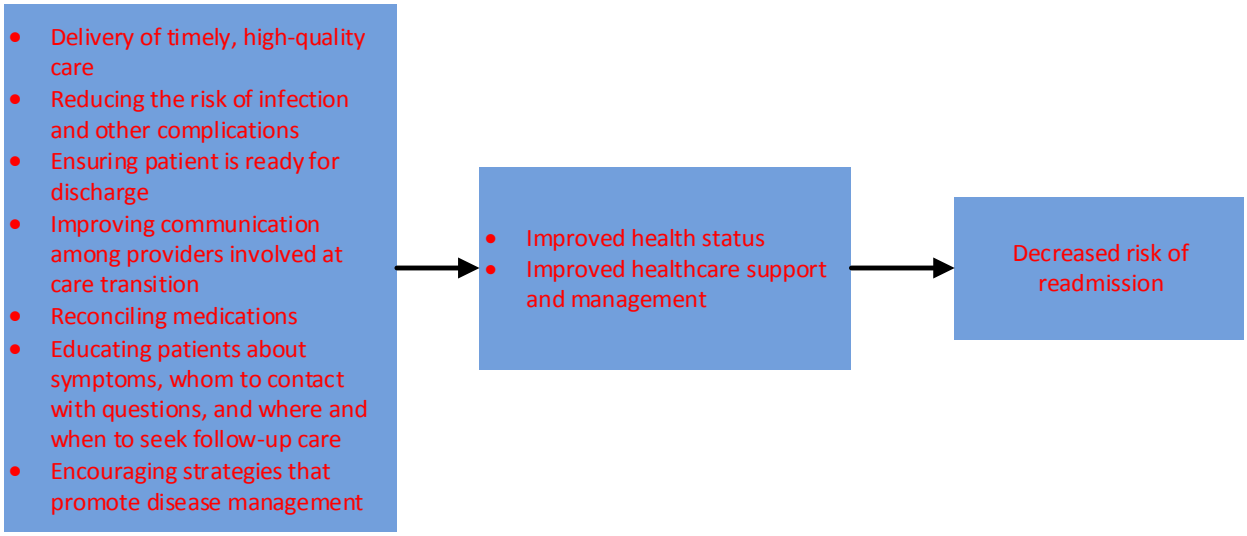
☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.



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 hospitalization for pneumonia. 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.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

In 2007, the Medicare Payment Advisory Commission (MedPAC) called for hospital-specific public reporting of readmission rates, and identified pneumonia as a priority condition (MedPAC, 2007). In 2010, pneumonia is the principal discharge diagnosis for more than 1 million hospitalizations each year in the United States (Lindenauer et al., 2012; FastStats: pneumonia, CDC). From 2003 to 2004, approximately 20% of pneumonia patients were rehospitalized within thirty days, representing the second-highest proportion of all rehospitalizations at 6.3% (Jencks et al., 2009). Among patients 65 years [of age] or older in the United States, pneumonia is the second leading cause of hospitalization (Fry et al., 2005), and based on 2005 Medicare data, MedPAC estimated that about 8.9% of Medicare pneumonia admissions were followed by a readmission within 15 days, accounting for more than 74,000 admissions at a cost of $533 million.

Pneumonia readmission is a costly event and represents an undesirable outcome of care from the patient's perspective, and highly disparate pneumonia readmission rates among hospitals suggest there is room for improvement (MedPAC, 2007; Lindenauer et al., 2010). Although many current hospital interventions are known to decrease the risk of readmission within 30 days of hospital discharge (Leppin et al., 2014), current process-based performance measures, cannot capture all the ways that care within the hospital might influence outcomes. Measurement of patient outcomes allows for a comprehensive view of quality of care that reflects complex aspects of care such as communication between providers and coordinated transitions to the outpatient environment. These aspects are critical to patient outcomes, and are more broad than what can be captured by individual process-of-care measures.

The pneumonia hospital-specific risk-standardized readmission rate (RSRR) measure is thus intended to inform quality-of-care improvement efforts, as individual process-based performance measures cannot encompass all the complex and critical aspects of care within a hospital that contribute to patient outcomes. As a result, many stakeholders, including patient organizations, are interested in outcomes measures that allow patients and providers to assess relative outcomes performance for hospitals (Bratzler et al., 2007).

The diagram above indicates some of the many care processes that can influence readmission risk by improving health status or improving healthcare management and support. Numerous studies have demonstrated that appropriate (guideline recommended care), high-quality and timely treatment for pneumonia patients can reduce the risk of readmission within 30 days of hospital discharge (Leppin et al., 2014; Hansen et al., 2011). Recent evidence of declining readmission rates provides further support for the concept that care processes during and following hospitalization can affect a patient's risk of readmission (Lee et al., 2014).

References:

Bratzler DW, Nsa W, Houck PM. Performance measures for pneumonia: are they valuable, and are process measures adequate. Current Opinion in Infectious Diseases. 2007; 20(2):182-189.

Centers for Disease Control and Prevention. FastStats: pneumonia. Available at: http://www.cdc.gov/nchs/fastats/pneumonia.htm. Accessed August 13, 2015.

Fry AM, Shay DK, Holman RC, et al. Trends in hospitalizations for pneumonia among persons aged 65 years or older in the United States, 1988–2002. JAMA. 2005; 294:2712–2719.

Hansen LO, Young RS, Hinami K, et al. Interventions to reduce 30-day rehospitalization: a systematic review. 2011; 155(8):520-8.

Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. N Engl J Med. 2009; 360(14):1418-28.

Leppin AL, Gionfriddo MR, Kessler M, et al. Preventing 30-day hospital readmissions: a systematic review and meta-analysis of randomized trials. JAMA Internal Med. 2014; 174(7):1095-107.

Lee JS, Nsa W, Hausmann LRM, et al. Quality of care for elderly patients hospitalized for pneumonia in the United States, 2006 to 2010. JAMA Intern Med. 2014; 174(11):1806-1814.

Lindenauer PK, Bernheim SM, Grady JN, et al. The performance of US hospitals as reflected in risk-standardized 30-day mortality and readmission rates for Medicare beneficiaries with pneumonia. 2010; 5(6):E12-8.

Lindenauer PK, Lagu T, Shieh MS, et al. Association of diagnostic coding with trends in hospitalizations and mortality of patients with pneumonia, 2003-2009. JAMA American Medical Association. 2012; 307(13):1405-1413.

Medicare Payment Advisory Commission. Report to the Congress: Promoting Greater Efficiency in Medicare. 2007.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

N/A. This measure is not an intermediate outcome, process, or structure performance measure.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

N/A. This measure is not an intermediate outcome, process, or structure performance measure.

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

N/A

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

N/A

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

N/A

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

N/A

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

N/A

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

N/A

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

N/A

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

N/A

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

N/A

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

N/A

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

N/A

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

N/A

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

N/A

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: Click here to enter date range

N/A

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

N/A

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

N/A

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

N/A

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

N/A

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

N/A

1a.8 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.8.1 What process was used to identify the evidence?

N/A

1a.8.2. Provide the citation and summary for each piece of evidence.

N/A

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

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all 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_0506_PN_Readmission_NQF_Evidence_Attachment_01-29-16_v1.1.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 pneumonia. 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 their patient case mix, and therefore promote hospital quality improvement and better inform consumers about care quality.

Pneumonia readmission is a priority area for outcomes 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.*

Distribution of Hospital Pneumonia RSRRs over Different Time Periods
Results for each data year
Characteristic//07/2011-06/2012//07/2012-06/2013//07/2013-06/2014//07-2011-06/2014
Number of Hospitals// 4,623 // 4,608 // 4,566 // 4,700
Number of Admissions// 493,792 // 514,408 // 461,077 // 1,469,277
Mean (SD)// 18.2 (1.0) // 17.4 (1.1) // 17.2 (0.9) // 17.6 (1.4)
Range (min. – max.)// 14.7-23.7 // 13.7-23.7 // 12.7-22.7 // 13.1-24.7
Minimum// 14.7 // 13.7 // 12.7 // 13.1
10th percentile// 17.1 // 16.2 // 16.1 // 16.0
20th percentile// 17.5 // 16.6 // 16.5 // 16.5

30th percentile// 17.7 // 16.9 // 16.7 // 16.9
40th percentile// 18.0 // 17.1 // 16.9 // 17.2
50th percentile// 18.1 // 17.3 // 17.1 // 17.5
60th percentile// 18.3 // 17.5 // 17.3 // 17.8
70th percentile// 18.5 // 17.7 // 17.5 // 18.1
80th percentile// 18.8 // 18.1 // 17.8 // 18.6
90th percentile// 19.4 // 18.8 // 18.4 // 19.5
Maximum// 23.7 // 23.7 // 22.7 // 24.7

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

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for 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.*
Distribution of Pneumonia RSRRs by Proportion of Dual Eligible Patients:
Dates of Data: July 2011 through June 2014
Data Source: Medicare FFS claims

Characteristic//Hospitals with a low proportion (≤11.2%) Dual Eligible patients//Hospitals with a high proportion (≥ 25.2%) Dual Eligible patients
Number of Measured Hospitals// 1,088 // 1,085
Number of Patients// 394,137 patients in low-proportion hospitals/266,712 in high-proportion hospitals
Maximum// 24.3 // 24.1
90th percentile// 19.2 // 19.9
75th percentile// 18.2 // 18.8
Median (50th percentile)// 17.3 // 17.8
25th percentile// 16.5 // 17.0
10th percentile// 15.8 // 16.2
Minimum // 13.3 // 14.3

Distribution of RSRRs by Proportion of African-American Patients:
Dates of Data: July 2011 through June 2014
Data Source: Medicare FFS claims

Characteristic// Hospitals with a low proportion (≤0.0%) African-American patients//Hospitals with a high proportion (≥8.3%) African-American patients
Number of Measured Hospitals// 1,275 // 1,085
Number of Patients//157,004 patients in low-proportion hospitals/428,198 in high-proportion hospitals
Maximum// 22.7 // 24.7
90th percentile// 18.5 // 20.2
75th percentile// 17.9 // 19.1
Median (50%)// 17.2 // 18.0
25th percentile// 16.6 // 17.1
10th percentile// 15.9 // 16.4
Minimum// 13.4 // 13.7

Distribution of Pneumonia RSRRs by Proportion of Patients with AHRQ SES Index Scores Equal to or Below 42.7:
Dates of Data: July 2011 through June 2014
Data Source: Medicare FFS claims and The American Community Survey (2008-2012) data

Characteristic// Hospitals with low proportion of patients with AHRQ SES index score equal to or below 42.7 (≤ 7.8%) // Hospitals with high proportion of patients with AHRQ SES index score equal to or below 42.7 (≥36.8%)
Number of Measured Hospitals// 1,085 // 1,085
Number of Patients// 308,131 patients in hospitals with low proportion of patients with AHRQ SES index score equal to or below 42.7 /268,306 patients in hospitals with high proportion of patients with AHRQ SES index score equal to or below 42.7
Maximum// 22.9 // 24.1
90th percentile// 18.8 // 20.0
75th percentile// 18.0 //18.8
Median (50th percentile)// 17.2 // 17.8
25th percentile// 16.5 // 17.1
10th percentile// 15.9 // 16.3
Minimum // 13.4 // 14.3

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, 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.**
In 2007, the Medicare Payment Advisory Commission (MedPAC) called for hospital-specific public reporting of readmission rates, and identified pneumonia as a priority condition (MedPAC, 2007). In 2010, pneumonia is the principal discharge diagnosis for more than 1 million hospitalizations each year in the United States (Lindenauer et al., 2012; FastStats: pneumonia, CDC). From 2003 to 2004, approximately 20% of pneumonia patients were rehospitalized within thirty days, representing the second-highest proportion of all rehospitalizations at 6.3% (Jencks et al., 2009). Among patients over 65 years [of age] or older in the United States, pneumonia is the second leading cause of hospitalization (Fry et al., 2005), and based on 2005 Medicare data, MedPAC estimated that about 8.9% of Medicare pneumonia admissions were followed by a readmission within 15 days, accounting for more than 74,000 admissions at a cost of $533 million (MedPAC, 2007).

Pneumonia readmission is a costly event and represents an undesirable outcome of care from the patient's perspective, and highly disparate pneumonia readmission rates among hospitals suggest there is room for improvement (MedPAC, 2007; Lindenauer et al., 2010). Although many current hospital interventions have been shown to decrease the risk of readmission within 30 days of hospital discharge (Leppin et al., 2014), current process-based performance measures, cannot capture all the ways that care within the hospital might influence outcomes. As a result, many stakeholders, including patient organizations, are interested in outcomes measures that allow patients and providers to assess relative outcomes performance for hospitals (Bratzler et al., 2007).

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

Bratzler DW, Nsa W, Houck PM. Performance measures for pneumonia: are they valuable, and are process measures adequate. Current Opinion in Infectious Diseases. 2007; 20(2):182-189.

Centers for Disease Control and Prevention. FastStats: pneumonia. Available at: http://www.cdc.gov/nchs/fastats/pneumonia.htm. Accessed August 13, 2015.

Fry AM, Shay DK, Holman RC, et al. Trends in hospitalizations for pneumonia among persons aged 65 years or older in the United States, 1988–2002. JAMA. 2005; 294:2712–2719.

Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. N Engl J Med. 2009; 360(14):1418-28.

Leppin AL, Gionfriddo MR, Kessler M, et al. Preventing 30-day hospital readmissions: a systematic review and meta-analysis of randomized trials. JAMA Internal Med. 2014; 174(7):1095-107.

Lindenauer PK, Bernheim SM, Grady JN, et al. The performance of US hospitals as reflected in risk-standardized 30-day mortality and readmission rates for Medicare beneficiaries with pneumonia. 2010; 5(6):E12-8.

Lindenauer PK, Lagu T, Shieh MS, et al. Association of diagnostic coding with trends in hospitalizations and mortality of patients with pneumonia, 2003-2009. JAMA American Medical Association. 2012; 307(13):1405-1413.

Medicare Payment Advisory Commission. Report to the Congress: Promoting Greater Efficiency in Medicare. 2007.

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):*

Pulmonary/Critical Care : Pneumonia

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

Care Coordination, Care Coordination : Readmissions, Safety, Safety : Complications, Safety : Healthcare Associated

Infections

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:** NQF_0506_PN_Readmission_S2b_Readmission_Data_Dictionary_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.
Annual Updates
1. Updated CC map.

a. Rationale: The ICD-9-CM CC map was updated annually to capture all relevant comorbidities coded in patient administrative claims data.

Planned Update for 2016 public reporting – (changes reflected in this application)
1. Expanded cohort to include patients with a principal diagnosis of aspiration pneumonia and those with a principal discharge diagnosis of sepsis (not including severe sepsis) who have a secondary discharge diagnosis of pneumonia (including aspiration pneumonia) coded as POA and no secondary discharge diagnosis of severe sepsis.

a. Rationale: The cohort was expanded to capture a broader population of patients admitted for pneumonia and to capture a consistent clinical cohort across hospitals. The cohort expansion responds to changing coding patterns in which patients with pneumonia are increasingly given a principal discharge diagnosis of sepsis. As hospitals increasingly use a principal discharge diagnosis code of sepsis in combination with a secondary discharge diagnosis of pneumonia that is POA, such patients would be excluded from the measure without the cohort expansion. Furthermore, variation in the use of sepsis coding across hospitals could lead to differential exclusion of pneumonia patients from the measures across hospitals which could bias efforts to comparatively assess hospital quality. (Please see updated 2015 Reevaluation and Re-Specification Report of the Hospital-Level 30-Day Risk-Standardized Measures Following Hospitalization for Pneumonia Readmission, version 8.2 for more details on the modifications made to this measure and final measure specifications. The report is posted on the CMS.gov website at http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Measure-Methodology.html. The pneumonia readmission report [version 8.2] can be found in the AMI, HF, PN, COPD, and Stroke Readmission Updates zip folder.)

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 readmission. We define readmission as an inpatient admission for any cause, with the exception of certain planned readmissions, within 30 days from the date of discharge from the index admission for patients 18 and older discharged from the hospital with a principal discharge diagnosis of pneumonia, including aspiration pneumonia or a principal discharge diagnosis of sepsis (not severe sepsis) with a secondary discharge diagnosis of pneumonia (including aspiration pneumonia) coded as POA and no secondary

discharge diagnosis of severe sepsis. If a patient has more than one unplanned admission (for any reason) within 30 days after discharge from the index admission, only the first one is counted as a readmission. The measure looks for a dichotomous yes or no outcome of whether each admitted patient has an unplanned readmission within 30 days. However, if the first readmission after discharge is considered planned, any subsequent unplanned readmission is not counted as an outcome for that index admission because the unplanned readmission could be related to care provided during the intervening planned readmission rather than during the index admission.

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 pneumonia hospitalization.

Denominator Time Window: This original measure was developed with 12 months of data. The re-specified measure with the expanded pneumonia cohort (version 8.2) was tested with three years of data. The time window can be specified from one to three years. Currently, the measure is publicly reported with three years of index hospitalizations.

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.
The measure counts readmissions to any acute care hospital for any cause within 30 days of the date of discharge of the index pneumonia admission, excluding planned readmissions as defined below.

Planned Readmission Algorithm (Version 4.0)
The planned readmission algorithm is a set of criteria for classifying readmissions as planned among the general Medicare population using Medicare administrative claims data. The algorithm identifies admissions that are typically planned and may occur within 30 days of discharge from the hospital.

The planned readmission algorithm has three fundamental principles:

1. A few specific, limited types of care are always considered planned (transplant surgery, maintenance chemotherapy/ immunotherapy, rehabilitation);
2. Otherwise, a planned readmission is defined as a non-acute readmission for a scheduled procedure; and
3. Admissions for acute illness or for complications of care are never planned.

The algorithm was developed in 2011 as part of the Hospital-Wide Readmission measure. In 2013, CMS applied the algorithm to its other readmission measures. In applying the algorithm to condition- and procedure-specific measures, teams of clinical experts reviewed the algorithm in the context of each measure-specific patient cohort and, where clinically indicated, adapted the content of the algorithm to better reflect the likely clinical experience of each measure's patient cohort. The planned readmission algorithm is applied to the pneumonia measure without modifications.

The planned readmission algorithm and associated code tables are attached in data field S.2b (Data Dictionary or Code Table).

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 over or (2)

patients aged 18 years or older. We have specifically tested the measure in both age groups.

The cohort includes admissions for patients aged 18 years and older discharged from the hospital with principal discharge diagnosis of pneumonia, including aspiration pneumonia or a principal discharge diagnosis of sepsis (not severe sepsis) with a secondary discharge diagnosis of pneumonia (including aspiration pneumonia) coded as POA and no secondary discharge diagnosis of severe sepsis; and with a complete claims history for the 12 months prior to admission. The measure will be publicly reported by CMS for those patients 65 years and older who are Medicare FFS beneficiaries admitted to non-federal hospitals.

Additional details are provided in S.9 Denominator Details.

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

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)*
To be included in the measure cohort used in public reporting, patients must meet the following inclusion criteria:

1. Principal discharge diagnosis of pneumonia, including aspiration pneumonia; or
Principal discharge diagnosis of sepsis (not including severe sepsis), with a secondary discharge diagnosis of pneumonia (including aspiration pneumonia) coded as POA but no secondary discharge diagnosis of severe sepsis.
2. Enrolled in Medicare fee-for-service (FFS)
3. Aged 65 or over
4. Not transferred from another acute care facility
5. Enrolled in Part A and Part B Medicare for the 12 months prior to the date of admission, and enrolled in Part A during the index admission.

This measure can also be used for an all-payer population aged 18 years and older. We have explicitly tested the measure in both patients aged 18 years and older; and those aged 65 years or over (see Testing Attachment for details).

International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) codes used to define the cohort for each measure are:

ICD-9 codes that define patients with pneumonia:
480.0 Pneumonia due to adenovirus
480.1 Pneumonia due to respiratory syncytial virus
480.2 Pneumonia due to parainfluenza virus
480.3 Pneumonia due to SARS-associated coronavirus
480.8 Pneumonia due to other virus not elsewhere classified
480.9 Viral pneumonia, unspecified
481 Pneumococcal pneumonia
482.0 Pneumonia due to Klebsiella pneumoniae
482.1 Pneumonia due to Pseudomonas
482.2 Pneumonia due to Hemophilus influenzae
482.30 Pneumonia due to Streptococcus, unspecified
482.31 Pneumonia due to Streptococcus, group A
482.32 Pneumonia due to Streptococcus, group B
482.39 Pneumonia due to other Streptococcus
482.40 Pneumonia due to Staphylococcus, unspecified
482.41 Methicillin susceptible pneumonia due to Staphylococcus aureus

482.42 Methicillin resistant pneumonia due to Staphylococcus aureus
482.49 Other Staphylococcus pneumonia
482.81 Pneumonia due to anaerobes
482.82 Pneumonia due to escherichia coli
482.83 Pneumonia due to other gram-negative bacteria
482.84 Pneumonia due to Legionnaires´ disease
482.89 Pneumonia due to other specified bacteria
482.9 Bacterial pneumonia, unspecified
483.0 Pneumonia due to mycoplasma pneumoniae
483.1 Pneumonia due to chlamydia
483.8 Pneumonia due to other specified organism
485 Bronchopneumonia, organism unspecified
486 Pneumonia, organism unspecified
487.0 Influenza with pneumonia
488.11 Influenza due to identified 2009 H1N1 influenza virus with pneumonia

ICD-9 codes that define patients with aspiration pneumonia:
507.0 Pneumonitis due to inhalation of food or vomitus

ICD-9 codes that define patients with sepsis (not including severe sepsis [995.92 or 785.52]) (Cohort requires principal discharge diagnosis of sepsis combined with a secondary discharge diagnosis of pneumonia or aspiration pneumonia coded as POA but no secondary discharge diagnosis of severe sepsis):
038.0 Streptococcal septicemia
038.10 Staphylococcal septicemia, unspecified
038.11 Methicillin susceptible Staphylococcus aureus septicemia
038.12 Methicillin resistant Staphylococcus aureus septicemia
038.19 Other staphylococcal septicemia
038.2 Pneumococcal septicemia [Streptococcus pneumoniae septicemia]
038.3 Septicemia due to anaerobes
038.40 Septicemia due to gram-negative organism, unspecified
038.41 Septicemia due to hemophilus influenzae [H. influenzae]
038.42 Septicemia due to escherichia coli [E. coli]
038.43 Septicemia due to pseudomonas
038.44 Septicemia due to serratia
038.49 Other septicemia due to gram-negative organisms
038.8 Other specified septicemias
038.9 Unspecified septicemia
995.91 Sepsis

ICD-10 codes that define patients with pneumonia:
J12.0 Adenoviral pneumonia
J12.1 Respiratory syncytial virus pneumonia
J12.2 Parainfluenza virus pneumonia
J12.81 Pneumonia due to SARS-associated coronavirus
J12.89 Other viral pneumonia
J12.9 Viral pneumonia, unspecified
J13 Pneumonia due to Streptococcus pneumoniae
J18.1 Lobar pneumonia, unspecified organism
J15.0 Pneumonia due to Klebsiella pneumoniae
J15.1 Pneumonia due to Pseudomonas
J14 Pneumonia due to Hemophilus influenzae

J15.4 Pneumonia due to other streptococci
J15.3 Pneumonia due to streptococcus, group B
J15.20 Pneumonia due to staphylococcus, unspecified
J15.211 Pneumonia due to Methicillin susceptible staphylococcus
J15.212 Pneumonia due to Methicillin resistant staphylococcus
J15.29 Pneumonia due to other staphylococcus
J15.8 Pneumonia due to other specified bacteria
J15.5 Pneumonia due to Escherichia coli
J15.6 Pneumonia due to other aerobic Gram-negative bacteria
A48.1 Legionnaires´ disease
J15.8 Pneumonia due to other specified bacteria
J15.9 Unspecified bacterial pneumonia
J15.7 Pneumonia due to Mycoplasma pneumoniae
J16.0 Chlamydial pneumonia
J16.8 Pneumonia due to other specified infectious organisms
J18.0 Bronchopneumonia, unspecified organism
J18.9 Pneumonia, unspecified organism
J11.00 Influenza due to unidentified influenza virus with unspecified type of pneumonia
J12.9 Viral pneumonia, unspecified
J10.08 Influenza due to other identified influenza virus

ICD-10 codes that define patients with aspiration pneumonia:
J69.0 Pneumonitis due to inhalation of food and vomit

ICD-10 codes that define patients with sepsis (not including severe sepsis [ICD-9 995.92 or 785.52]) (Cohort requires principal discharge diagnosis of sepsis combined with a secondary discharge diagnosis of pneumonia or aspiration pneumonia coded as POA but no secondary discharge diagnosis of severe sepsis):
A40.9 Streptococcal sepsis, unspecified
A41.2 Sepsis due to unspecified staphylococcus
A41.01 Sepsis due to Methicillin susceptible Staphylococcus
A41.02 Sepsis due to Methicillin resistant Staphylococcus
A41.1 Sepsis due to other specified staphylococcus
A40.3 Sepsis due to Streptococcus pneumoniae
A41.4 Sepsis due to anaerobes
A41.50 Gram-negative sepsis, unspecified
A41.3 Sepsis due to Hemophilus influenzae
A41.51 Sepsis due to Escherichia coli [E. coli]
A41.52 Sepsis due to Pseudomonas
A41.53 Sepsis due to Serratia
A41.59 Other Gram-negative sepsis
A41.89 Other specified sepsis
A41.9 Sepsis, unspecified organism

An ICD-9 to ICD-10 crosswalk is attached in field S.2b. (Data Dictionary or Code Table).

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
The readmission measures exclude index admissions for patients:

1. Discharged against medical advice (AMA);
2. Without at least 30 days post-discharge enrollment in FFS Medicare;
3. Admitted within 30 days of a prior index admission.

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)*
1. Discharges against medical advice (AMA) are identified using the discharge disposition indicator in claims data.

2. Admissions without at least 30 days post-discharge enrollment in FFS Medicare are determined by examining the Medicare Enrollment Database (EDB).

3. Pneumonia admissions within 30 days of discharge from a qualifying pneumonia index admission are identified by comparing the discharge date from the index admission with subsequent admission dates.

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 employs a hierarchical logistic regression model to create a hospital-level 30-day RSRR. In brief, the approach simultaneously models data at the patient and hospital levels to account for the variance in patient outcomes within and between hospitals (Normand & Shahian, 2007). At the patient level, the model adjusts the log-odds of readmission within 30 days of admission for age, sex, and selected clinical covariates. At the hospital level, the approach models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of readmission at the hospital, after accounting for patient risk. If there were no differences among hospitals, then after adjusting for patient risk, the hospital intercepts should be identical across all hospitals.

Candidate and Final Risk-adjustment Variables:
Candidate variables were patient-level risk-adjustors that were expected to be predictive of readmission, based on empirical analysis, prior literature, and clinical judgment, including age, sex, and indicators of comorbidity and disease severity. For each patient, covariates are obtained from claims records extending 12 months prior to and including the index admission. For the measure currently implemented by CMS, these risk-adjusters are identified using both inpatient and outpatient Medicare FFS claims data. However, in the all-payer hospital discharge database measure, the risk-adjustment variables can be obtained only from inpatient claims in the prior 12 months and the index admission.

The model adjusts for case-mix differences based on the clinical status of patients 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 (Pope et al., 2000). A file that contains a list of the ICD-9-CM codes and their groupings into CCs is attached in data field S.2b (Data Dictionary or Code Table). In addition, only comorbidities that convey information about the patient at admission or in the 12 months prior, and not complications that arise during the course of the index

hospitalization, are included in the risk adjustment. Hence, we do not risk adjust for CCs that may represent adverse events of care when they are only recorded in the index admission.

The final set of risk adjustment variables is:

Demographics
Male
Age-65 (years, continuous) for patients aged 65 or over cohorts; or Age (years, continuous) for patients aged 18 and over cohorts.

Comorbidities
History of Coronary Artery Bypass Graft (CABG) (ICD-9 codes V45.81, 36.10–36.16)
History of infection (CC1, 3-6)
Septicemia/sepsis (CC 2)
Metastatic cancer or acute leukemia (CC 7)
Lung, upper digestive tract, and other severe cancers (CC 8)
Other major cancers (CC 9-10)
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 gastrointestinal disorders (CC 36)
Severe hematological disorders (CC 44)
Iron deficiency or other unspecified anemias and blood disease (CC 47)
Dementia or other specified brain disorders (CC 49-50)
Drug/alcohol abuse/dependence/psychosis (CC 51-53)
Major psychiatric disorders (CC 54-56)
Other psychiatric disorders (CC 60)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)
Cardio-respiratory failure or shock (CC 78-79)
Congestive heart failure (CC 80)
Acute coronary syndrome (CC 81-82)
Coronary atherosclerosis or angina (CC 83-84)
Valvular or rheumatic heart disease (CC 86)
Specified arrhythmias and other heart rhythm disorders (CC 92-93)
Stroke (CC 95-96)
Vascular or circulatory disease (CC 104-106)
Chronic obstructive pulmonary disease (COPD) (CC 108)
Fibrosis of lung or other chronic lung disorders (CC 109)
Asthma (CC 110)
Pneumonia (CC 111-113)
Pleural effusion/pneumothorax (CC 114)
Other lung disorders (CC 115)
End-stage renal disease or dialysis (CC 129-130)
Renal failure (CC 131)
Urinary tract infection (CC 135)
Other urinary tract disorders (CC 136)
Decubitus ulcer or chronic skin ulcer (CC 148-149)
Vertebral fractures (CC 157)
Other injuries (CC 162)
Respirator dependence/tracheostomy (CC 77)

References:
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.

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

Pope GC, et al. 2000. Principal Inpatient Diagnostic Cost Group Models for Medicare Risk Adjustment. Health Care Financing Review 21(3): 93-118.

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*)

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.)*
The measure estimates hospital-level 30-day, all-cause, RSRRs following hospitalization for pneumonia using hierarchical logistic regression models. In brief, the approach simultaneously models data at the patient and hospital levels to account for variance in patient outcomes within and between hospitals (Normand and Shahian, 2007). At the patient level, it models the log-odds of readmission within 30 days of index admission using age, sex, selected clinical covariates, and a hospital-specific intercept. At the hospital level, it models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of a readmission at the hospital, after accounting for patient risk. The hospital-specific intercepts are given a distribution 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.

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

readmission rates or worse quality.

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

This calculation transforms the ratio of predicted over expected into a rate that is compared to the national observed readmission rate. The hierarchical logistic regression models are described fully in the original methodology report (Krumholz et al., 2008).

Reference:
Krumholz H, Normand S-LT, Keenan P, et al. Hospital 30-Day Pneumonia Readmission Measure Methodology. 2008.

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

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.

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 survey or patient-reported data.

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.
Administrative claims

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:

1. 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.

2. 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).

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

4. Data sources for the all-payer update: For our analyses to examine use in all-payer data, we used all-payer data from California in addition to CMS data for Medicare FFS 65+ patients in California hospitals. 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 2009, there were 3,193,904 adult discharges from 446 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 as well as CMS Medicare FFS data for California hospitals, we performed analyses to determine whether the pneumonia mortality measure can be applied to all adult patients, including not only FFS Medicare patients aged 65+ 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/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_0506_PN_Readmission_NQF_Testing_Attachment_01-29-16_v1.1.docx

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

Measure Number (*if previously endorsed*)**:** 0506
Measure Title: Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization
Date of Submission: 1/29/2016
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☒ Outcome (*including PRO-PM*)
☐ 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 as well as the Medicare Enrollment Database (EDB). Additionally, census as well as claims data were used to assess socioeconomic factors and race (dual eligible and African American race variables obtained through enrollment data; Agency for Healthcare Research and Quality [AHRQ] socioeconomic status [SES] index 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? Click here to enter date range

The dates used vary by testing type; see Section 1.7 for details.

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

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

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

For this measure, hospitals are the measured entities. All non-federal, acute care inpatient US hospitals (including territories) with Medicare fee-for-service (FFS) beneficiaries aged 65 years or over are included. The number of measured entities (hospitals) varies by testing type; see Section 1.7 for details.

1.6. How many and which patients 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 hospitals, and number of admissions used in each type of testing are as follows:

For reliability testing (Section 2a2)
The reliability of the model was tested by randomly selecting 50% of the Medicare patients aged 65 years or over from the expanded pneumonia cohort (measure version 8.2) and applying the re-specified risk-adjusted model for this group. We then applied the same model for the remaining 50% of patients and compared the two. Thus, for reliability testing, we randomly split **Dataset 1** into two samples. In each year of measure maintenance, we also re-fit the model and compared the frequencies and model coefficients of risk variables (condition categories for patient comorbidities) and model fit across 3 years (**Dataset 1** below).

Dataset 1 (expanded cohort, measure version 8.2): Medicare Part A Inpatient and Outpatient and Part B Outpatient claims
Dates of Data: July 1, 2011 – June 30, 2014 (expanded cohort)
Number of Admissions: 1,469,277
Patient Descriptive Characteristics: average age=81.0, % male=46.6
Number of Measured Hospitals: 4,700

First half of split sample
-Number of Admissions: 733,434
-Number of Measured Hospitals: 4,670

Second half of split sample
-Number of Admissions: 735,843
-Number of Measured Hospitals: 4,700

For validity testing (Section 2b2)
Split samples of **Dataset 1**

For testing of measure exclusions (Section 2b3)
Dataset 1

For testing of measure risk adjustment (Section 2b4)
Dataset 1

Dataset 2 (all payer dataset): California Patient Discharge Data

Dates of Data: January 1, 2009 – December 31, 2009
Number of Admissions: 78,780 (all 18+ total); 29,244 (FFS 65+); 13,251 (non-FFS 65+); 30,957 (all 18-64)
Patient Descriptive Characteristics: mean age=67.8, %male=49.9% (all 18+ total); mean age=80.7, %male=48.5% (FFS 65+); mean age=80.5, %male=49.5% (non-FFS 65+); mean age=48.4, %male=51.2% (all 18-64)
Number of Measured Hospitals: 317 non-Federal acute care hospitals

The measure was applied to California Patient Discharge Data, a large, linked all-payer database of patient hospital admissions. Records are linked by a unique patient identification number, allowing us to determine patient history from previous hospitalizations.

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

For testing of socioeconomic status (SES) factors and race in risk models (Section 2b4.3)
Dataset 1 and **Dataset 3:** 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 SES.

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 1**)
- Validated AHRQ SES index score is a composite of 7 different variables found in the census data (**Dataset 3**)

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 SES variables.

The 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 pneumonia is more limited though studies suggest a possible increased risk of readmission in particular with the inclusion of race variables (Calvillo-King et al., 2013; Joynt et al., 2011; Lindenauer et al., 2013; McHugh et al., 2010; Mather et al., 2014; Vidic et al., 2015).

The causal pathways for SES and race variables' effects are described below in Section 2b4.3.

The SES and race variables used for analysis were:

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

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 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, 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 variable that describes the average SES of people living in defined geographic areas (Bonito et al., 2008). Its value as a proxy for patient-level information is dependent on having the most granular level data with respect to communities that patients live in. In this submission, we present analysis using the census block level, the most granular level possible using American Community Survey data. We used 2009-2013 American Community Survey data and mapped patients' 9-digit ZIP codes via vendor software to the AHRQ SES Index at the census block group level. Given the variation in cost of living across the country, the median income and median property value components of the AHRQ SES Index were adjusted by regional price parity values published by the Bureau of Economic Analysis (BEA). This provides a better marker of low SES neighborhoods in high expense geographic areas. We then calculated an AHRQ SES Index score for census block groups that can be linked to 9-digit ZIP codes. In the PN measure cohort, we were able to assign an AHRQ SES Index score to 99.5% of patient admissions. 86.5% of patient admissions had calculated AHRQ SES Index scores linked to their 9-digit ZIP codes. 13.0% of patient admissions had only valid 5-digit ZIP codes; we utilized the data for the median 9-digit ZIP code within that 5-digit ZIP code.

References:

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.

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

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. J Gen Intern Med. 2013 Feb; 28(2):269-82. doi: 10.1007/s11606-012-2235-x. Epub 2012 Oct 6.

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.

Gilman M, Adams EK, Hockenberry JM, et al. 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 KE, Orav EJ, Jha AK. Thirty-day readmission rates for Medicare beneficiaries by race and site of care. JAMA. 2011 Feb 16; 305(7):675-81. doi: 10.1001/jama.2011.123.

Lindenauer PK, Lagu T, Rothberg MB, et al. Income inequality and 30 day outcomes after acute myocardial infarction, heart failure, and pneumonia: retrospective cohort study. BMJ. 2013 Feb 14; 346:f521. doi: 10.1136/bmj.f521.

Mather JF, Fortunato GJ, Ash JL, et al. Prediction of pneumonia 30-day readmissions: a single-center attempt to increase model performance. Respir Care. 2014 Feb; 59(2):199-208. doi: 10.4187/respcare.02563. Epub 2013 Aug 13.

McHugh MD, Carthon JM, Kang XL. Medicare readmissions policies and racial and ethnic health disparities: a cautionary tale. Policy Polit Nurs Pract. 2010 Nov; 11(4):309-16. doi: 10.1177/1527154411398490.

Vidic A, Chibnall JT, Hauptman PJ. Heart failure is a major contributor to hospital readmission penalties. J Card Fail. 2015 Feb; 21(2):134-7. doi: 10.1016/j.cardfail.2014.12.002. Epub 2014 Dec 9.

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 in 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 used 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**)

The frequency of some model variables increased and others decreased between 2011 and 2014, which may reflect an increase or decrease rate of comorbidities in the FFS population. For example, there was a notable increase in percent frequency for "septicemia/sepsis (CC 2)" (11.5% to 12.8%), "drug/alcohol abuse/dependence/psychosis (CC 51-53)" (15.5% to 16.8%), "other psychiatric disorders (CC 60)" (20.0% to 24.5%), "cardio-respiratory failure or shock (CC 78-79)" (23.7% to 25.6%), "pneumonia (CC 111-113)" (52.2% to 53.4%), and "renal failure (CC 131)" (30.7% 32.7%). There was a notable decrease in percent frequency for "severe hematological disorders (CC 44)" (3.6% to 2.2%), "dementia or other specified brain disorders (CC 49-50)" (37.9% to 36.6%), "congestive heart failure (CC 80)" (40.2% to 39.0%), "coronary atherosclerosis or angina (CC 83-84)" (50.1% to 48.8%), "fibrosis of lung or other chronic lung disorders (CC 109)" (16.0% to 13.5%), "other lung disorders (CC 115)" (47.2% to 44.4%), "urinary tract infection (CC 135)" (32.7% to 31.6%), and "other urinary tract disorders (CC 136)" (26.4% to 25.1%). Examination of the odds ratios for each risk variable in the model shows that, overall, the odds ratios for individual risk variables remained relatively constant across three years.

These frequencies are from the expanded cohort and re-specified model (measure version 8.2).

Measure Score Reliability Results (**Dataset 1**)
There were 1,469,277 admissions in the combined 3-year sample, with 733,434 in one sample and 735,843 in the other randomly selected sample. The agreement between the two RSRRs for each hospital was 0.73, which according to the conventional interpretation is "substantial" (Landis and Koch, 1977).

Note that this analysis was limited to hospitals with 12 or more cases in each split sample. The intra-class correlation coefficient is based on a split sample of three years of data, resulting in a volume of patients in each sample equivalent to only 1.5 years of data, whereas the measure is reported with the full three years of data.

Reference:
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 over time of the risk factor frequencies and odds ratios suggests that the underlying data elements are reliable. Additionally, the ICC score demonstrates substantial agreement of measure scores across samples using a conservative approach to assessment.

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)*

The measure's validity is demonstrated in three manners. The first is clinical and face validity of the cohort expansion. As discussed in the 2015 Reevaluation and Re-Specification Report of the Hospital-Level 30-Day Risk-Standardized Measures Following Hospitalization for Pneumonia (Mortality, version 9.2; Readmission, version 8.2) (Lindenauer et al., 2015), made publicly available to support the FY 2016 IPPS rule, the cohort expansion is based on changes in clinical and coding practices that have led to greater numbers of patients with pneumonia being coded with sepsis or aspiration pneumonia as a principal discharge diagnosis. These are patients that the

measure is intended to assess, as they fit within the broad clinical category of pneumonia patients and are often treated by the same groups of physicians and staff, using similar treatment strategies. Moreover, virtually all patients hospitalized with pneumonia meet criteria for sepsis. The expansion was also supported by findings in the literature (Lindenauer et al., 2012; Rothberg et al., 2014).

Second, for a number of claims-based outcome measures, including the original version of this measure, we validated the administrative model with a medical-record based model. In this earlier study, we demonstrated that the rates calculated using the risk adjustment model with claims and medical record data were highly correlated (Krumholz et al., 2008). These analyses, though based on an earlier version of this measure, demonstrated that using comorbidity information from administrative claims data is a valid approach to risk adjustment and specifically, that claims-based risk adjustment adequately assesses the difference in case mix among hospitals. The claims-based measure produced results which were highly correlated with those produced through manual chart audit (Krumholz et al., 2008; Lindenauer et al., 2011). The revised pneumonia readmission measure utilizes the same approach as the original (now, currently publicly reported) measure. When developing the expanded cohort for the readmission measure, we re-examined the risk ratios for the risk variables used in the original (or current) measure, which showed that the variables remained predictive of the outcome (that is, readmission). Also, model performance characteristics were similar to those of the current pneumonia readmission measure.

As we demonstrated in our analyses in the 2015 Reevaluation Report (Lindenauer et al., 2015), although the revision is bringing in a large portion of patients currently not included in the measure, the revised version of the measure likely has greater validity in that it has mitigated biases introduced by hospital coding patterns. We confirmed that the approach to risk adjustment was effective, as hospital coding frequency was no longer associated with performance on the revised measure.

Third, as part of measure validation, we tested the performance of the pneumonia readmission model developed in the first half of a randomly split sample of pneumonia hospitalizations from **Dataset 1** (representing 733,434 admissions from 4,670 hospitals) against the second half of the randomly split sample of pneumonia hospitalizations (representing 735,843 admissions from 4,700 hospitals).

References:

Krumholz H, Normand S, Keenan P, et al. Hospital 30-Day Pneumonia Readmission Measure Methodology. 2008; http://www.qualitynet.org/dcs/BlobServer?blobkey=id&blobnocache=true&blobwhere=1228873654295&blobheader=multipart%2Foctet-stream&blobheadername1=Content-Disposition&blobheadervalue1=attachment%3Bfilename%3DPneumo_ReadmMeasMethod.pdf&blobcol=urldata&blobtable=MungoBlobs. Accessed December 9, 2015.

Lindenauer P, Lagu T, Shieh M, Pekow P, Rothberg M. Association of diagnostic coding with trends in hospitalizations and mortality of patients with pneumonia, 2003-2009. Jama. Apr 4 2012; 307(13):1405-1413.

Lindenauer P, Normand S, Drye E, et al. Development, validation, and results of a measure of 30-day readmission following hospitalization for pneumonia. Journal of hospital medicine. Mar 2011; 6(3):142-150.

Lindenauer P, Ross J, Strait K, et al. 2015 Reevaluation and Re-Specification Report of the Hospital-Level 30-Day Risk-Standardized Measures Following Hospitalization for Pneumonia Mortality, version 9.2; Pneumonia Readmission, version 8.2. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Measure-Methodology.html. Accessed 12 November, 2015.

Rothberg M, Pekow P, Priya A, Lindenauer P. Variation in diagnostic coding of patients with pneumonia and its association with hospital risk-standardized mortality rates: a cross-sectional analysis. Annals of internal medicine. Mar 18 2014; 160(6):380-388.

ICD-9 to ICD-10 Conversion

Statement of Intent

[X] Goal was to convert this measure to a new code set, fully consistent with the intent of the original measure.

[ ] Goal was to take advantage of the more specific code set to form a new version of the measure, but fully consistent with the original intent.

[ ] The intent of the measure has changed.

Process of Conversion

ICD-10 codes were identified using 2015 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. An ICD-9 to ICD-10 crosswalk is attached in field S.2b. (Data Dictionary or Code Table).

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

The performance of the first half of the split sample and second half of the split sample from **Dataset 1** was similar. The areas under the receiver operating characteristic (ROC) curve for the two models are 0.6419 and 0.6436, respectively.

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?*)

The results between the first half of the split sample and second half of the split sample from **Dataset 1** proved to be similar for each of the model testing that was performed. The ROC results were nearly identical and in line with other readmission models.

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*)

All exclusions were determined by careful clinical review and have been made based on clinically relevant decisions to ensure accurate calculation of the measure. To ascertain impact of exclusions on the cohort, we examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion (**Dataset 1**). These exclusions are consistent with similar NQF-endorsed outcome measures. Rationales for the exclusions are detailed in data field S.10 (Denominator Exclusions).

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*)

In **Dataset 1:**

Exclusion	**N**	**%**	**Distribution across hospitals (N=4,340): Min, 25th, 50th, 75th percentile, max**
1. Discharged against medical advice (AMA)	11,621	0.55%	(0.00, 0.00, 0.33, 0.76, 9.09)
2. Without at least 30 days post-discharge enrollment in FFS Medicare for index admissions	169,803	7.98%	(0.00, 5.41, 7.21, 9.09, 86.95)
3. Pneumonia admission within 30 days of a prior pneumonia index admission	64,916	3.05%	(0.00, 1.94, 2.82, 3.61, 14.29)

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*)

Exclusion 1 (patients who are discharged AMA) accounts for 0.55% of all index admissions excluded from the initial index cohort. This exclusion is needed for acceptability of the measure to hospitals, who do not have the opportunity to deliver full care and prepare the patient for discharge. Given that a very small percent of patients are being excluded, it is unlikely that is exclusion affects the measure score.

Exclusion 2 (patients without at least 30 days post-discharge enrollment in FFS Medicare for index admissions in non-VA hospitals) accounts for 7.98% of all index admissions excluded from

the initial index cohort. This exclusion is needed since the 30-day readmission outcome cannot be assessed in this group since claims data are used to determine whether a patient was readmitted.

For **exclusion 3** (patients with admissions within 30 days of a prior index admission) if a patient has an admission within 30 days of discharge from the index admission, that admission is not included in the cohort so that admission can be both an index admission and readmission. This exclusion accounts for 3.05% of all index admissions excluded from the initial index cohort.

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 41 risk factors**

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

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

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (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 discharge 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.

As part of measure reevaluation in 2015, the pneumonia cohort was expanded and the risk model was re-specified. The revised pneumonia readmission measure (version 8.2):

1. Retains the clinical comorbidity variables included in the current readmission (version 8.0) risk model.

2. Incorporates the following new risk-adjustment variable if present in the 12 months prior to the index admission:
- Respiratory dependence/tracheostomy (CC77)

Although this risk variable was not included in the original measure development and validation, during measure re-evaluation we determined that this risk variable was common (that is, with a prevalence of greater than 10% in the population) and had strong associations with readmission (odds ratio [OR] > 1.5) in the expanded pneumonia cohort.

3. Modifies one clinical risk variable as follows:
- add **Respiratory Arrest (CC78)** to the cardio-respiratory failure or shock risk variable (that is, Respiratory Arrest will be added to the currently defined **Cardio-respiratory failure or shock risk variable (CC79)** in measure version 8.0, which will now be redefined in the model as **Cardio-respiratory failure or shock** (**CC78-79)** in measure version 8.2.)

Similar to the rationale for including CC77 noted above, the pneumonia readmission measure includes the clinical comorbidity risk variable **CC78** because this was common (that is, with a prevalence of greater than 10% in the population) and had strong associations with readmission (OR > 1.5). This risk variable also had high levels of face validity in terms of the clinical

expectation that this condition would be associated with worse outcomes if it occurred during the 12 months prior to the index admission.

The final set of risk-adjustment variables is:

- Age-65 (years, continuous) for patients aged 65 or over cohorts; or Age (years, continuous) for patients aged 18 and over cohorts
- Male
- History of CABG
- History of infection
- Septicemia/sepsis
- Metastatic cancer or acute leukemia
- Lung, upper digestive tract, and other severe cancers
- Other major cancers
- Diabetes mellitus (DM) or DM complications
- Protein-calorie malnutrition
- Disorders of fluid/electrolyte/acid-base
- Other gastrointestinal disorders
- Severe hematological disorders
- Iron deficiency or other unspecified anemias and blood disease
- Dementia or other specified brain disorders
- Drug/alcohol abuse/dependence/psychosis
- Major psychiatric disorders
- Other psychiatric disorders
- Hemiplegia, paraplegia, paralysis, functional disability
- Cardio-respiratory failure or shock
- Congestive heart failure
- Acute coronary syndrome
- Coronary atherosclerosis or angina
- Valvular or rheumatic heart disease
- Specified arrhythmias and other heart rhythm disorders
- Stroke
- Vascular or circulatory disease
- Chronic obstructive pulmonary disease (COPD)
- Fibrosis of lung or other chronic lung disorders
- Asthma
- Pneumonia
- Pleural effusion/pneumothorax
- Other lung disorders
- End-stage renal disease or dialysis
- Renal failure
- Urinary tract infection
- Other urinary tract disorders
- Decubitus ulcer or chronic skin ulcer
- Vertebral fractures
- Other injuries
- Respiratory dependence/tracheostomy

Socioeconomic Status (SES) Factors and Race

We selected variables representing 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 Pneumonia Readmission

To examine the relationship between SES and race variables and hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization, 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 pneumonia readmission. Seventeen studies were reviewed by title and abstract, and eleven studies were excluded from full-text review. Among studies reviewed, there was evidence that SES and race increased the risk of pneumonia readmission (Lindenauer et al., 2013; Mather et al., 2014), with a noted risk associated with race in particular (Joynt et al., 2011; McHugh et al., 2010). However, other studies including a systematic review showed that there may be a significant association but that results have been inconclusive (Calvillo-King et al., 2013; Vidic et al., 2015).

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 measure 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 socioeconomic status (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 1.8, the following SES variables and race were considered:

- African American race (as compared to all others)
- Dual eligible status
- AHRQ SES index score

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:

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 2014; 7:391-7.

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. J Gen Intern Med. 2013 Feb; 28(2):269-82. doi: 10.1007/s11606-012-2235-x. Epub 2012 Oct 6.

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.

Gilman M, Adams EK, Hockenberry JM, et al. 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 KE, Orav EJ, Jha AK. Thirty-day readmission rates for Medicare beneficiaries by race and site of care. JAMA. 2011 Feb 16; 305(7):675-81. doi: 10.1001/jama.2011.123.

Krumholz HM, Brindis RG, Brush JE, et al. Standards for statistical models used for public reporting of health outcomes. Circulation. 2006; 113: 456-462. Available at: http://circ.ahajournals.org/content/113/3/456.full.pdf+html . Accessed January 14, 2016.

Lindenauer PK, Lagu T, Rothberg MB, et al. Income inequality and 30 day outcomes after acute myocardial infarction, heart failure, and pneumonia: retrospective cohort study. BMJ. 2013 Feb 14; 346:f521. doi: 10.1136/bmj.f521.

Mather JF, Fortunato GJ, Ash JL, et al. Prediction of pneumonia 30-day readmissions: a single-center attempt to increase model performance. Respir Care. 2014 Feb; 59(2):199-208. doi: 10.4187/respcare.02563. Epub 2013 Aug 13.

McHugh MD, Carthon JM, Kang XL. Medicare readmissions policies and racial and ethnic health disparities: a cautionary tale. Policy Polit Nurs Pract. 2010 Nov; 11(4):309-16. doi: 10.1177/1527154411398490.

Normand S-LT, Shahian DM. Statistical and Clinical Aspects of Hospital Outcomes Profiling. 2007/05 2007:206-226.

Reames BN, Birkmeyer NJ, Dimick JB, et al. Socioeconomic disparities in mortality after cancer surgery: failure to rescue. JAMA surgery 2014; 149:475-81.

Skinner J, Chandra A, Staiger D, et al. 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.

Vidic A, Chibnall JT, Hauptman PJ. Heart failure is a major contributor to hospital readmission penalties. J Card Fail. 2015 Feb; 21(2):134-7. doi: 10.1016/j.cardfail.2014.12.002. Epub 2014 Dec 9.

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

Below is a table showing the final variables in the model with associated odds ratios.

Final Model Variables (variables meeting criteria in field 2b4.3) (**Dataset 1**)

Variable	07/2011-06/2014 OR (95% CI)
Age minus 65 (years above 65, continuous)	1.00 (1.00, 1.00)
Male	1.06 (1.05, 1.07)
History of Coronary Artery Bypass Graft (CABG) (ICD-9 codes V45.81, 36.10-36.16)	0.97 (0.95, 0.98)
History of infection (CC 1, 3-6)	1.04 (1.03, 1.05)
Septicemia/Sepsis (CC 2)	1.06 (1.04, 1.07)
Metastatic cancer or acute leukemia (CC 7)	1.17 (1.15, 1.20)
Lung, upper digestive tract, and other severe cancers (CC 8)	1.17 (1.15, 1.19)
Other major cancers (CC 9-10)	1.04 (1.02, 1.05)
Diabetes mellitus (DM) or DM complications (CC 15-19, 119-120)	1.09 (1.08, 1.10)
Protein-calorie malnutrition (CC 21)	1.13 (1.11,1.14)
Disorders of fluid/electrolyte/acid-base (CC 22-23)	1.14 (1.13, 1.15)
Other gastrointestinal disorders (CC 36)	1.09 (1.08, 1.10)
Severe hematological disorders (CC 44)	1.21 (1.18, 1.24)
Iron deficiency or other unspecified anemias and blood disease (CC 47)	1.19 (1.17, 1.20)
Dementia or other specified brain disorders (CC 49-50)	1.00 (0.99, 1.01)
Drug/alcohol abuse/dependence/psychosis (CC 51-53)	1.09 (1.08, 1.10)
Major psychiatric disorders (CC 54-56)	1.03 (1.01, 1.04)
Other psychiatric disorders (CC 60)	1.05 (1.04, 1.07)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)	1.08 (1.07, 1.10)
Cardio-respiratory failure or shock (CC 78-79)	1.15 (1.14, 1.16)
Congestive heart failure (CC 80)	1.16 (1.15, 1.17)
Acute coronary syndrome (CC 81-82)	1.07 (1.05, 1.09)
Coronary atherosclerosis or angina (CC 83-84)	1.05 (1.04, 1.06)
Valvular or rheumatic heart disease (CC 86)	1.07 (1.06, 1.08)
Specified arrhythmias and other heart rhythm disorders (CC 92-93)	1.09 (1.08, 1.10)
Stroke (CC 95-96)	1.04 (1.02, 1.05)
Vascular or circulatory disease (CC 104-106)	1.05 (1.04, 1.06)
Chronic obstructive pulmonary disease (COPD) (CC 108)	1.18 (1.16, 1.19)

Variable	07/2011-06/2014 OR (95% CI)
Fibrosis of lung or other chronic lung disorders (CC 109)	1.10 (1.09, 1.12)
Asthma (CC 110)	0.99 (0.98, 1.01)
Pneumonia (CC 111-113)	1.04 (1.03, 1.05)
Pleural effusion/pneumothorax (CC 114)	1.09 (1.08, 1.11)
Other lung disorders (CC 115)	1.04 (1.03, 1.05)
End-stage renal disease or dialysis (CC 129-130)	1.22 (1.19, 1.25)
Renal failure (CC 131)	1.13 (1.11, 1.14)
Urinary tract infection (CC 135)	1.05 (1.03, 1.06)
Other urinary tract disorders (CC 136)	1.05 (1.04, 1.06)
Decubitus ulcer or chronic skin ulcer (CC 148-149)	1.10 (1.09, 1.11)
Vertebral fractures (CC 157)	1.07 (1.05, 1.09)
Other injuries (CC 162)	1.04 (1.03, 1.05)
Respiratory dependence/tracheostomy (CC 77)	1.12 (1.09, 1.16)

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 pneumonia cohort varies across measured entities. The median percentage of dual eligible patients is 17.5% (interquartile range [IQR] 11.2%- 25.2%). The median percentage of African-American patients is 1.9% (IQR 0.0%- 8.3%). The median percentage of patients with an AHRQ SES index score adjusted for cost of living at the census block group level equal to or below 42.7 is 19.0% (IQR 7.8%-36.8%).

Empirical association with the outcome (univariate)

The patient-level observed pneumonia readmission rate is higher for dual-eligible patients, 20.0%, compared with 17.1% for all other patients. The readmission rate for African-American patients was also higher at 22.2% compared with 17.2% for patients of all other races. Similarly the readmission rate for patients with an AHRQ SES index score equal to or below 42.7 was 19.3% compared with 17.1% for patients with an AHRQ SES index score above 42.7.

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 multivariable model that includes all of the claims-based clinical variables, the effect size of each of these variables is small. We also find that the c-statistic is essentially unchanged with the addition of any of these variables into the model. Furthermore we find that 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 0.005% (interquartile range [IQR] -0.018% – 0.024%, minimum -0.267% – maximum 0.129%) with a correlation coefficient between RSRRs for each hospital with and without dual eligibility added of 0.99961. The median absolute change in hospitals' RSRRs when adding a race indicator is 0.035% (IQR -0.038% – 0.086%, minimum -1.337% – maximum 0.226%) with a correlation coefficient between RSRRs for each hospital with and without race added of 0.99608. The median absolute

change in hospitals' RSRRs when adding a low AHRQ SES Index score indicator to the model is 0.0342% (IQR -0.0254% – 0.0806%, minimum -0.5159% – maximum 0.2296%) with a correlation coefficient between RSRRs for each hospital with and without an indicator for a low AHRQ SES Index score adjusted for cost of living at the census block group level is 0.9981.

Contextual Effect Analysis

As described in 2b4.3, we performed a decomposition analysis for each SES and race variable to assess whether there was a corresponding contextual effect. In order to better interpret the magnitude of results, we performed the same analysis for selected clinical risk factors. The results are described in the first table below (the decomposition table).

Both the patient-level and hospital-level dual eligible, race, and low AHRQ SES Index effects were significantly associated with pneumonia readmission in the decomposition analysis. **That the hospital level effects were significant indicates that if the dual eligible, race, or low AHRQ SES Index variables are 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**.

To assess the relative contributions of the patient- and hospital-level effects, we calculated a range of predicted probabilities of readmission for the SES or race variables and clinical covariates (comorbidities), as described in section 2b4.3. The results are presented in the figure and second table below (table of predicted probabilities for SES and race variables).

For SES variables, the hospital-level effect (P95-P5) is greater than the patient-level effect (delta) (second table below; the table of predicted probabilities for SES and race variables). For the race variable, the patient-level effect (delta) is greater than the hospital-level effect (P95-P5) (second table below; the table of predicted probabilities for SES and race variables). For clinical variables, the patient-level effect (delta) is greater than the hospital-level effect (P95-P5) for lung cancer and COPD (third table below; the table of predicted probabilities for clinical variables). The hospital-level effect (P95-P5) is greater than the patient-level effect (delta) for renal failure (third table below; the table of predicted probabilities for clinical variables). There is a consistent pattern demonstrating that SES variables have a much greater hospital-level effect than patient-level effect. Notably, the race variable had a slightly greater patient-level effect. The clinical variables had the opposite pattern, with a greater effect at the patient level than at the hospital level for lung cancer and COPD. However, renal failure had a similar hospital-level and patient-level effect. In sum, including SES variables into the model would predominantly adjust for a hospital-level effect, which is an important signal of hospital quality.

In the context of our conceptual model, we find clear evidence supporting the first two mechanisms by which SES might be related to poor outcomes. First we find that, although unadjusted rates of readmission are higher for patients of low SES or African-American race, the addition of SES to the readmission risk model, which already adjusts for clinical factors, makes very little difference. In particular, there is little-to-no change in model performance or hospital results with the addition of SES. This suggests that the model already largely accounts for the differences in clinical risk factors (degree of illness and comorbidities) among patients of varied SES.

Second, the predominance of the hospital-level effect of SES and race variables in the decomposition analyses suggests the risk associated with low SES is in large part due to lower quality of care at hospitals where more patients with these risk factors are treated; hospitals caring for socially- and economically-disadvantaged patients have higher readmission risk for all of their patients. Patients with low SES or African-American race indicators tend to receive care more frequently at lower quality hospitals compared with patients with high SES indicators. Direct adjustment for patient SES would essentially "over adjust" the measure, that is to say, it would be adjusting for an endogenous factor, one that influences the outcome through the site of treatment (hospital), as much as through an attribute of the patient.

In comparison, we did not observe the same predominance of the hospital-level effect among the clinical covariates, reinforcing the sense that SES and race factors have a distinct causal pathway in their impact on readmission risk.

Summary

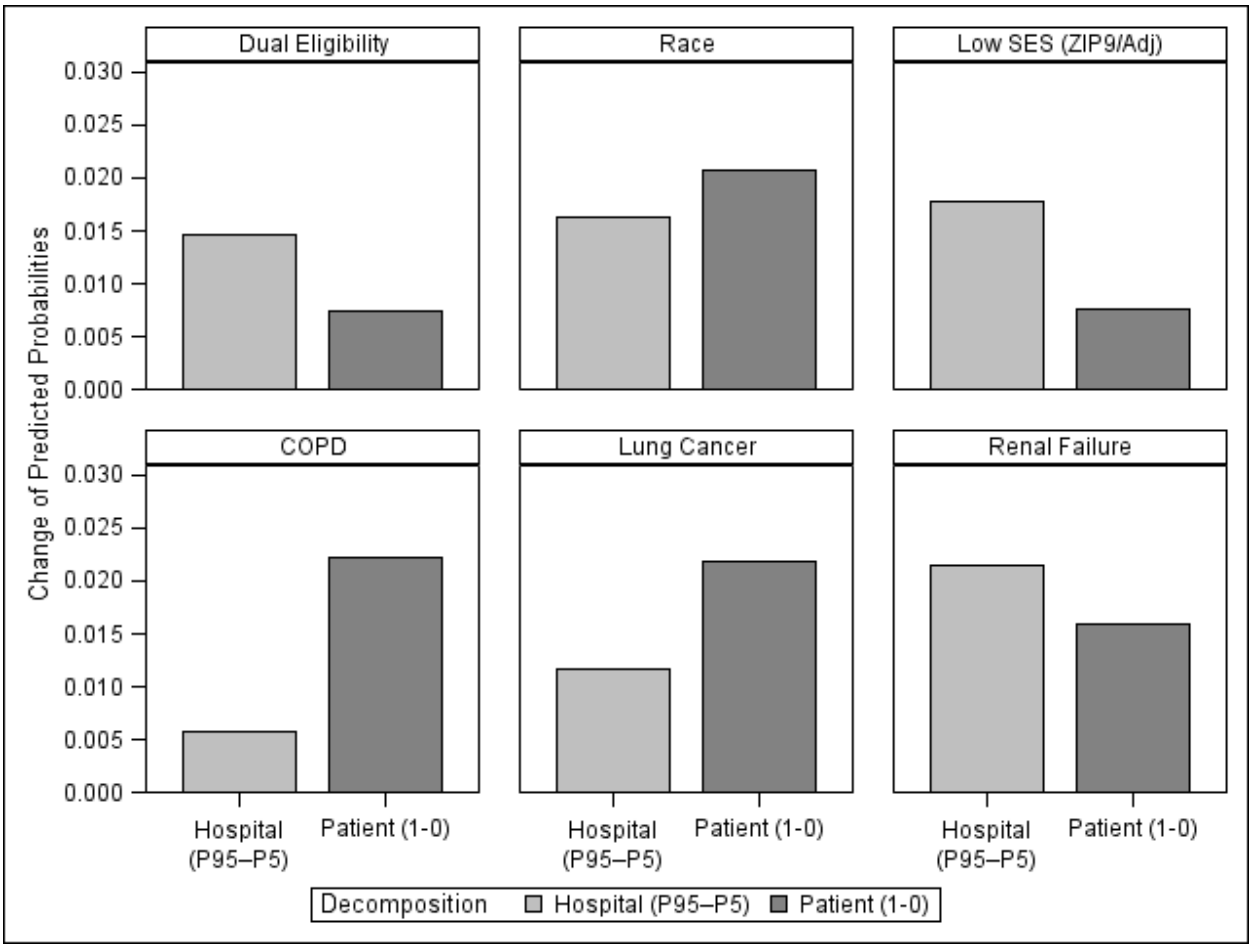
We found wide variation in the distribution of the three SES and race factors we examined, and we found that all three had some association with readmission risk. However, adjustment for these factors did not have an appreciable impact on hospital RSRRs, suggesting that existing clinical risk factors capture much of the risk related to low SES and African-American race. More importantly, we found that for all three factors there was a greater hospital-level effect, compared with the patient-level effect, indicating that patient-level adjustment alone would adjust for quality differences between hospitals. Therefore, we did not include SES or race factors in our final risk model.

Pneumonia Readmission Decomposition Analysis

Parameter	Estimate (Standard Error)	P-value
Dual Eligible – Patient-Level	0.0517 (0.0059)	<0.0001
Dual Eligible – Hospital-Level	0.2791 (0.0338)	<0.0001
African American – Patient-Level	0.1416 (0.0088)	<0.0001
African American – Hospital-Level	0.3381 (0.0294)	<0.0001
Low SES census block group (AHRQ SES index, linked to 9-digit ZIP – Adjusted for Cost of Living) – Patient-Level	0.0524 (0.0057)	<0.0001
Low SES census block group (AHRQ SES index, linked to 9-digit ZIP – Adjusted for Cost of Living) – Hospital-Level	0.1784 (0.0191)	<0.0001
Renal Failure – Patient-Level	0.1142 (0.0055)	<0.0001
Renal Failure – Hospital-Level	0.5530 (0.0500)	<0.0001
Metastatic Cancer – Patient-Level	0.1548 (0.0108)	<0.0001

Metastatic Cancer – Hospital-Level	0.6319 (0.1000)	<0.0001
Lung Cancer – Patient-Level	0.1519 (0.0092)	<0.0001
Lung Cancer – Hospital-Level	0.8398 (0.1170)	<0.0001
COPD – Patient-Level	0.1599 (0.0050)	<0.0001
COPD – Hospital-Level	0.1198 (0.0374)	0.0013

Change of Predicted Probabilities for SES and Race Compared with Clinical Variables in the PN Readmission Measure (July 2011-June 2014)



*Low SES (ZIP9/Adj) measured by linking patients' 9-digit ZIP codes to AHRQ SES Index at the census block group level, adjusted for cost of living

Predicted Probabilities for SES and Race Variables in the PN Readmission Measure (July 2011-June 2014)

Hospital SES/Race Risk Factor Percentile	Dual Eligibility				Race				Low SES census block group (AHRQ SES index, linked to 9-digit ZIP – Adjusted for Cost of Living)			
	Var_.J bar	Var_ij=0 for all patients	Var_ij=1 for all patients	Delta (Patient Effect)	Var_.J bar	Var_ij=0 for all patients	Var_ij=1 for all patients	Delta (Patient Effect)	Var_.J bar	Var_ij=0 for all patients	Var_ij=1 for all patients	Delta (Patient Effect)
0%	0.0000	0.1673	0.1744	0.0071	0.0000	0.1702	0.1904	0.0202	0.0000	0.1683	0.1755	0.0072
5%	0.0513	0.1692	0.1764	0.0071	0.0000	0.1702	0.1904	0.0202	0.0000	0.1683	0.1755	0.0072
10%	0.0714	0.1700	0.1772	0.0072	0.0000	0.1702	0.1904	0.0202	0.0223	0.1689	0.1761	0.0072
25%	0.1124	0.1716	0.1788	0.0072	0.0000	0.1702	0.1904	0.0202	0.0783	0.1702	0.1775	0.0073
50%	0.1747	0.1740	0.1812	0.0073	0.0188	0.1711	0.1913	0.0202	0.1905	0.1730	0.1803	0.0073
75%	0.2516	0.1770	0.1843	0.0074	0.0830	0.1741	0.1946	0.0205	0.3677	0.1774	0.1848	0.0075
90%	0.3545	0.1810	0.1885	0.0075	0.2120	0.1802	0.2012	0.0210	0.5831	0.1828	0.1905	0.0077
95%	0.4205	0.1837	0.1912	0.0076	0.3285	0.1859	0.2074	0.0215	0.7026	0.1859	0.1937	0.0078
100%	0.8788	0.2028	0.2110	0.0081	1.0000	0.2212	0.2455	0.0243	1.0000	0.1938	0.2018	0.0080
P95 – P5 (Hospital Effect)	-	0.0144	0.0149	-	-	0.0157	0.0170	-	-	0.0176	0.0182	-

Predicted Probabilities for Clinical Variables in the PN Readmission Measure (July 2011-June 2014)

Hospital SES/Race Risk Factor Percentile	Renal Failure				Lung Cancer				Chronic Obstructive Pulmonary Disease			
	Var_.J bar	Var_ij=0 for all patients	Var_ij=1 for all patients	Delta (Patient Effect)	Var_.J bar	Var_ij=0 for all patients	Var_ij=1 for all patients	Delta (Patient Effect)	Var_.J bar	Var_ij=0 for all patients	Var_ij=1 for all patients	Delta (Patient Effect)
0%	0.0000	0.1468	0.1612	0.0144	0.0000	0.1662	0.1875	0.0213	0.1000	0.1562	0.1777	0.0215
5%	0.1429	0.1567	0.1718	0.0152	0.0063	0.1669	0.1883	0.0214	0.3636	0.1602	0.1822	0.0219
10%	0.1756	0.1590	0.1744	0.0154	0.0198	0.1684	0.1899	0.0215	0.4000	0.1608	0.1828	0.0220
25%	0.2335	0.1632	0.1789	0.0157	0.0370	0.1704	0.1921	0.0217	0.4649	0.1618	0.1839	0.0221
50%	0.2941	0.1677	0.1837	0.0160	0.0570	0.1727	0.1946	0.0219	0.5336	0.1629	0.1851	0.0222
75%	0.3435	0.1714	0.1877	0.0163	0.0740	0.1746	0.1968	0.0221	0.6016	0.1640	0.1863	0.0223
90%	0.3877	0.1748	0.1913	0.0165	0.0909	0.1766	0.1989	0.0223	0.6625	0.1650	0.1874	0.0224
95%	0.4203	0.1773	0.1940	0.0167	0.1027	0.1780	0.2004	0.0224	0.7076	0.1657	0.1882	0.0225
100%	0.6176	0.1932	0.2110	0.0178	0.5375	0.2353	0.2624	0.0272	0.9074	0.1689	0.1917	0.0228
P95 – P5 (Hospital Effect)	-	0.0207	0.0222	-	-	0.0111	0.0122	-	-	0.0054	0.0060	-

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

Discrimination Statistics
(1) Area under the receiver operating characteristic (ROC) curve (the c-statistic) is the probability that predicting the outcome is better than chance, which is a measure of how accurately a statistical model is able to distinguish between a patient with and without an outcome)
(2) Predictive ability (discrimination in predictive ability measures the ability to distinguish high-risk subjects from low-risk subjects; therefore, we would hope to see a wide range between the lowest decile and highest decile.2)

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

We tested the performance of the model for **Dataset 1** described in section 1.7.

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

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 expanded measure cohort version 8.2 (**Dataset 1**) the results are summarized below:

c-statistic = 0.63

Predictive ability (lowest decile %, highest decile %) = (9.3, 32.7)

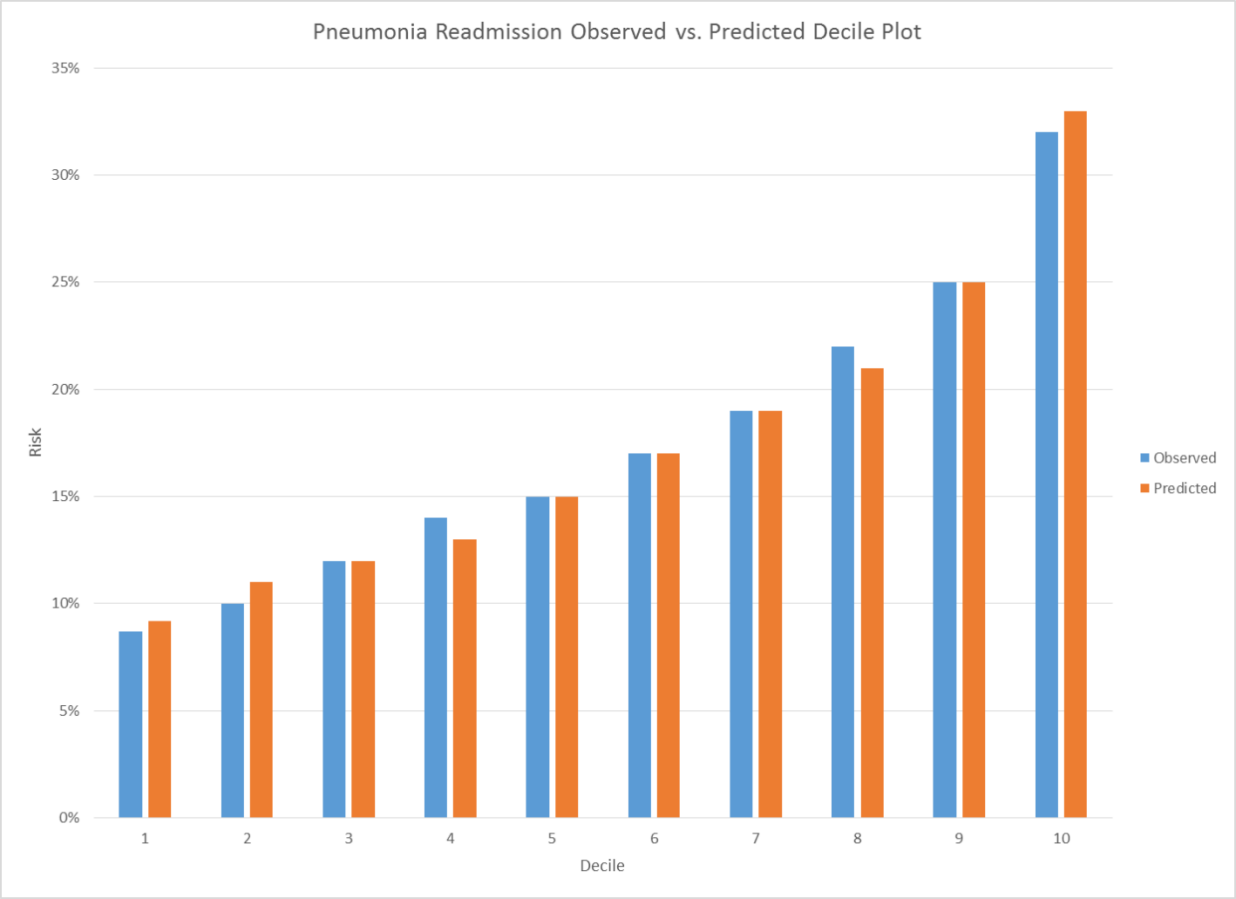
For comparison of model with and without inclusion of SES and race factors, see above section.

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

For the expanded cohort (**Dataset 1**) the results are summarized below:
1st half of split sample: Calibration: (0.0230, 0.9911)
2nd half of split sample: Calibration: (0.0231, 0.9900)

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. Below, we present the risk decile plot showing the distributions for Medicare FFS data from July 2011 to June 2014 (**Dataset 1**).



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-statistics of 0.63 indicate fair model discrimination (**Dataset 1**). 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 one, there is potential evidence of over-fitting. The calibration value of close to 0 at one end and close to 1 to the other end indicates 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 good 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*)

This measure is fully risk-adjusted using a hierarchical logistic regression model to calculate hospital RSRRs accounting for differences in hospital case-mix.

Application to Patients Aged 18 and Older (**Dataset 2**)
We applied the model to all-payer data from California. The analytic sample included 69,247 cases aged 18 and older in the 2009 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.

To help determine whether the measure could be applied to a population of patients aged 18+, we examined the interaction terms between age (18-64 vs. 65+) and each of the other risk factors. Specifically, we fit the model in all patients 18+ with and without interaction terms and (a) conducted a reclassification analysis to compare risk prediction at the patient level; (b) compared the c-statistic; and (c) compared hospital-level risk-standardized rates (scatterplot, correlation coefficient, and R2) to assess whether the model with interactions is different from the current model in profiling hospital rates.

When the model was applied to all patients 18 and over (18+), overall discrimination was good (c-statistic=0.658). 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, (a) the reclassification analysis demonstrated good patient-level risk prediction (5.1% to 29.3% vs. 6.0% to 28.4%, respectively, from the bottom decile to the top decile of the prediction values); (b) the c-statistic was nearly identical (0.665 vs. 0.658); and (c) hospital-level risk-standardized rates were highly correlated (r=0.995). Thus, the inclusion of the 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 and older.

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 (because it 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*)

Analyses of Medicare FFS data show substantial variation in RSRRs among hospitals. Using data from July 2011-June 2014 (**Dataset 1**), the median hospital RSRR was 17.5%, with a range of 13.1% to 24.7%. The interquartile range was 16.7%-18.4%.

Out of 4,700 hospitals in the U.S., 86 performed "better than the U.S. national rate," 4,061 performed "no different from the U.S. national rate," and 193 performed "worse than the U.S. national rate." 360 were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing.

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 the quality of care received across hospitals for pneumonia that support measurement to reduce the variation.

Note: The expansion of the cohort to include patients with a principal discharge diagnosis of sepsis who had pneumonia that was present on admission and patients with a principal discharge diagnosis of aspiration pneumonia (version 8.2), resulted in a modest increase in the readmission rate (17.6%) for the 3-year period that includes admissions from July 2011 through June 2014 when compared to the rate from the currently endorsed and publically reported measure, version 8.0 (17.0%). For the measure with the expanded cohort, between July 2011 and June 2014, the readmission has decreased from 18.2% in July 2011 to June 2012 to 17.2% in July 2013 to June 2014.

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 criterion is directed 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).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of 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 demonstrating comparability of 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.**

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)
	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 (Hospital IQR) Program, Centers for Medicare and Medicaid Services (CMS)

Purpose: The 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 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 Hospital 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 entities included in the program, as well as the number of patients included in the measure, varies by reporting year. For the data period between 2011-2014, the number of hospitals included in the measure with the expanded cohort was 4,700 and the number of admissions was 1,469,277.

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**

The median hospital 30-day, all-cause, RSRR for the re-specified pneumonia readmission measure with the expanded cohort (version 8.2) for the 3-year period between July 2011 and June 2014 was 17.5% (IQR 16.7% - 18.3%). The median RSRR decreased by 1.0 absolute percentage points from July 2011-June 2012 (median RSRR: 18.1%) to July 2013-June 2014 (median RSRR: 17.1%).

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. In response to research demonstrating changing coding patterns that could introduce bias in the measure, we did update the cohort as described in this Submission form. We are committed to ongoing monitoring of 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)**
0231 : Pneumonia Mortality Rate (IQI #20)
0279 : Bacterial Pneumonia Admission Rate (PQI 11)
0468 : Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following pneumonia hospitalization
0708 : Proportion of Patients with Pneumonia that have a Potentially Avoidable Complication (during the episode time window)
1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)
2579 : Hospital-level, risk-standardized payment associated with a 30-day episode of care for pneumonia

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?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
We did not include in our list of related measures any non-outcome (e.g., process) measures with the same target population as our measure. Because this is an outcome measure, clinical coherence of the cohort takes precedence over alignment with related non-outcome measures. Furthermore, non-outcome measures are limited due to broader patient exclusions. This is because they typically only include a specific subset of patients who are eligible for that measure (for example, patients who receive a specific medication or undergo a specific procedure).

5b. Competing Measures

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

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
N/A

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** 2015_Measures_Reevaluation_Condition-Specific_Readmission_AUS_Report_FINAL_508_Compliant-635895832865227558.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.
The working group involved in the initial measure development is detailed in the original technical report available at www.qualitynet.org.

Our measure development team consisted of the following members:

Harlan M. Krumholz, MD, SM
Sharon-Lise T. Normand, PhD*
Patricia S. Keenan, PhD, MHS
Mayur M. Desai, PhD, MPH
Zhenqiu Lin, PhD
Elizabeth E. Drye, MD, SM
Kanchana R. Bhat, MPH
Geoffrey C. Schreiner, BS

*Harvard Medical School, Department of Health Care Policy

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2008
Ad.3 **Month and Year of most recent revision:** 07, 2015
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?** 07, 2017

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

Ad.8 **Additional Information/Comments:** N/A

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 #: 1789
De.2. **Measure Title:** Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)
Co.1.1. **Measure Steward:** Centers for Medicare & Medicaid Services (CMS)
De.3. **Brief Description of Measure:** The measure estimates a hospital-level risk-standardized readmission rate (RSRR) of unplanned, all-cause readmission after admission for any eligible condition within 30 days of hospital discharge. The measure reports a single summary RSRR, derived from the volume-weighted results of five different models, one for each of the following specialty cohorts based on groups of discharge condition categories or procedure categories: surgery/gynecology; general medicine; cardiorespiratory; cardiovascular; and neurology, each of which will be described in greater detail below. The measure also indicates the hospital-level standardized risk ratios (SRR) for each of these five specialty cohorts. The outcome is defined as unplanned readmission for any cause within 30 days of the discharge date for the index admission (the admission included in the measure cohort). A specified set of planned readmissions do not count in the readmission outcome. CMS annually reports the measure for patients who are 65 years or older, are enrolled in fee-for-service (FFS) Medicare, and hospitalized in non-federal hospitals.
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 all-cause unplanned readmission rates among Medicare beneficiaries 65 yars and older admitted to all non-federal US acute care hospitals. 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 their patient case mix and hospital service mix, and therefore promote hospital quality improvement and better inform consumers about care quality.

Hospital-wide readmission is a priority area for outcomes 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 readmission. We define readmission as an inpatient admission for any cause, with the exception of certain planned readmissions, within 30 days from the date of discharge from an eligible index admission. If a patient has more than one unplanned admission (for any reason) within 30 days after discharge from the index admission, only one is counted as a readmission. The measure looks for a dichotomous yes or no outcome of whether each admitted patient has an unplanned readmission within 30 days. However, if the first readmission after discharge is considered planned, any subsequent unplanned readmission is not counted as an outcome for that index admission because the unplanned readmission could be related to care provided during the intervening planned readmission rather than during the index admission.
S.7. **Denominator Statement:** The measure includes admissions for Medicare beneficiaries who are 65 years and older and are discharged from all non-federal, acute care inpatient US hospitals (including territories) with a complete claims history for the 12 months prior to admission.

Additional details are provided in S.9 Denominator Details.
S.10. **Denominator Exclusions:** The measure excludes index admissions for patients:

1. Admitted to Prospective Payment System (PPS)-exempt cancer hospitals;

2. Without at least 30 days post-discharge enrollment in FFS Medicare;
3. Discharged against medical advice (AMA);
4. Admitted for primary psychiatric diagnoses;
5. Admitted for rehabilitation; or
6. Admitted for medical treatment of cancer.

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Administrative claims
S.26. **Level of Analysis:** Facility

IF Endorsement Maintenance – Original Endorsement Date: Apr 24, 2012 **Most Recent Endorsement Date:** Sep 13, 2012

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

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

De.4. **IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** N/A

Maintenance of Endorsement -- 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. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

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.

Summary of evidence provided by the developer:

- As a rationale for measuring this health outcome, the developer suggests that hospitals are able to influence readmission rates through a broad range of clinical activities including communication between providers, prevention of, and response to, complications, patient safety and coordinated transitions to the outpatient environment.
- The developer notes that there is no new evidence since the last review.

Question for the Committee:

- *Since there is no updated evidence, does the Committee agree the underlying rationale for the measure remains reasonable and there is no need for repeat discussion and vote on evidence?*

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.

- The developer provides performance data from three measurement periods, covering approximately 22,000,000 admissions.
- The data show that during the measurement period of 07/2013-06/2014, readmission rates ranged from a minimum of 11.4% to a maximum of 20.1%, with the 10th percentile at 14.6%, the 50th percentile at 15.4%, and the 90th

percentile at 16.5%.

Disparities

- To help in assessment of potential disparities, the developers also provided performance scores (using 2011-2014 data) for hospitals serving a low proportion of dual eligible patients vs. those serving a high proportion of dual eligible patients, performance scores for hospitals serving a low proportion of African-American patients vs. those serving a high proportion of African-American patients, and performance scores for hospitals serving a low proportion of patients with AHRQ SES Index Score index score equal to or below 45.9 vs. those serving a high proportion of patients with an AHRQ SES index score equal to or below 45.9.

- Hospitals serving a low proportion (=9.8%) Dual Eligible patients had a slightly lower median readmission rates (-0.3%) compared to hospitals serving a high proportion (=22.6%) Dual Eligible patients. Hospitals serving a low proportion (=2.2%) African-American patients had a slightly lower median readmissions rates (-0.3%) compared to hospitals serving a high proportion (=9.4%) African-American patients. Finally, hospitals serving a low proportion of patients below AHRQ SES index score of 45.0 had slightly lower median readmissions rates (-0.2%) compared to hospitals serving a high proportion of patients below AHRQ SRS index score of 45.0.

- By proportion of **Dual Eligible Patients:**

  // Low proportion (=9.8%) Dual Eligible patients//Hospitals with a high proportion (=22.6%) Dual Eligible patients
  Number of Measured Hospitals// 1,257 // 1,219
  Number of Patients// 2,137,895 patients in low-proportion hospitals // 927,007 in high-proportion hospitals
  Maximum// 18.7 // 20.1
  90th percentile// 16.2 // 16.8
  75th percentile// 15.7 // 16.0
  Median (50th percentile)// 15.3 // 15.6
  25th percentile// 14.8 // 15.2
  10th percentile// 14.3 // 14.9
  Minimum // 11.5 // 12.2

- By proportion of **African-American Patients**:

  // Low proportion (=2.2%) African-American patients//Hospitals with a high proportion (=9.4%) African-American patients
  Number of Measured Hospitals// 1,156 // 1,180
  Number of Patients// 222,648 patients in low-proportion hospitals/ 2,294,715 in high-proportion hospitals
  Maximum// 19.1 // 19.9
  90th percentile// 16.0 // 17.1
  75th percentile// 15.6 // 16.3
  Median (50th percentile)// 15.4 // 15.7
  25th percentile// 15.1 // 15.2
  10th percentile// 14.8 // 14.8
  Minimum // 12.9 // 12.2

- By Proportion of Patients with AHRQ SES Index Scores Equal or Below 45.9:

  // Low proportion of patients below AHRQ SES index score of 45.0 (=5.0%)// Hospitals with a high proportion of patients below AHRQ SES index score of 45.0 (=57.1%)
  Number of Measures Hospitals// 1,209 // 1,217
  Number of Patients// 1,651,852 patients in hospitals with low proportion of patients below AHRQ SES index score of 45.0 //795,899 patients in hospitals with high proportion of patients below AHRQ SES index score of 45.0

Maximum// 19.9 // 20.1
90th percentile// 16.2 // 16.6
75th percentile// 15.7 // 16.0
Median (50th percentile)// 15.3 // 15.5
25th percentile// 14.9 // 15.2
10th percentile// 14.5 // 14.8
Minimum // 11.5 // 13.0

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)

1. Importance to Measure and Report

1a. Evidence to Support Measure Focus

Comments: **No specific link with readmissions and specific actions to reduce readmissions. Although there are multiple ongoing efforts in response to readmission reduction programs, some evidence supporting links of those efforts to readmission rate reductions would have been informative. Further, is there a threshold below which further efforts to reduce rates are either ineffective or harmful?

**In the evidence criteria document, the developer identifies six healthcare actions that may lead to improved health status and a decreased risk of readmission. The six actions are:

- ensuring patients are clinically ready for discharge
- reducing the risk of infection
- reconciling medications
- improving communication among providers involved in transition of care
- encouraging strategies that promote disease management principles
- educating patients about symptoms, who to contact with questions and when to seek follow-up care

The rationale cites several studies where the stated healthcare actions were shown to influence readmission rates.

1b. Performance Gap

Comments: **The distance between the 10th and 90th percentile in RSRR is 1.9%. The major gap appears to be at the extremes which would raise concerns about outliers in the distribution. The impact of proportion of dual eligibles, African-American patients and AHRQ SES, scores greater than or equal to 45 appear to be relatively negligible. However, in light of the penalties currently being assigned by CMS to hospitals with greater than expected readmission rates, it would be useful to know whether the observed differences in these subgroups would have led to a change in the proportion receiving penalties.

**Performance data for the measure was provided. Using data spanning three measurement periods and totaling 22M admissions, the developer found a range in outcomes of 11.4% to 20.1%.

Dual-eligibles, african americans and SES were analyzed as sub-groups looking for potential disparities. Small lower readmissions rates were found in hospitals having lower proportions of patients falling into each sub-group were found.

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

Comments: **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.

- This measure calculates 30-day readmissions for patients with an eligible index admission.
- The measure produces a risk-standardized readmission rate (RSRR), which is calculated as the ratio of the number of "predicted" to the number of "expected" readmission at a given hospital, multiplied by the national observed readmission rate.
- The denominator includes Medicare beneficiaries who are 65 years and older and are discharged from all non-federal, acute care inpatient US hospitals (including territories) with a complete claims history for the 12 months prior to admission.
- The numerator includes patients were readmitted to any acute care hospital for any cause within 30 days of the date of discharge of the index admission, excluding planned readmissions
- The denominator population is defined using ICD-9 and ICD-10 codes; a list of applicable codes is included in the submission.
- The data sources for this measure may include Medicare Part A and B claims and the Medicare Enrollment Database (EDB).
- This measure was developed with 12 months of data and is currently publicly reported with one year of data.
- The measure is risk-adjusted using a statistical risk model (see details below).

Questions for the Committee :

- *Specific questions on the specifications, codes, definitions, etc.*
- *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.

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

- Datasets used for testing included Medicare Parts A and B claims and the Medicare Enrollment Database (EDB). Additionally, census data were used to assess socio-demographic factors.
- **Data element reliability**:
  - With regard to data element reliability, the developer notes that the measure has been developed to avoid the use of claims data elements that are thought to be coded inconsistently across hospitals or providers, instead using fields that are consequential for payment and which are audited by CMS.
  - In addition, the developer compared frequencies and odds ratios of variables from their risk model across three years of data in order to assess the consistency of those variables over time.
- **Performance score reliability:**
  - The developer defines performance score reliability as the degree to which repeated measurements of the same entity agree with each other.
  - In line with this thinking, the developer's approach to assessing score-level reliability was to consider the extent to which assessments of a hospital using different but randomly-selected subsets of patients produce similar measures of hospital performance. The developers refer to this as a "test-retest" approach; it may also be called a "split-half" method. This is generally considered an appropriate method of testing reliability.

Results of reliability testing

- **Data element reliability**:
  - Summarizing the results of this analysis, the developer notes that the frequency of some data elements may increase or decrease slightly from year-to-year.
  - The developer notes these changes may reflect small changes in rates of comorbidity in the fee-for-service population.
- **Performance score reliability:**
  - A total of 6,843,808 admissions in the 2015 publicly reported measure, with 3,420,728 in one sample and 3,423,080 in the other randomly-selected sample. Two risk-standardized readmission rates (RSRR) were calculated for each hospital: one from each of the two separate samples.
    - The agreement between the two RSRRs for each hospital (as measured by an intra-class correlation coefficient (ICC)) was 0.80; the developer states that according to the conventional interpretation, this is considered a "substantial" level of agreement.

Guidance from the Reliability Algorithm

- Question 1. Submitted specifications are precise, unambiguous, and complete. Measure can be consistently implemented.
- Question 2. Empirical reliability testing was conducted using statistical tests with the measure as specified.
- Question 3. Empirical validity testing of patient-level data was conducted.
- Question 4. Reliability testing was conducted with computed performance measure scores for each measured entity.
- Question 5. Random split-half correlation was used to assess the proportion of variability due to real differences among the measured entities.
- Question 6. The ICC was 0.80 which is considered a substantial level of agreement.

Questions for the Committee:

- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

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.

- This measure estimates 30-day readmissions for any cause with the exclusion of certain planned readmissions.
- As a rationale for measuring this health outcome, the developers suggest that hospitals are able to influence readmission rates through a broad range of clinical activities, including ensuring patients are ready for discharge, reducing the risk of infection, reconciling medications, improving communication among providers, promoting disease management, and educating patients.

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.

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**

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

- The developer demonstrated measure validity through prior validity testing done on their other claims-based measures, through the use of established measure development guidelines, and examination of content validity by comparing hospital performance with that on other quality measures.
- Validity of Claims Data:
  - CMS validated six other NQF-endorsed readmission measures currently in public reporting (acute myocardial infarction [AMI], heart failure, and pneumonia mortality and readmission) with models that used chart-abstracted data for risk-adjustment.
  - 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.
- Validity Indicated by Established Measure Development Guidelines:
  - The developer states that this measure was developed in consultation with national guidelines, with outside experts, and with the public.
- Validation Against Other Outcomes Measures:
  - The developer examined whether hospitals that perform well according to other measures and ranking systems had lower hospital-wide risk-standardized readmission rates than remaining hospitals when applying our measure to the Medicare FFS population.
  - The developers found significant correlation between patient satisfaction and RSRR as measured by the HWR measure. "Top performers" as defined by Thomson Reuters have lower RSRRs as measured by the HWR measure. However, the developers also found that hospitals identified by The Joint Commission as having superior performance on all four categories of clinical process measures have identical performance as those with lower performance, consistent with published studies.

Table 1. Correlation between RSRR (2009 Medicare FFS data) and HCAHPS response (N=3,723 hospitals)

HCAHPS Question	Correlation
Pain was 'sometimes' or 'never' well controlled	0.34
Patients 'sometimes' or 'never' received help as soon as they wanted	0.34
Nurses 'sometimes' or 'never' communicated well	0.33
'NO' patients would not recommend the hospital	0.32
Patients were 'sometimes' or 'never' given information about what to do during their recovery at home	0.32
Patients who gave a rating of '6' or lower	0.31
Doctors 'sometimes' or 'never' communicated well	0.21

p value for all correlations <0.001REVISED Hospital-Wide Readmission NQF Application January 5, 2012 32

2. Thomson Reuters Top 100 Hospitals

Table 2. shows the RSRRs distribution for the top performers in comparison to the rest of hospitals.

Table 2. Distribution of RSRRs (2009 Medicare FFS data) for the Thomson Reuters Top 100 Hospitals

vs. others

	On List	Not On List
Number	100	3017
Mean (SD)	16.19 (1.39)	16.65 (1.28)
Minimum	13.77	12.51
Lower Quartile	15.05	15.79
Median	16.06	16.51
Upper Quartile	16.99	17.35
Maximum	19.81	22.69

3. The Joint Commission's Top Performers on Key Quality Measures program

Table 3. shows the distribution of risk-standardized readmission rates of the 158 top performers compared to other hospitals.

Table 3. Distribution of RSRR (2009 Medicare FFS data) for The Joint Commission's Top Performers vs. Others

	On List	Not On List
Number	158	4630
Mean (SD)	16.66 (0.99)	16.61 (1.16)
Minimum	14.18	12.51
Lower Quartile	16.01	15.87
Median	16.64	16.49
Upper Quartile	17.17	17.21
Maximum	19.91	22.69

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?*
- *Other specific question of the validity testing?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- Patients in the following categories are excluded from the measure:
  - Admitted to Prospective Payment System (PPS)-exempt cancer hospitals;
  - Without at least 30 days post-discharge enrollment in FFS Medicare;
  - Discharged against medical advice (AMA);
  - Admitted for primary psychiatric diagnoses;
  - Admitted for rehabilitation; or
  - Admitted for medical treatment of cancer.
- To determine the impact of exclusions, the developer examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion.
- The number and percentage of patients excluded for each criterion are as follows:
  - Admitted to PPS-exempt cancer hospitals; **19,823 (0.28%)**
  - Without at least 30 days post-discharge enrollment in FFS Medicare; **36,640 (0.52%)**
  - Discharged against medical advice (AMA); **26,665 (0.38%)**
  - Admitted for primary psychiatric diagnoses; **19,691 (0.28%)**
  - Admitted for rehabilitation; or 7,512 (.10%)
- The developer also provides the distribution across hospitals for each exclusion criterion.

Questions for the Committee:

- *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: **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 measure employs a hierarchical logistic regression model (a form of hierarchical generalized linear model [HGLM]) to create a hospital-level 30-day risk-standardized readmission rate (RSRR).
- The developer notes that this approach simultaneously models data at the patient and hospital levels to account for the variance in patient outcomes both within and between hospitals.
- Variables considered for inclusion in the model were developed from a "starter" set of variables drawn from previous readmission measures (AMI, heart failure, pneumonia, hip and knee arthroplasty, and stroke). The developer then reviewed all remaining CMS-CCs and determined on a clinical basis whether they were likely to be relevant to an all-condition measure.
- For each patient, covariates were obtained from Medicare claims extending 12 months prior to and including the index admission.
- The measure does not adjust for CCs that were possible adverse events of care and that were only recorded in the index admission.
- The final set of 33 risk-adjustment variables is included in the testing attachment; the odds ratio associated with each variable is also provided.
- The developers also considered a number of variables related to sociodemographic status (SDS) for potential inclusion in the risk-adjustment model. Candidate SDS variables were selected for examination based on a review of literature and national data sources.
- **Conceptual analysis of the need for SDS adjustment:**
    - The developers note there is a large body of literature linking various SES factors and African-American race to worse health status and higher readmission risk with income, education, and occupational level being the most commonly examined variables. The developers state that the literature directly examining how SES factors or race might influence the likelihood of older, insured, Medicare patient of being readmitted within 30 days of an admission for heart failure is more limited.
    - The developers state that few studies directly address causal pathways for SDS factors to affect 30-day readmission rates or examine the role of the hospital in these pathways.
    - There are at least four potential pathways for SDS factors to affect 30-day readmission rates:
        - One potential pathway is the relationship to health status at the time of admission. SDS factors 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. The developers note that this pathway should be largely accounted for by their clinical risk-adjustment model.
        - The next potential path way is that patients with low income and African-American patient are more likely to be seen in lower quality hospitals, which can contribute to increased risk of readmission.
        - The third major pathway is that a patient's race or SDS status cause them to experience differential, lower quality care or may not receive the differentiated care they require.

        - Finally, some SES risk factors may affect the likelihood of readmission without directly affecting health status at admission or the quality of care received during the hospitalization. Patients may have worse outcomes due to competing economic priorities or a lack of access to care outside the hospital.

- **Empirical analysis of SDS factors:**
    - The developers considered African-American race, dual-eligible status-i.e. enrolled in both Medicare and Medicaid, and AHRQ SES index score. The developers assessed the relationship between the SES variables and race with the outcome and examined the incremental effect in a multivariable mode.
    - The developer stated that they examined all patient-level indicators of both SES and race/ethnicity that are reliably available for all Medicare beneficiaries and linkable to claims data and selected those that are most valid.
    - The developer noted that the AHRQ-validated SES index score is a widely-used variable that describes the average socioeconomic status of people living in defined geographic areas. The developer notes that its value as a proxy for patient-level SDS is depend on having the most granular level data.
        - These variables are linked to patients by zip code and census block; however, the data are only linked at a 5-digit zip code level—nine-digit zip code data, which may provide a more granular view of patient sociodemographic status, were not available.
        - However, the developers note they are currently performing analyses at the census block level (the most granular level possible in this dataset) and hope to present the results of this analysis to the committee.
    - The developer assessed the relationship between the SDS variables and the 30-day readmission rate and examined the incremental effect of SDS in a multivariable model, evaluating the extent to which the addition of any one of these variables improved model performance or changed hospital results.
    - The developer notes that 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. Therefore, the developers performed a decomposition analysis to assess the independent effects of the SES and race variables at the patient level and the hospital level.
    - The developers' analysis found that the prevalence of SDS factors in the hospital-wide readmission cohort does vary across measured entities.
    - With regard to the empirical association of each SDS variable with the outcome (bivariate), the analysis found that patient-level observed hospital-wide readmission rate for Medicaid patients was higher, at 19.3% compared with 14.8% for all other patients. The readmission rate for African-American patients was also higher at 19.2% compared with 15.1% for patients of all other races. Similarly the readmission rate for patients in the lowest SES quartile by AHRQ index was 16.8% compared with 15.1% for all other patients.
    - With regard to the strength and significance of the SDS variables in the context of a multivariable model, the developers' analysis found that the effect size of each of these variables is small, the c-statistic (i.e., predictive value) is unchanged with the addition of any of these variables into the model, and the addition of any of these variables into the model has little to no effect on hospital performance.
        - The median absolute change in hospitals' RSRRs when adding a Medicaid indicator is 0.004% (interquartile range [IQR] -0.017% – 0.024%, minimum -0.309% – maximum 0.135%) with a correlation coefficient between RSRRs for each hospital with and without Medicaid added of 0.998.
        - The median absolute change in hospitals' RSRRs when adding a race indicator is 0.011% (IQR -0.010% – 0.033%, minimum -0.671% – maximum 0.130%) with a correlation coefficient between

RSRRs for each hospital with and without race added of 0.998.
        - The median absolute change in hospitals' RSRRs when adding a low SES AHRQ indicator is 0.007% (IQR -0.033% – 0.036%, minimum -0.322% – maximum 0.135%) with a correlation coefficient between RSRRs for each hospital with and without low SES added of 0.997.
    - The developers state that the patient-level and hospital-level dual eligible, race, and low AHRQ SES Index effects were significantly associated with hospital-wide readmission in the decomposition analysis. The developers note that if the dual eligible, race, or low AHRQ SES Index variables are 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 developers state that given these findings and complex pathways that could explain any relationship between SDS and readmission, they did not incorporate SDS variables into the measure.
- **Risk Model Diagnostics:**
    - To assess the overall performance of their risk-adjustment model, the developers computed three summary statistics, including:
        - *Area under the receiver operating characteristic (ROC) curve* (also known as a c-statistic, which measures the probability that the model's prediction of the outcome is better than chance)
        - *Predictive ability* (the model's ability to distinguish high-risk subjects from low-risk subjects)
        - *Over-fitting indices (model calibration)* (to ensure that the model is not only describing the relationship between predictive variables and outcome in the development dataset but also providing valid predictions in new patients)
    - For the current measure cohort, the findings from this analysis are as follows:
        - C-statistic:
            - Medicine cohort: 0.643
            - Surgical cohort: 0.675
            - Cardiorespiratory cohort: 0.636
            - Cardiovascular cohort: 0.658
            - Neurology cohort: 0.622
            - The developers state the c-statistics indicate fair model discrimination for each of the models.
        - Predictive ability (lowest decile %, highest decile %):
            - Medicine cohort: 9%-33%
            - Surgical cohort: 5%-27%
            - Cardiorespiratory cohort: 10%-35%
            - Cardiovascular cohort: 7%-31%
            - Neurology cohort: 8%-26%
            - The developers state that this indicates a wide range between the lowest decile and highest decile, indicating the ability to distinguish high-risk subjects from low-risk subjects.
        - Overfitting indices (model calibration) [presented as (γ0, γ1)]:
            - The developer states that if the γ0 in the validation samples are substantially far from zero and the γ1 is substantially far from one, there is potential evidence of over-fitting. The calibration value of close to 0 at one end and close to 1 to the other end indicates good calibration of the model.
                - Medicine cohort: (0.132, 1.118)
                - Surgical cohort: (0.104, 1.076)
                - Cardiorespiratory cohort: (0.193, 1.184)
                - Cardiovascular cohort: (0.145, 1.109)
                - Neurology cohort: (0.201, 1.163)
                - The developer's overall interpretation of the results of their analysis is that the findings demonstrate the risk-adjustment model adequately controls for

differences in patient characteristics (case mix).

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 rationale that there is no conceptual basis for adjusting this measure for SDS factors?*
- *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)*:

- For public reporting of this measure, CMS characterizes the uncertainty associated with the RSMR by estimating the 95% interval estimate.
- If the RSMR'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 RSMR as "no different than the U.S. national rate" or "the difference is uncertain."
- In the 2015 public reporting year, out of 4,772 hospitals in the U.S., 178 performed "better than the U.S. national rate," 4,078 performed "no different from the U.S. national rate," 337 performed "worse than the U.S. national rate," and1779 were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing.
- The developer's interpretation of this data is that the variation in rates and number of performance outliers suggests there remain differences in the quality of care received across hospitals for HF that support measurement to reduce the variation.

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

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)

2. Scientific Acceptability of Measure Properties

2a1. & 2b1. Specifications

Comments: **Conceptual rationale for use as quality measure was mutability in response to QI efforts (e.g. discharge planning, medication reconciliation, etc.)

**As this is a maintenance measure, all elements are clearly defined. The complicated nature of the calculation of this measure may

force many hospitals to rely only on CMS for results, vs gathering data and calculating results more frequently. This may impact improvement in poorer performing hospitals. I noted no inconsistencies between specification and evidence.

2a2. Reliability Testing

Comments: **Data element reliability - developer compared frequencies and ORs of variables in risk models across 3 years to assess consistency over time.

Performance score reliability was assessed using split half reliability on two separate patient samples. ICC agreement = .80.

**Yes, the developer tested with adequate scope. The measure was tested using medicare part A claims for one calendar year (7/1/13 - 6/30/14). This sample included 4,772 hospitals and over 6.8M admissions.

2b2. Validity Testing

Comments: **Compared RSRRs to other measures of quality (e.g. HCAHPS, Reuters Top Performers) - found associations in expected direction; no differences in RSRRs for hospitals classified as top performers on Joint Commission measures. Unclear whether these findings constitute evidence of validity.

**Because the developer used Medicare claims data they were able to use data elements that are "consequential for payment and audited". In their testing description, the developer cited an example of discharge disposition as being widely understood to be an unreliable field in claims data. Therefore, the developer uses other fields within claims to derive the most reliable data from them. The developer described a test/re-test approach to validity in their submitted documentation. The agreement between RSRRs in two samples from dataset 1 was .80, which they deemed to be "substantial".

I agree that this measure is an indicator of quality.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **2b3. Exclusions appear not to substantively affect conclusions about generalizing of measure.

2b4. Conceptual model linking SDS to readmission provided and supported, however the developers noted that the effect size for the variables used (dual eligible patients, African-American patients and AHRQ SES) was small, the c-statistic was unchanged and the inclusion of these variables did not meaningfully change hospitals RSRR. While other variables may better represent patient mix, the measures available to the developer for this analysis did not appear to affect hospital comparisons.

2b5. Of 4772 hospitals included in the analysis, 515 (10.8%) performed better (3.7%) or worse (7.1%) than expected. While this proportion suggests opportunity for some improvement, the absolute differences defining "better" or "worse" than expected may be small.

**no

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 based on administrative claims data (e.g., DRG, ICD-9/10), which the developers note are routinely generated and collected as part of hospitals' billing processes.
- The developer indicates that all data elements are in defined fields in electronic claims.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Is the data collection strategy ready to be put into operational use?*

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

Committee pre-evaluation comments
Criteria 3: Feasibility

3. Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **The measure uses administrative claims data already in existence and has been collected for several years.
**I have no concerns about putting the measure into operational use, given that it has already been approved and leverages administrative claims 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**
OR
Planned use in an accountability program? ☐ **Yes** ☐ **No**

Accountability program details
- Hospital Inpatient Quality Reporting (IQR) Program http://cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html

Improvement results
- The developer reports: "there has been significant progress in 30-day RSRR for unplanned, all-cause readmissions. The median 30-day RSRR decreased by 0.7 absolute percentage points from the 2013 public reporting period (median RSRR: 15.9%) to the 2015 public reporting period (median RSRR: 15.2%)."

Unexpected findings (positive or negative) during implementation
- The developer noted that there are no unexpected findings to report.

Potential harms
- The developer noted that there were no unintended consequences during development, testing or re-specification. They are committed to ongoing monitoring of potential unintended consequences over time.

Feedback
- During the 2012-2013 MAP review, MAP supported this measure for inclusion in the IQR and PQRS programs. The group agreed that the new specifications are an improvement over the existing finalized measure.

***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

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **The developers note a decrease of 0.7% since the 2013 public reporting period as evidence for usability. There were no unintended consequences reported.

**The measure is used in the IQR program.

The developer noted a .7 percentage point decrease in the all-cause 30-day RSRR from the 2013 to the 2015 reporting period.

The results of this measure could be further analyzed to provide insight into disparities in care-- as noted in this analysis there may be opportunities to improve care among dual-eligibles, african-americans, and patients with a low SES score (45.9 or lower for the AHRQ SES score).

Criterion 5: Related and Competing Measures

Related or competing measures

- 1768: Plan All-Cause Readmissions (PCR)

Harmonization

- This measure and the NCQA Plan All-Cause Readmissions (PCR) Measure #1768 are related measures, but are not competing because they don't have the same measure focus and same target population. Each of these measures has different specifications. In addition, both have been previously harmonized to the extent possible under the guidance of the National Quality Forum Steering Committee in 2011.

Pre-meeting public and member comments

Comment by: Ms. Elizabeth Godsey

Organization: Vizient, Inc.

Comment May 05, 2016: Vizient, Inc., the largest member-owned health care company in the country, is dedicated to serving members & customers through innovative data-driven solutions, expertise & collaborative opportunities that lead to improved patient outcomes & lower costs. Vizient requests CMS to review & provide follow-up analysis on more applied/practical alternate modeling approaches to account for within & across hospital variation besides hierarchical modeling. While hierarchical modeling is a valid technique controlling for within & across hospital variation, the approach lacks a tangible, practical framework of an observed to expected ratio that hospitals need to drive patient care. The predicted to expected approach complicates the public's & provider's understanding of how the actual observed values impacts hospital performance. Through numerous member discussions, we heard repeatedly, Oh, you mean that number does really reflect my actual readmissions? How can I improve that number? Even more concerning is the focus the current measure places on improving documentation & coding rather than patient care. Currently, providers see the only direct way to improve the measure is through documentation & coding capture of co-morbidities which count toward the predicted & expected value calculations. We hope this was not the original intention of the measure & this misguided focus is simply an unintended artifact of an overly complicated modeling technique. We recommend analyzing & provide results comparing a model that uses hospital characteristics, such as teaching status or bed size to account for structural differences across hospitals & provide an observed to expected ratio which is much more meaningful for the public & providers. While in the past, CMS has commented they would not incorporate these features due to NQF restrictions; it is important to point out NQF has endorsed other risk adjustment models that incorporate these characteristics (NHSN) & consider these factors in the 30-day risk adjustment as well. Also, we would ask CMS & NQF to institute discrimination performance thresholds for the models given the importance these models bare on CMS's performance programs & public reporting. Currently, no model performs > 0.70, a standard considered fair-good practical performance threshold & while the c-stat does not fully evaluate the model, it certainly should require basic

performance standards. Additionally, we ask CMS to provide performance statistics, like AIC, BIC & the Somers' D, Gamma & Tau-a association of predicted probabilities & observed counts for a more comprehensive assessment. Using these standards & model diagnostics, NQF can provide CMS with recommendations for improvement. Until minimum discrimination thresholds are instituted, we recommend NQF remove endorsement of the readmission measures.

Comment by: Ms. Elizabeth Godsey
Organization: Vizient, Inc.
Comment May 05, 2016: Vizient, Inc., the largest member-owned health care company in the country, is dedicated to serving members & customers through innovative data-driven solutions, expertise & collaborative opportunities that lead to improved patient outcomes & lower costs. For the readmission measures considered, CMS presented patient-level & hospital specific SES factor beta coefficients & p-values, yet overall model performance were not presented. We request the actual model performance results for model evaluation. For the AHRQ SES Index variable, we request further information on how the binary classification for a measure that ranges between 0-100 was determined & the impact of transforming into a binary representation vs. actual value had on the model performance. This detail along with the overall model performance information would provide the public with the necessary information to truly assess CMS's comment 'Given these findings & 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.' Regarding the complex pathways associated with 30-day readmissions as stated by CMS, we strongly ask CMS to entirely re-evaluate the utility of the 30-day measures. As stated by CMS, factors influencing readmissions are blurred between providers & patients 30-days post discharge resulting in a limited insights in how providers can improve care. We believe CMS's efforts to remove the planned readmissions PR4 logic is a strong step in true opportunity identification; however, more refinement is needed. We recommend a shorter, more actionable 7 day post-discharge readmission timeframe to pinpoint opportunities providers truly can influence & thus, mitigate many of SES confounding factors. The 7-day window provides clearer opportunities for patient stabilization & post-acute discharge planning which the 30-day window doesn't reflect. We recommend CMS provide a 7-day readmission risk adjustment for review. Also, the hospital wide readmission measure evaluates all readmissions within the 30-day window post inpatient discharge & considers readmit cases to also be eligible as the index admission; however, the condition specific measures evaluate only 1 readmit within the 30-day window & cannot be eligible as an index. We ask CMS for the rationale why the different approaches for the same measure as this adds unnecessary complexity which are impractical to manage. We recommend a consistent approach across all readmission measure calculations & recommend evaluating & counting all readmits that occur within the 30-day window so providers have a clear understanding of the # readmits are truly occurring. We support considering a readmit as an index for the next 30-day cycle to again, assist organizations in tracking & improving complete patient care.

Comment by: Ms. Elizabeth Godsey
Organization: Vizient, Inc.
Comment May 05, 2016: Vizient, Inc., the largest member-owned health care company in the country, is dedicated to serving members & customers through innovative data-driven solutions, expertise & collaborative opportunities that lead to improved patient outcomes & lower costs. Vizient agrees with the minor changes to clinical cohorts due to changes in CCS procedure codes categories, specifically those for obstetrics procedures which are now in many different CCS categories. Vizient recommends CMS and NQF reviewing this criterion and provide the appropriate ICD-10 translations to address the debridement of wound; infection or burn procedure codes for ccs 169 impacting the planned procedure exclusion criteria.

Comment by: Ms. Elizabeth Godsey
Organization: Vizient, Inc.
Comment May 05, 2016: Vizient, Inc., the largest member-owned health care company in the country, is dedicated to serving members & customers through innovative data-driven solutions, expertise & collaborative opportunities that lead to improved patient outcomes & lower costs. Vizient agrees with the minor changes to clinical cohorts due to changes in CCS procedure codes categories, specifically those for obstetrics procedures which are now in many different CCS categories. Vizient recommends CMS and NQF reviewing this criterion and provide the appropriate ICD-10 translations to address the debridement of wound; infection or burn procedure codes for ccs 169 impacting the planned procedure exclusion criteria.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 1789

Measure Title: Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: N/A

Date of Submission: 1/29/2016

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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, hospital-wide, all-cause, unplanned readmission

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

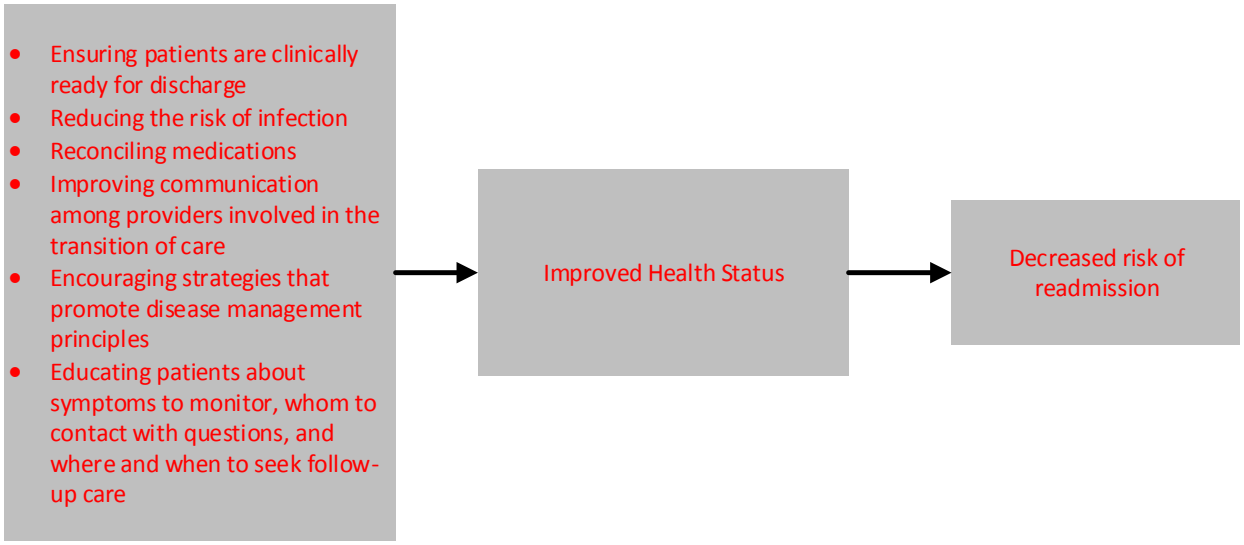
☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.



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 of unplanned, all-cause readmission after admission for any eligible condition within 30 days of hospital discharge. 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.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

The diagram above indicates some of the many care processes that can influence readmission risk. In general, randomized controlled trials have shown that improvement in the following areas can directly reduce readmission rates: quality of care during the initial admission; improvement in communication with patients, their caregivers, and their clinicians; patient education; predischarge assessment; and coordination of care after discharge. Evidence that hospitals have been able to reduce readmission rates through these quality-of-care initiatives illustrates the degree to which hospital practices can affect readmission rates. Successful randomized trials have reduced 30-day readmission rates by 20-40% [1-11]. Since 2008, 14 Medicare Quality Improvement Organizations have been funded to focus on care transitions, applying lessons learned from clinical trials. Several have been notably successful in reducing readmissions. The strongest evidence supporting the efficacy of improved discharge processes and enhanced care at transitions is a randomized controlled trial by the Project RED (Re-Engineered Discharge) intervention, in which a nurse was assigned to each patient as a discharge advocate, responsible for patient education, follow-up, medication reconciliation, and preparing individualized discharge instructions sent to the patient's primary care provider and there was a follow-up phone call from a pharmacist within 4 days of discharge demonstrated a 30% reduction in 30-day readmissions [1]. Hospital processes that reflect the quality of inpatient and outpatient care such as discharge planning, medication reconciliation, and coordination of outpatient care have been shown to reduce readmission rates [12]. Although readmission rates are also influenced by hospital system characteristics, such as the bed capacity of the local health care system, these hospital characteristics should not influence quality of care [13]. Therefore, this measure does not risk adjust for such hospital characteristics.

Studies have estimated the rate of preventable readmissions to be as low as 12% and as high as 76% [14, 15]. Given that studies have shown readmissions to be related to quality of care, and that interventions have been able to reduce 30-day readmission rates, it is reasonable to consider an all-condition readmission rate as a quality measure.

The hospital-wide risk-standardized readmission rate (RSRR) measure is thus intended to inform quality-of-care improvement efforts, as individual process-based performance measures cannot encompass all the complex and critical aspects of care within a hospital that contribute to patient outcomes. As a result, many stakeholders, including patient organizations, are interested in outcomes measures that allow patients and providers to assess relative outcomes performance for hospitals

References:

1. Jack BW, Chetty VK, Anthony D, Greenwald JL, Sanchez GM, Johnson AE, et al. A reengineered hospital discharge program to decrease rehospitalization: a randomized trial. Ann Intern Med 2009;150(3):178-87.

2. Coleman EA, Smith JD, Frank JC, Min SJ, Parry C, Kramer AM. Preparing patients and caregivers to participate in care delivered across settings: the Care Transitions Intervention. J Am Geriatr Soc 2004;52(11):1817-25.

3. 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 2009;57(3):395-402.

4. 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.

5. Koehler BE, Richter KM, Youngblood L, Cohen BA, Prengler ID, Cheng D, 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. J Hosp Med 2009;4(4):211-218.

6. Mistiaen P, Francke AL, Poot E. Interventions aimed at reducing problems in adult patients discharged from hospital to home: a systematic metareview. BMC Health Serv Res 2007;7:47.

7. 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 1994;120(12):999-1006.

8. Naylor MD, Brooten D, Campbell R, Jacobsen BS, Mezey MD, Pauly MV, et al. Comprehensive discharge planning and home follow-up of hospitalized elders: a randomized clinical trial. Jama 1999;281(7):613-20.

9. 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 2002;17(3):186-92.

10. Weiss M, Yakusheva O, Bobay K. Nurse and patient perceptions of discharge readiness in relation to postdischarge utilization. Med Care 2010;48(5):482-6.

11. 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.

12. Nelson EA, Maruish ME, Axler JL. Effects of Discharge Planning and Compliance With Outpatient Appointments on Readmission Rates. Psychiatr Serv. July 1 2000;51(7):885-889.

13. Fisher ES, Wennberg JE, Stukel TA, Sharp SM. Hospital Readmission Rates for Cohorts of Medicare Beneficiaries in Boston and New Haven. New England Journal of Medicine. 1994;331(15):989-995.

14. Benbassat J, Taragin M. Hospital readmissions as a measure of quality of health care: advantages and limitations. Archives of Internal Medicine 2000;160(8):1074-81.

15. Medicare Payment Advisory Commission (U.S.). Report to the Congress promoting greater efficiency in Medicare. Washington, DC: Medicare Payment Advisory Commission, 2007.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

N/A. This measure is not an intermediate outcome, process, or structure performance measure.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

N/A

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

N/A

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

N/A

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

N/A

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

N/A

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

N/A

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

N/A

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

N/A

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

N/A

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

N/A

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

N/A

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency

of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

N/A

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade: .

N/A

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

N/A

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

N/A

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

N/A

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

N/A

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

N/A

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

N/A

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

N/A

1a.8 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.8.1 What process was used to identify the evidence?

N/A

1a.8.2. Provide the citation and summary for each piece of evidence.

N/A

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

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all 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_1789_HWR_NQF_Evidence_Attachment_02-15-16_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 all-cause unplanned readmission rates among Medicare beneficiaries 65 yars and older admitted to all non-federal US acute care hospitals. 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 their patient case mix and hospital service mix, and therefore promote hospital quality improvement and better inform consumers about care quality.

Hospital-wide readmission is a priority area for outcomes 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.*
Distribution of HWR RSRRs over Different Time Periods
Results for each data year
Characteristic//07/2011-06/2012//07/2012-06/2013//07/2013-06/2014/
Number of Hospitals/4,821/ /4,794/ /4,772/
Number of Admissions/7,678,216/ /7,279,853/ /6,843,808/
Mean (SD)/16.2(1.1)/15.6(0.92)/ /15.5 (0.8)/
Range (min. – max.)/10.9-22.6/ /11.0-21.4/ /11.4-20.1/
Minimum/10.9/ /11.0/ /11.4/
10th percentile/15.1/ / 14.6/ /14.6/
20th percentile/15.4/ /14.9/ /14.9/
30th percentile/15.7/ /15.2/ /15.1/
40th percentile/15.9/ /15.4/ /15.3/
50th percentile/16.1/ /15.5/ /15.4/
60th percentile/16.4/ /15.7/ /15.6/
70th percentile/16.6/ /15.9/ /15.8/
80th percentile/17.0/ /16.2/ /16.0/
90th percentile/17.5/ /16.8/ /16.5/
Maximum/22.6/ /21.4/ /20.1/

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

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for 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.*
Distribution of HWR RSRRs by Proportion of Dual-Eligible Patients:
Dates of Data: July 2013 through June 2014
Data Source: Medicare FFS claims

Characteristic//Hospitals with a low proportion (=9.8%) Dual Eligible patients//Hospitals with a high proportion (=22.6%) Dual Eligible patients
Number of Measured Hospitals// 1,257 // 1,219
Number of Patients// 2,137,895 patients in low-proportion hospitals // 927,007 in high-proportion hospitals
Maximum// 18.7 // 20.1
90th percentile// 16.2 // 16.8
75th percentile// 15.7 // 16.0
Median (50th percentile)// 15.3 // 15.6
25th percentile// 14.8 // 15.2
10th percentile// 14.3 // 14.9
Minimum // 11.5 // 12.2

Distribution of HWR RSRRs by Proportion of African-American Patients:
Dates of Data: July 2013 through June 2014
Data Source: Medicare FFS claims

Characteristic// Hospitals with a low proportion (=2.2%) African-American patients//Hospitals with a high proportion (=9.4%) African-American patients
Number of Measured Hospitals// 1,156 // 1,180
Number of Patients// 222,648 patients in low-proportion hospitals/ 2,294,715 in high-proportion hospitals
Maximum// 19.1 // 19.9
90th percentile// 16.0 // 17.1
75th percentile// 15.6 // 16.3
Median (50th percentile)// 15.4 // 15.7
25th percentile// 15.1 // 15.2
10th percentile// 14.8 // 14.8
Minimum // 12.9 // 12.2

Distribution of HWR RSRRs by Proportion of Patients with AHRQ SES Index Scores Below 45.0:
Dates of Data: July 2013 through June 2014
Data Source: Medicare FFS claims and the American Community Survey (2008-2012) data

Characteristic//Hospitals with a low proportion of patients below AHRQ SES index score of 45.0 (=5.0%)// Hospitals with a high proportion of patients below AHRQ SES index score of 45.0 (=57.1%)
Number of Measures Hospitals// 1,209 // 1,217
Number of Patients// 1,651,852 patients in hospitals with low proportion of patients below AHRQ SES index score of 45.0 //795,899 patients in hospitals with high proportion of patients below AHRQ SES index score of 45.0
Maximum// 19.9 // 20.1
90th percentile// 16.2 // 16.6
75th percentile// 15.7 // 16.0
Median (50th percentile)// 15.3 // 15.5
25th percentile// 14.9 // 15.2
10th percentile// 14.5 // 14.8
Minimum // 11.5 // 13.0

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.**
During 2003 and 2004, almost one fifth of Medicare beneficiaries – over 2.3 million patients – were rehospitalized within 30 days of discharge from an acute care hospital (Jencks et al., 2009). Jencks et. al. estimated that readmissions within 30 days of discharge cost Medicare more than $17 billion annually (Jencks et al., 2009). A 2006 Commonwealth Fund report further estimated that if national readmission rates were lowered to the levels achieved by the top performing regions, Medicare would save $1.9 billion annually (The Commonwealth Fund, 2006). In a 2007 report to the Congress, the Medicare Payment Advisory Commission (MedPAC) estimated that in 2005, 17.6% of hospital patients were readmitted within 30 days of discharge and that 76% of these readmissions were potentially preventable; the average payment for a "potentially preventable" readmission was estimated at approximately $7,200 (MedPAC, 2007).

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. New England Journal of Medicine 2009;360(14):1418-28.

Why Not the Best? Results from a National Scorecard on U.S. Health System Performance. Fund Report. Harrisburg, PA: The Commonwealth Fund, 2006.

Medicare Payment Advisory Commission (U.S.). Report to the Congress promoting greater efficiency in Medicare. Washington, DC: Medicare Payment Advisory Commission, 2007.

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 : Acute Myocardial Infarction, Cardiovascular : Atrial Fibrillation, Cardiovascular : Congestive Heart Failure, Cardiovascular : Hyperlipidemia, Cardiovascular : Hypertension, Cardiovascular : Ischemic Heart Disease, Coronary Artery

Disease, Cardiovascular : Percutaneous Coronary Intervention (PCI), Cardiovascular : Screening, Endocrine, Endocrine : Diabetes, Endocrine : Screening, Endocrine : Thyroid Disorders, Gastrointestinal (GI), Gastrointestinal (GI) : Appendicitis, Gastrointestinal (GI) : Cirrhosis, Gastrointestinal (GI) : Gall Bladder Disease, Gastrointestinal (GI) : Gastroenteritis, Gastrointestinal (GI) : Gastro-Esophageal Reflux Disease (GERD), Gastrointestinal (GI) : GI Bleeding, Gastrointestinal (GI) : Peptic Ulcer, Gastrointestinal (GI) : Polyps, Gastrointestinal (GI) : Screening, GU/GYN, GU/GYN : Incontinence, GU/GYN : Screening, Infectious Diseases, Infectious Diseases : Hepatitis, Infectious Diseases : Human Immunodeficiency Virus/Acquired Immune Deficiency Syndrome (HIV/AIDS), Infectious Diseases : Respiratory, Infectious Diseases : Sexually Transmitted, Infectious Diseases : Tuberculosis, Musculoskeletal, Musculoskeletal : Hip/Pelvic Fracture, Musculoskeletal : Joint Surgery, Musculoskeletal : Low Back Pain, Musculoskeletal : Osteoarthritis, Musculoskeletal : Osteoporosis, Musculoskeletal : Rheumatoid Arthritis, Neurology, Neurology : Brain Injury, Neurology : Cognitive Impairment/Dementia, Neurology : Delirium, Neurology : Stroke/Transient Ischemic Attack (TIA), Pulmonary/Critical Care, Pulmonary/Critical Care : Asthma, Pulmonary/Critical Care : Chronic Obstructive Pulmonary Disease (COPD), Pulmonary/Critical Care : Critical Care, Pulmonary/Critical Care : Dyspnea, Pulmonary/Critical Care : Pneumonia, Pulmonary/Critical Care : Sleep Apnea, Renal, Renal : Chronic Kidney Disease (CKD), Renal : End Stage Renal Disease (ESRD), Surgery, Surgery : Cardiac Surgery, Surgery : General Surgery, Surgery : Perioperative, Surgery : Thoracic Surgery, Surgery : Vascular Surgery

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Care Coordination : Readmissions, Safety, Safety : Complications, Safety : Healthcare Associated Infections, Safety : Medication Safety, Safety : Readmissions

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:** NQF_1789_HWR_NQF_Data_Dictionary_01-29-16_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.
Annual Updates
1. Each year we update to the most current version of the Agency for Healthcare Research and Quality Clinical Classifications Software (AHRQ CCS) software by identifying any changes from the previous version that might impact the measure.
2. In addition, we have updated the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) Hierarchical Condition Categories (HCC) map annually to capture any changes that might impact the measure's risk model. The version of the HCC map used for this measure has not been updated since 2013.

Updates by year
2015
1. Respecified the measure by updated to CMS Planned Readmission Algorithm (Version 4.0).
Rationale: Version 4.0 incoprporates additional improvements made following a validation study of the algorithm using data from a medical record review. These changes required additional input from clinical experts and were, therefore, not included in the changes made in version 3.0. The changes improve the accuracy of the algorithm by more correctly classifying planned and unplanned readmissions.

2014
1. Updated to CMS Planned Readmission Algorithm (Version 3.0).
Rationale: Version 3.0 incorporates improvements made following a validation study of the algorithm using data from a medical record review. These changes improve the accuracy of the algorithm by decreasing the number of readmissions that the algorithm mistakenly designated as planned by removing two procedure categories and adding several acute diagnoses.

2013

1. Updated to CMS Planned Readmission Algorithm (Version 2.1).
Rationale: Version 2.1 incorporated improvements to the original algorithm made following an extensive review by clinical experts and stakeholder feedback submitted during the HWR measure's public comment period and 2012 dry run.

3. Removed procedure CCS 61 (Other or procedures on vessels other than head and neck) from the list of procedures qualifying an admission for the surgery cohort.
Rationale: This procedure CCS was removed from the surgical cohort because patients undergoing this procedure are typically admitted primarily for cardiovascular or medical care.

4. Modified the planned readmission algorithm handling of admissions to psychiatric and rehabilitation hospitals.
Rationale: Psych and rehab hospitals in Maryland have the same provider ID number as acute care hospitals. Therefore, readmissions are not counted if the patient has a principal diagnosis code beginning with a "V57" (indication of admission to a rehab unit) or if all three of the following criteria are met: (1) the admission being evaluated as a potential readmission has a psychiatric principal discharge diagnosis code (ICD-9 codes 290-319); (2) the index admission has a discharge disposition code to a psychiatric hospital or psychiatric unit from the index admission; and (3) the admission being evaluated as a potential readmission occurred during the same day as or the day following the index discharge. The criteria for identifying such admissions are available in the 2010 Measures Maintenance Technical Report: Acute Myocardial Infarction, Heart Failure, and Pneumonia 30-Day Risk-Standardized Readmission Measures.

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 readmission. We define readmission as an inpatient admission for any cause, with the exception of certain planned readmissions, within 30 days from the date of discharge from an eligible index admission. If a patient has more than one unplanned admission (for any reason) within 30 days after discharge from the index admission, only one is counted as a readmission. The measure looks for a dichotomous yes or no outcome of whether each admitted patient has an unplanned readmission within 30 days. However, if the first readmission after discharge is considered planned, any subsequent unplanned readmission is not counted as an outcome for that index admission because the unplanned readmission could be related to care provided during the intervening planned readmission rather than during the index admission.

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 admission.

Denominator Time Window: This measure was developed with 12 months of data and is currently publicly reported with one year of data.

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.
The measure counts readmissions to any acute care hospital for any cause within 30 days of the date of discharge of the index admission, excluding planned readmissions as defined below.

Planned Readmission Algorithm (Version 4.0)
The Planned Readmission Algorithm is a set of criteria for classifying readmissions as planned among the general Medicare population using Medicare administrative claims data. The algorithm identifies admissions that are typically planned and may occur within 30 days of discharge from the hospital.

The Planned Readmission Algorithm has three fundamental principles:

1. A few specific, limited types of care are always considered planned (obstetric delivery, transplant surgery, maintenance chemotherapy/immunotherapy, rehabilitation);
2. Otherwise, a planned readmission is defined as a non-acute readmission for a scheduled procedure; and

3. Admissions for acute illness or for complications of care are never planned.

The algorithm was developed in 2011 as part of the Hospital-Wide Readmission measure. In 2013, CMS applied the algorithm to its other readmission measures.

The Planned Readmission Algorithm and associated code tables are attached in data field S.2b (Data Dictionary or Code Table).

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The measure includes admissions for Medicare beneficiaries who are 65 years and older and are discharged from all non-federal, acute care inpatient US hospitals (including territories) with a complete claims history for the 12 months prior to admission.

Additional details are provided in S.9 Denominator Details.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

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)*
To be included in the measure cohort patients must be:

1. Enrolled in Medicare fee-for-service (FFS) Part A for the 12 months prior to the date of admission and during the index admission;
2. Aged 65 or over;
3. Discharged alive from a non-federal short-term acute care hospital; and
4. Not transferred to another acute care facility.

The measure aggregates the ICD-9 principal diagnosis and all procedure codes of the index admission into clinically coherent groups of conditions and procedures (condition categories or procedure categories) using the AHRQ CCS. There are a total of 285 mutually exclusive AHRQ condition categories, most of which are single, homogenous diseases such as pneumonia or acute myocardial infarction. Some are aggregates of conditions, such as “other bacterial infections.” There are a total of 231 mutually exclusive procedure categories. Using the AHRQ CCS procedure and condition categories, the measure assigns each index hospitalization to one of five mutually exclusive specialty cohorts: surgery/gynecology, cardiorespiratory, cardiovascular, neurology, and medicine. The rationale behind this organization is that conditions typically cared for by the same team of clinicians are expected to experience similar added (or reduced) levels of readmission risk.

The measure first assigns admissions with qualifying AHRQ procedure categories to the Surgery/Gynecology Cohort. This cohort includes admissions likely cared for by surgical or gynecological teams.

The measure then sorts admissions into one of the four remaining specialty cohorts based on the AHRQ diagnosis category of the principal discharge diagnosis:

The Cardiorespiratory Cohort includes several condition categories with very high readmission rates such as pneumonia, chronic obstructive pulmonary disease, and heart failure. These admissions are combined into a single cohort because they are often clinically indistinguishable and patients are often simultaneously treated for several of these diagnoses.

The Cardiovascular Cohort includes condition categories such as acute myocardial infarction that in large hospitals might be cared for by a separate cardiac or cardiovascular team.

The Neurology Cohort includes neurologic condition categories such as stroke that in large hospitals might be cared for by a separate neurology team.

The Medicine Cohort includes all non-surgical patients who were not assigned to any of the other cohorts.

The full list of the specific diagnosis and procedure AHRQ CCS categories used to define the specialty cohorts are attached in data field S.2b (Data Dictionary or Code Table).

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
The measure excludes index admissions for patients:

1. Admitted to Prospective Payment System (PPS)-exempt cancer hospitals;
2. Without at least 30 days post-discharge enrollment in FFS Medicare;
3. Discharged against medical advice (AMA);
4. Admitted for primary psychiatric diagnoses;
5. Admitted for rehabilitation; or
6. Admitted for medical treatment of cancer.

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)*
1. Admitted to a PPS-exempt cancer hospital, identified by the Medicare provider ID.
2. Admissions without at least 30 days post-discharge enrollment in FFS Medicare are determined using data captured in the Medicare Enrollment Database (EDB).
3. Discharges against medical advice (AMA) are identified using the discharge disposition indicator in claims data.
4. Admitted for primary psychiatric disease, identified by a principal diagnosis in one of the specific AHRQ CCS categories listed in the attached data dictionary.
5. Admitted for rehabilitation care, identified by the specific ICD-9 diagnosis codes included in CCS 254 (Rehabilitation care; fitting of proestheses; and adjustment of devices).
6. Admitted for medical treatment of cancer, identified by the specific AHRQ CCS categories listed in the attached data dictionary.

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 employs a hierarchical logistic regression model to create a hospital-level 30-day RSRR. In brief, the approach simultaneously models data at the patient and hospital levels to account for the variance in patient outcomes within and between hospitals (Normand & Shahian, 2007). At the patient level, the model adjusts the log-odds of readmission within 30 days of discharge for age and selected clinical covariates. At the hospital level, the approach models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of readmission at the hospital, after accounting for patient risk. If there were no differences among hospitals, then after adjusting for patient risk, the hospital intercepts should be identical across all hospitals.

We use a fixed, common set of variables in all our models for simplicity and ease of data collection and analysis. However, we estimate a hierarchical logistic regression model for each specialty cohort separately, and the coefficients associated with each variable may vary across specialty cohorts.

Candidate and Final Risk-adjustment Variables: Candidate variables were patient-level risk-adjustors that were 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 claims records extending 12 months prior to and including the index admission. For the measure currently implemented by CMS, these risk-adjusters are identified using inpatient Medicare FFS claims data.

The model adjusts for case-mix differences based on the clinical status of patients 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 (Pope et al., 2000). A file that contains a list of the ICD-9-CM codes and their groupings into CCs is attached in data field S.2b (Data Dictionary or Code Table). In addition, only comorbidities that convey information about the patient at admission or in the 12 months prior, and not complications that arise during the course of the index hospitalization, are included in the risk adjustment. Hence, we do not risk adjust for CCs that may represent adverse events of care when they are only recorded in the index admission. The models also include a condition-specific indicator for all AHRQ CCS categories with sufficient volume (defined as those with more than 1,000 admissions nationally each year for Medicare FFS data) as well as a single indicator for conditions with insufficient volume in each model.

The final set of risk adjustment variables are listed in the attached Data Dictionary.

Demographics
Age-65 (years, continuous) for patients aged 65 or over cohorts; or Age (years, continuous) for patients aged 18 and over cohorts

Comorbidities
Metastatic cancer or acute leukemia (CC 7)
Severe cancer (CC 8-9)
Other cancers (CC 10-12)
Severe hematological disorders (CC 44)
Coagulation defects and other specified hematological disorders (CC 46)
Iron deficiency or other unspecified anemias and blood disease (CC 47)
End-stage liver disease (CC 25-26)
Pancreatic disease (CC 32)
Dialysis status (CC 130)
Renal failure (CC 131)
Transplants (CC 128, 174)
Severe infection (CC 1, 3-5)
Other infectious diseases and pneumonias (CC 6, 111-113)
Septicemia/shock (CC 2)
Congestive heart failure (CC 80)
Coronary atherosclerosis or angina, cerebrovascular disease (CC 81-84, 89, 98-99, 103-106)
Specified arrhythmias and other heart rhythm disorders (CC 92-93)
Cardio-respiratory failure or shock (CC 79)
Chronic obstructive pulmonary disease (COPD) (CC 108)
Fibrosis of lung or other chronic lung disorders (CC 109)
Protein-calorie malnutrition (CC 21)
Disorders of fluid/electrolyte/acid-base (CC 22-23)
Rheumatoid arthritis and inflammatory connective tissue disease (CC 38)
Diabetes mellitus (DM) or DM complications (CC 15-20, 119-120)
Decubitus ulcer or chronic skin ulcer (CC 148-149)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)
Seizure disorders and convulsions (CC 74)
Respirator dependence/tracheostomy status (CC 77)
Drug/alcohol psychosis or dependence (CC 51-52)
Psychiatric comorbidity (CC 54-56, 58, 60)
Hip fracture/dislocation (CC 158)

Principal Diagnoses
Refer to the 2015 Measure Updates and Specifications: Hospital-Wide All-Cause Unplanned Readmission - Version 4.0 referenced here for the full lists of principal diagnosis AHRQ CCS categories included in each specialty cohort risk adjustment model.

References:
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.

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

Pope GC, et al. 2000. Principal Inpatient Diagnostic Cost Group Models for Medicare Risk Adjustment. Health Care Financing Review 21(3): 93-118.

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*)

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.)*
The measure estimates hospital-level 30-day all-cause RSRRs using hierarchical logistic regression models. In brief, the approach simultaneously models data at the patient and hospital levels to account for variance in patient outcomes within and between hospitals (Normand et al., 2007). At the patient level, it models the log-odds of hospital readmission within 30 days of discharge using age, selected clinical covariates, and a hospital-specific effect. At the hospital level, the approach models the hospital-specific effects as arising from a normal distribution. The hospital effect represents the underlying risk of a readmission at the hospital, after accounting for patient risk. The hospital-specific effects are given a distribution to account for the clustering (non-independence) of patients within the same hospital (Normand et al., 2007). If there were no differences among hospitals, then after adjusting for patient risk, the hospital effects should be identical across all hospitals.

Admissions are assigned to one of five mutually exclusive specialty cohort groups consisting of related conditions or procedures. For each specialty cohort group, the standardized readmission ratio (SRR) is calculated as the ratio of the number of "predicted" readmissions to the number of "expected" readmissions at a given hospital. For each hospital, the numerator of the ratio is the number of readmissions within 30 days predicted based on the hospital's performance with its observed case mix and service mix, and the denominator is the number of readmissions expected based on the nation's performance with that hospital's case mix and service mix. This approach is analogous to a ratio of "observed" to "expected" used in other types of statistical analyses. It conceptually allows a particular hospital's performance, given its case mix and service mix, to be compared to an average hospital's performance with the same case mix and service mix. Thus, a lower ratio indicates lower-than-expected readmission rates or better quality, while a higher ratio indicates higher-than-expected readmission rates or worse quality.

For each specialty cohort, the "predicted" number of readmissions (the numerator) is calculated by using the coefficients estimated by regressing the risk factors (found in Table D.9) and the hospital-specific effect on the risk of readmission. The estimated hospital-specific effect for each cohort is added to the sum of the estimated regression coefficients multiplied by patient characteristics. The results are log transformed and summed over all patients attributed to a hospital to get a predicted value. The "expected" number of readmissions (the denominator) is obtained in the same manner, but a common effect using all hospitals in our sample is added in place of the hospital-specific effect. The results are log transformed and summed over all patients in the hospital to get an expected value. To assess hospital performance for each reporting period, we re-estimate the model coefficients using the data in that period.

The specialty cohort SRRs are then pooled for each hospital using a volume-weighted geometric mean to create a hospital-wide composite SRR. The composite SRR is multiplied by the national observed readmission rate to produce the RSRR. The statistical modeling approach is described fully in Appendix A and in the original methodology report (Horwitz et al., 2012).

References:

Horwitz L, Partovian C, Lin Z, et al. Hospital-Wide All-Cause Unplanned Readmission Measure: Final Technical Report. 2012; http://www.qualitynet.org/dcs/BlobServer?blobkey=id&blobnocache=true&blobwhere=1228889825199&blobheader=multipart%2Foctet-stream&blobheadername1=Content-Disposition&blobheadervalue1=attachment%3Bfilename%3DDryRun_HWR_TechReport_081012.pdf&blobcol=urldata&blobtable=MungoBlobs. Accessed 30 April, 2014.

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

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.

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 survey or patient-reported data.

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.
Administrative claims

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:

1. Medicare Part A claims data for calendar years 2007 and 2008 were combined and then randomly split into two equal subsets (development sample and validation sample). Risk variable selection was done using the development sample, the risk models for each of the five specialty cohorts in the measure were applied to the validation sample and the models' performance was compared. In addition we re-tested the models in Medicare Part A claims data from calendar year 2009 to look for temporal stability in the models' performance. The number of measured entities and index admissions are listed below by specialty cohort.

2. 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 and following discharge from index 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/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_1789_HWR_NQF_Testing_Attachment_01-29-16_v1.1.docx

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

Measure Number (*if previously endorsed*)**:** 1789
Measure Title: Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)
Date of Submission: 1/29/2016
Type of Measure: Quality Outcome Measure

☐ Composite – ***STOP – use composite testing form***	☒ Outcome (*including PRO-PM*)
☐ 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 Part A inpatient claims and the Medicare Enrollment Database (EDB). Census as well as claims data were used to assess socioeconomic factors and race (dual-eligible and African American race variables were 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?

The dates used vary by testing type; see Section 1.7 for details.

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

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

☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

For this measure, hospitals are the measured entities. All non-federal, acute care inpatient US hospitals (including territories) with Medicare fee-for-service (FFS) beneficiaries aged 65 years and older are included. The number of measured entities (hospitals) varies by testing type; see Section 1.7 for details.

1.6. How many and which patients 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:

For reliability testing (Section 2a2)
The reliability of the model was tested by randomly selecting 50% of the Medicare patients aged 65 years or over within each hospital in the most recent 1-year measure cohort and calculating the measure results for each hospital. We then calculate the measure results for the remaining 50% of patients within each hospital and compare the two. Thus, for reliability testing, we randomly split **Dataset 1** into two samples. In each year of measure reevaluation, we also re-fit the model and examine frequencies and model coefficients of risk variables (condition categories for patient comorbidities) and model fit in the new year of data (**Dataset 1** below).

Dataset 1 (2015 public reporting cohort version 4.0): Medicare Part A Inpatient Claims and Medicare Enrollment Database
Dates of Data: July 1, 2013 – June 30, 2014
Number of index admissions: 6,843,808

Number of hospitals: 4,772
Average age of patients: 78.3

For testing of measure exclusions (Section 2b3)
Dataset 1 (2015 public reporting cohort version 4.0): Medicare Part A Inpatient Claims and Medicare Enrollment Database
Dates of Data: July 1, 2013 – June 30, 2014
Number of index admissions: 6,843,808

Number of hospitals: 4,772
Average age of patients: 78.3

For testing of measure risk adjustment (Section 2b4)
Dataset 2: Medicare Part A claims data for calendar years 2007 and 2008 were combined and then randomly split into two equal subsets (development sample and validation sample). Risk variable selection was done using the development sample, the risk models for each of the five specialty cohorts in the measure were applied to the validation sample and the models' performance was compared. In addition we re-tested the models in Medicare Part A claims data from calendar year 2009 to look for temporal stability in the models' performance. The number of measured entities and index admissions are listed below by specialty cohort.

Medicine model:
Development sample: 3,085,962 admissions to 4,954 hospitals
Validation sample: 3,082,357 admissions to 4,946 hospitals
2009 sample: 3,032, 518 admissions to 4,908 hospitals
Surgery/gynecology model:
Development sample: 2, 208753 admissions to 4,354 hospitals
Validation sample: 2,208,482 admissions to 4,353 hospitals
2009 sample: 2,109,292 admissions to 4,232 hospitals
Cardiorespiratory model:
Development sample: 1,396562 admissions to 4,810 hospitals
Validation sample: 1,396,855 admissions to 4,806 hospitals
2009 sample: 1,331,539 admissions to 4,718 hospitals
Cardiovascular model:
Development sample: 860,485 admissions to 4,702 hospitals
Validation sample: 861,925 admissions to 4,703 hospitals
2009 sample: 809,520 admissions to 4,641 hospitals
Neurology model:
Development sample: 461,225 admissions to 4,699 hospitals
Validation sample: 461,262 admissions to 4,686 hospitals
2009 sample: 452,743 admissions to 4,609 hospitals

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

For testing of socioeconomic status (SES) factors and race in risk models (Section 2b4.3)
Dataset 1 and **Dataset 3**: 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 1**)
- Validated AHRQ SES index score is a composite of 7 different variables found in the census data (**Dataset 3**)

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 (SES) and "race" as separate terms.

We selected 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 AB et al., 2014; Eapen ZJ et al. 2015; Gilman M et al., 2014; Hu J et al., 2014; Joynt KE and Jha AK, 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 across multiple conditions is more limited, available studies suggest a consistent association between SES/race variables and risk of readmission (Aseltine RH et al., 2015; Gu Q et al., 2014; Arbaje AI et al., 2008). 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 1**)
- African-American race (**Dataset 1**)
- 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 3**)

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 5-digit zip code and census block levels only. The data are not currently available at the 9-digit zip code level. In this submission, we present analysis using the 5-digit level. However, we are currently

performing analysis at the census block level, the most granular level possible. We hope to present the results of the census block-level analysis to the committee.

References:

Arbaje AI, Wolff JL, Yu Q, Powe NR, Anderson GF, Boult C. Post discharge environmental and socioeconomic factors and the likelihood of early hospital readmission among community-dwelling Medicare beneficiaries. The Gerontologist. 2008;48(4):495-504.

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.

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.

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.

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.

Gu Q, Koenig L, Faerberg J, Steinberg CR, Vaz C, Wheatley MP. The Medicare Hospital Readmissions Reduction Program: potential unintended consequences for hospitals serving vulnerable populations. Health services research. 2014;49(3):818-837.

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.

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, 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 in each new data year.

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 V, et al., 2002).

For test-retest reliability, we randomly sampled half of patients within each hospital in the most recent year of data, 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 P and Fleiss J, 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 by Shrout P and Fleiss J (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 CC, 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
The frequency of some model variables are assessed in each data year. From year-to-year the frequency of individual variables may increase or decrease slightly. These changes may reflect small changes in rates of comorbidity in the fee-for-service population. For details please see the attached 2015 Measure Updates and Specifications Report. Reports from previous years can be found on QualityNet.

Measure Score Reliability Results
There were 6,843,808 admissions in the 2015 public reported measures (**Dataset 1**), with 3,420,728 in one sample and 3,423,080 in the other randomly selected sample. The agreement between the two RSRRs for each hospital was 0.80, which according to the conventional interpretation is "substantial" (Landis J & Koch G, 1977).

Reference:
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 ICC score demonstrates substantial agreement across samples using a conservative approach to assessment.

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:
Measure validity is demonstrated through prior validity testing done on our other claims-based measures, through use of established measure development guidelines, and examination of content validity by comparing hospital performance with that on other quality measures.

Validity of Claims Data:
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 the corresponding results and elements from medical records. CMS validated the six NQF-endorsed measures currently in public

reporting (acute myocardial infarction [AMI], heart failure, 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 AMI patients (Cooperative Cardiovascular Project data), (Krumholz HM, et al., 2006) heart failure patients (National Heart Failure data), (Krumholz HM, et al., 2006, Keenan PS, et al., 2006), and pneumonia patients (National Pneumonia Project dataset), (Bratzler DW, et al., 2011). 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.
We have also completed two national, multi-site validation efforts for two procedure-based complications measures (for primary elective hip/knee arthroplasty and implantable cardioverter defibrillator [ICD]). Both projects demonstrated strong agreement between complications coded in claims and abstracted medical chart data. These validation efforts suggest that such claims data variables are valid across a variety of conditions.

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 National Quality Forum (NQF) guidance for outcomes measures (National Quality Forum, 2012), CMS Measure Management System guidance, and the guidance articulated in the American Heart Association scientific statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz HM, et al., 2006).

Validation Against Other Outcomes Measures:

In order to test the construct validity of the HWR measure, we examined whether hospitals considered "top performers" according to other measures and ranking systems had lower hospital-wide risk-standardized readmission rates than remaining hospitals when applying our measure to the Medicare FFS population. This type of validity testing tests the assumption that hospitals considered top performers have developed an organizational culture of excellence that will manifest itself in better outcomes including lower hospital-wide readmission rates. However, there are multiple challenges associated with this approach:
1. There are many measures and ranking systems available, using a variety of criteria in order to define and select top performers, including: adherence to core processes of care, complications and safety indexes, resource utilization, outcomes, patient satisfaction, and even reputation. "Top performers" on one measure are not the same as "top performers" on another. Moreover, most of these measures are not themselves validated.
2. In many cases, the methodology for identifying "top performers" is proprietary and not transparent.
3. The starting set of hospitals from which different ranking systems select the top performers usually includes only a subset of all acute care hospitals included in the HWR measure; in most cases it is not possible to replicate this starting set exactly.
4. We have not found a ranking system which specifically measures factors most relevant to readmission risk, such as medication reconciliation, patient education, post-discharge follow up, or communication with outpatient clinicians.
After reviewing ranking systems, we selected the following three to use for construct validity testing because they are widely used and their methodology is available to the public:

1. Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey score
http://www.hcahpsonline.org/home.aspx
2. Thomson Reuters 100 top hospitals
http://100tophospitals.com/Portals/2/assets/TOP%2015313%200315%20100%20Top%20Study_web.pdf
3. Joint Commission list of Top Performers on Key Quality Measures
http://www.jointcommission.org/accreditation/top_performers.aspx

1. HCAHPS
From the 27 questions in the HCAHPS survey, we selected seven that we felt were most likely to be correlated with readmission rates based on clinical judgment and previously reported results by others (Akamigbo A, 2010, Jha, AK, et al., 2008). Based on previous results we expected to see that patient satisfaction is significantly correlated with hospital readmission rates. For this analysis, we compared 2009 HCAHPS results to 2009 Medicare FFS RSRRs. See results in Section 2b2.3.

2. Thomson Reuters Top 100 Hospitals
Given that this measure includes several elements theoretically related to readmission risk, including complications, patient safety, readmissions, and HCAHPS, we felt this measure was a reasonable candidate for construct validity testing. However, since the measure also contains other components such as core measures, expenses, and profitability that would not be expected to correlate with readmission, we expected the analysis to show at best small improvements in readmission performance among top performers. See results in Section 2b2.3.

3. The Joint Commission's Top Performers on Key Quality Measures program
Of the Joint Commission's list of 405 top performers, we selected only those 158 hospitals with superior performance in *all four* adult measure sets (HWR is for patients 18 years and older), on the assumption that these hospitals demonstrated hospital-wide performance excellence. We calculated their hospital-wide readmission rates and compared them to those of other hospitals. However, since numerous studies have shown that there is little relationship between performance on core process measures and outcomes including mortality and readmission rates we expected the Joint Commission's top performers to have similar risk-standardized readmission rates as other hospitals, (Bradley E H, et al., 2006; Werner R, et al., 2006; Fonarow GC, et al. 2007; Fonarow GC and Peterson E, 2009; Jha AK, et al., 2009; Patterson M, et al., 2010; Shwartz, M et al., 2011). See results in Section 2b2.3.

References:

Akamigbo A. The Relationship Between Hospital Readmissions and HCAHPS Scores 2010; https://cahps.ahrq.gov/news-and-events/events/UGM12/CAHPS/AkamigboA.pdf

Bradley EH, Herrin J, Elbel, B, McNamara, RL, Magid, DJ, Nallamothu, BK, Krumholz, HM. Hospital quality for acute myocardial infarction: Correlation among process measures and relationship with short-term mortality. Journal of the American Medical Association 2006;296:72-78.

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

Fonarow GC et al. Association between performance measures and clinical outcomes for patients hospitalized with heart failure. Journal of the American Medical Association 2007; 297(1):61-70.

Fonarow GC and Peterson E. Heart failure performance measures and outcomes. Journal of the American Medical Association 2009; 302(7): 792-794.

Jha AK, Orav EJ, Zheng J, and Epstein AM. Patients' perception of hospital care in the United States. New England Journal of Medicine 2008;359 (18):1921-31.

Jha AK, Orav EJ, Zheng J, and Epstein AM. Public reporting of discharge planning and rates of readmissions. New England Journal of Medicine 2009; 361(27):2637-45.

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

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

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

Krumholz HM, Brindis RG, Brush JE, et al. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. Circulation. January 24, 2006 2006;113(3):456-462.

National Quality Forum. National voluntary consensus standards for patient outcomes, first report for phases 1 and 2: A consensus report REVISED Hospital-Wide Readmission NQF Application January 5, 2012 31

Patterson M, et al. Process of care performance measures and long-term outcomes in patients hospitalized with heart failure. Medical Care 2010; 48(3):210-216.

Shwartz M, et al. How well can we identify the high-performing hospital? Medical Care Research and Review 68;3 (2011):290-310

Werner R and Bradlow E. Relationship between Medicare's hospital compare performance measures and mortality rates. Journal of the American Medical Association 2006;296(22):2694-270.

International Classification of Diseases, Ninth Revision (ICD-9) to International Classification of Diseases, Tenth Revision (ICD-10) Conversion

Statement of Intent
[X] Goal was to convert this measure to a new code set, fully consistent with the intent of the original measure.
[ ] Goal was to take advantage of the more specific code set to form a new version of the measure, but fully consistent with the original intent.
[ ] The intent of the measure has changed.

Process of Conversion
ICD-10 codes were initially identified using 2015 General Equivalence Mappings (GEM) 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. An ICD-9 to ICD-10 crosswalk is attached in field S.2b. (Data Dictionary or Code Table).

We have also examined the updated ICD-9 Map to AHRQ Clinical Classification Software (CCS) crosswalk to the ICD-10 CCS map provided by AHRQ in preparation for the inclusion of ICD 10 data in this measure. Please refer to the ICD-10 CCS map on the AHRQ website.

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

Content validity results from the analyses described above, are presented describing comparisons of hospital performance on the HWR measure and other selected quality metrics.

1. HCAHPS

Table 1. shows the correlation (Pearson correlation coefficient) between RSRR and the proportion of patients who responded in that given manner to the question. The analysis includes the 3,723 hospitals that publicly report HCAHPS results.

Table 1. Correlation between RSRR (2009 Medicare FFS data) and HCAHPS response (N=3,723 hospitals)

HCAHPS Question	Correlation
Pain was 'sometimes' or 'never' well controlled	0.34
Patients 'sometimes' or 'never' received help as soon as they wanted	0.34
Nurses 'sometimes' or 'never' communicated well	0.33
'NO' patients would not recommend the hospital	0.32
Patients were 'sometimes' or 'never' given information about what to do during their recovery at home	0.32
Patients who gave a rating of '6' or lower	0.31
Doctors 'sometimes' or 'never' communicated well	0.21

p value for all correlations <0.001REVISED Hospital-Wide Readmission NQF Application January 5, 2012 32

2. Thomson Reuters Top 100 Hospitals

Table 2. shows the RSRRs distribution for the top performers in comparison to the rest of hospitals.

Table 2. Distribution of RSRRs (2009 Medicare FFS data) for the Thomson Reuters Top 100 Hospitals vs. others

	On List	Not On List
Number	100	3017
Mean (SD)	16.19 (1.39)	16.65 (1.28)
Minimum	13.77	12.51
Lower Quartile	15.05	15.79
Median	16.06	16.51
Upper Quartile	16.99	17.35
Maximum	19.81	22.69

3. The Joint Commission's Top Performers on Key Quality Measures program

Table 3. shows the distribution of risk-standardized readmission rates of the 158 top performers compared to other hospitals.

Table 3. Distribution of RSRR (2009 Medicare FFS data) for The Joint Commission's Top Performers vs. Others

	On List	Not On List
Number	158	4630
Mean (SD)	16.66 (0.99)	16.61 (1.16)
Minimum	14.18	12.51
Lower Quartile	16.01	15.87
Median	16.64	16.49
Upper Quartile	17.17	17.21
Maximum	19.91	22.69

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?*)

In summary, these three construct validity analyses demonstrated results consistent with our expectations. There is a significant correlation between patient satisfaction and RSRR as measured by the HWR measure. "Top performers" as defined by Thomson Reuters have lower RSRRs as measured by the HWR measure. On the other hand, hospitals identified by The Joint Commission as having superior performance on all four categories of clinical process measures have identical performance as those with lower performance, consistent with published studies.

_____

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*)

All exclusions were determined by careful clinical review and have been made based on clinically relevant decisions and to ensure accurate calculation of the measure. To ascertain impact of exclusions on the cohort, we examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion (**Dataset 1**). These exclusions are consistent with similar NQF-endorsed outcome measures. Rationales for the exclusions are detailed in data field S.10 (Denominator Exclusions).

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*)

In Dataset 1(2015 Public Reporting Cohort):

Exclusion	N	%	Distribution across hospitals (N=4,802): Minimum, 25th percentile, 50th percentile, 75th percentile, maximum
Admitted to PPS-Exempt Cancer Hospitals	19823	0.28	(0.00, 0.00, 0.00, 0.00, 100.00)
Without 30 Days of Post-Discharge Enrollment	36640	0.52	(0.00, 2.3, 3.1, 4.0, 22.2)
Discharged against medical advice (AMA)	26665	0.38	(0.00, 0.20, 0.50, 1.00, 21.1)
Admitted for Primary Psychiatric Diagnosis	19691	0.28	(0.00, 0.00, 0.10, 0.40, 100.00)
Admitted for Rehabilitation	7152	0.10	(0.00, 0.00, 0.00, 0.00, 100.00)
Admitted for Medical Treatment of Cancer	152288	2.15	(0.00, 0.60, 1.30, 1.90, 55.00)

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*)

Exclusions applied to the HWR measure cohort

1. Patients admitted to Inpatient Prospective Payment System (IPPS)-exempt cancer hospitals account for 0.28% of all index admissions excluded from the initial cohort. Admissions for treatment of cancer are associated with a very different mortality and readmission risk compared with admissions to other IPPS hospitals for treatment of other diseases. Additionally, outcomes for these admissions do not correlate well with outcomes for other types of admissions. (Patients with cancer who are admitted for other diagnoses or for surgical treatment of their cancer remain in the measure).
2. Patients without at least 30 days post-discharge enrollment in FFS Medicare following discharge account for 0.52% of all index admissions excluded from the initial cohort. This exclusion is needed since the 30-day readmission outcome cannot be assessed in patients who do not maintain enrollment for at least 30 days following discharge.
3. Patients discharged against medical advice (AMA) account for 0.38% of all index admissions excluded from the initial index cohort. This exclusion is needed for acceptability of the measure to hospitals, who do not have the opportunity to adequately deliver full care and prepare the patient for discharge.
4. Patients admitted for primary psychiatric diagnoses account for 0.28% of all index admissions excluded from the initial cohort. This exclusion is needed because these patients are typically cared for in separate psychiatric or rehabilitation centers which are not comparable to acute care hospitals.
5. Patients admitted for rehabilitation account for 0.10% of all index admissions excluded from the initial cohort. This exclusion is needed because patients admitted for rehabilitation are not admitted for treatment of acute illness and the care provided in rehabilitation centers is not comparable to care provided in acute care hospitals.
6. Patients admitted for medical treatment of cancer account for 2.15% of all index admissions excluded from the initial cohort. Admissions for treatment of cancer are associated with a very different mortality and readmission risk compared with admissions to other IPPS hospitals for treatment of other diseases. Additionally, outcomes for these admissions do not correlate well with outcomes for other types of admissions. (Patients with cancer who are admitted for other diagnoses or for surgical treatment of their cancer remain in 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 33 risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (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 HM, et al., 2006). The measure estimates hospital-level 30-day all-cause RSRRs using hierarchical logistic regression models. In brief, the approach simultaneously models data at the patient and hospital levels to account for variance in patient outcomes within and between hospitals, (Normand S-LT, Shahian DM, 2007).

At the patient level, it models the log-odds of hospital readmission within 30 days of discharge using age, selected clinical covariates, and a hospital-specific intercept. At the hospital level, the approach models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of a readmission at the hospital, after accounting for patient risk. The hospital-specific intercepts are given a distribution to account for the clustering (non-independence) of patients within the same hospital (Normand S-LT, Shahian DM, 2007). If there were no differences among hospitals, then after adjusting for patient risk, the hospital intercepts should be identical across all hospitals.

Admissions are assigned to one of five mutually exclusive specialty cohort groups consisting of related conditions or procedures. For each specialty cohort group, the standardized readmission ratio (SRR) is calculated as the ratio of the number of "predicted" readmissions to the number of "expected" readmissions at a given hospital. For each hospital, the numerator of the ratio is the number of readmissions within 30 days predicted based on the hospital's performance with its observed case mix and service mix, and the denominator is the number of readmissions expected based on the nation's performance with that hospital's case mix and service mix. This approach is analogous to a ratio of "observed" to "expected" used in other types of statistical analyses. It conceptually allows a particular hospital's performance, given its case mix and service mix, to be compared to an average hospital's performance with the same case mix and service mix. Thus, a lower ratio indicates lower-than-expected readmission rates or better quality, while a higher ratio indicates higher-than-expected readmission rates or worse quality.

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

The specialty cohort SRRs are then pooled for each hospital using a volume-weighted geometric mean to create a hospital-wide composite SRR. The composite SRR is multiplied by the national observed readmission rate to produce the RSRR.

Data Source

The HWR risk-adjustment models use only inpatient claims data (history and current) in order to make it feasible to implement with Medicare data, and to make it applicable to all-payer data, which are typically restricted to inpatient claims.

The HWR measure uses CMS-CCs (Horwitz L, Partovian C, Lin Z, et al. 2012), the grouper used in previous CMS risk-standardized outcomes measures, to group ICD-9-CM codes into comorbid risk adjustment variables,

since four CMS condition-specific claims-based readmission models that use this grouper to define variables for risk adjustment have been validated against models that use chart-abstracted data for risk adjustment (Pope G, et al., 2000, Keenan PS, Normand SL, Lin Z, et al., 2008, Krumholz HM, Lin Z, Drye EE, et al. 2011).

Approach to Variable Selection:

In order to select the comorbid risk variables, we developed a "starter" set of 30 variables drawn from previous readmission measures (AMI, heart failure, pneumonia, hip and knee arthroplasty, and stroke). Next we reviewed all the remaining CMS-CCs and determined on a clinical basis whether they were likely to be relevant to an all-condition measure. We selected 11 additional risk variables to consider.

Using data from the index admission and any admission in the prior 12 months, we ran a standard logistic regression model for every discharge condition category with the full set of candidate risk adjustment variables. We compared odds ratios for different variables across different condition categories (excluding condition categories with fewer than 700 readmissions due to the number of events per variable constraints). We selected the final set of comorbid risk variables based on the following principles:

• We excluded risk variables that were statistically significant for very few condition categories, given that they would not contribute much to the overall models.

• We excluded risk variables that behaved in clinically incoherent ways. For example, we dropped risk variables that sometimes increased risk and sometimes decreased risk, when we could not identify a clinical rationale for the differences.

• We excluded risk variables that were predominantly protective when we felt this protective effect was not clinically reasonable but more likely reflected coding factors. For example, drug/alcohol abuse without dependence (CC 53) and delirium and encephalopathy (CC 48) were both protective for readmission risk although clinically they should increase patients' severity of illness.

• Where possible, we grouped together risk variables that were clinically coherent and carried similar risks across condition categories. For example, we combined coronary artery disease (CCs 83-84) with cerebrovascular disease (CCs 98, 99, and 103).

• We examined risk variables that had been combined in previous CMS publicly reported measures, and in one instance separated them: for cancers, the previous measures generally pool 5 categories of cancers (CCs 8 to 12), together. In our analysis, lung cancer (CC 8) and other severe cancers (CC 9) carried higher risks, so we separated them into a distinct risk variable and grouped other major cancers (CC 10), benign cancers (CC 11), and cancers of the urinary and GI tracts (CC 12) together. Consistent with other publicly reported measures, we also left metastatic cancer/leukemia (CC 7) as a separate risk variable.

Complications occurring during hospitalization are not comorbid illnesses, may reflect hospital quality of care, and therefore should not be used for risk adjustment. Hence, conditions that may represent adverse outcomes due to care received during the index hospital stay are not included in the risk-adjusted model; see Table 5 in Section 2a1.13. CCs on this list were not counted as a risk variable in our analyses if they appeared only on the index admission.

Service mix adjustment:

The measure includes many different discharge condition categories that differ in their baseline readmission risks. In addition, hospitals differ in their relative distribution of these condition categories (service mix). To adjust for service mix, the measure uses an indicator variable for the discharge condition category in addition to risk variables for comorbid conditions. The models include a condition-specific indicator for all condition categories with sufficient volume (defined as those with more than 1,000 admissions nationally in a given year for Medicare FFS data) as well as a single indicator for conditions with insufficient volume in each model.

Socioeconomic Status (SES) Factors and Race

SES factors and race for examination were 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.

SES and Race Variables and HF Readmission

To examine the relationship between SES and race variables and hospital 30-day, hospital-wide, all-cause, unplanned readmission following hospitalization, 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 (VA) databases as the primary data source, and articles not explicitly focused on SES or race and readmission across multiple conditions. One hundred and sixty nine articles were initially reviewed, and one hundred and fifty five studies were excluded from full-text review based on the above criteria. Studies indicate that SES/race variables were associated with increased risk of readmission across multiple major illnesses and conditions (Aseltine RH, et al., 2015; Mitchell SE, et al., 2012; Odonkor CA, et al., 2015; Herrin J, et al., 2015; Gu Q, et al., 2014, Kim H, et al., 2010; Kangovi S, et al., 2012; Iloabuchi TC, 2014; Beck AF, et al., 2012; Arbaje AI, et al., 2008; Hu J, 2014; Nagasako EM, et al., 2014; Joynt, KE, et al., 2013), though there may not be a significant effect on hospital-level profiling (Blum AB, et al., 2014).

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 ZJ, et al., 2015; Hu J, 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 AB, 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 M, et al., 2014; Joynt KE and Jha AK, 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 AK, et al., 2011; Reames BN, et al., 2014). Similarly African-American patients have been shown to have less access to high quality facilities compared with white patients (Skinner J, 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 AN, 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:

Arbaje AI, Wolff JL, Yu Q, Powe NR, Anderson GF, Boult C. Postdischarge environmental and socioeconomic factors and the likelihood of early hospital readmission among community-dwelling Medicare beneficiaries. The Gerontologist. 2008;48(4):495-504

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.

Beck AF, Simmons JM, Huang B, Kahn RS. Geomedicine: area-based socioeconomic measures for assessing risk of hospital reutilization among children admitted for asthma. American journal of public health. 2012;102(12):2308-2314.

Blum AB, NN 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.

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.

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.

Gu Q, Koenig L, Faerberg J, Steinberg CR, Vaz C, Wheatley MP. The Medicare Hospital Readmissions Reduction Program: potential unintended consequences for hospitals serving vulnerable populations. Health services research. 2014;49(3):818-837.

Herrin J, St Andre J, Kenward K, Joshi MS, Audet AM, Hines SC. Community factors and hospital readmission rates. Health services research. 2015;50(1):20-39.

Horwitz L, Partovian C, Lin Z, et al. Hospital-Wide All-Cause Unplanned Readmission Measure: Final Technical Report. 2012;

https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1228772504318

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.

Iloabuchi TC, Mi D, Tu W, Counsell SR. Risk factors for early hospital readmission in low-income elderly adults. Journal of the American Geriatrics Society. 2014;62(3):489-494.

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.

Kangovi S, Grande D, Meehan P, Mitra N, Shannon R, Long JA. Perceptions of readmitted patients on the transition from hospital to home. Journal of hospital medicine. 2012;7(9):709-712.

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

Kim H, Ross JS, Melkus GD, Zhao Z, Boockvar K. Scheduled and unscheduled hospital readmissions among patients with diabetes. The American journal of managed care. 2010;16(10):760-767.

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.

Krumholz HM, Lin Z, Drye EE, et al. An administrative claims measure suitable for profiling hospital performance based on 30-day all-cause readmission rates among patients with acute myocardial infarction. Circulation 2011;4(2):243-52.

Mitchell SE, Sadikova E, Jack BW, Paasche-Orlow MK. Health literacy and 30-day postdischarge hospital utilization. Journal of health communication. 2012;17 Suppl 3:325-338.Nagasako EM, Reidhead M, Waterman B, Dunagan WC. Adding socioeconomic data to hospital readmissions calculations may produce more useful results. Health affairs (Project Hope). 2014;33(5):786-791

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

Odonkor CA, Hurst PV, Kondo N, Makary MA, Pronovost PJ. Beyond the Hospital Gates: Elucidating the Interactive Association of Social Support, Depressive Symptoms, and Physical Function with 30-Day Readmissions. *American journal of physical medicine & rehabilitation / Association of Academic Physiatrists.* 2015;94(7):555-567

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

Reames BN, Birkmeyer NJ, Dimick JB, Ghaferi AA. Socioeconomic disparities in mortality after cancer surgery: failure to rescue. JAMA surgery 2014; 149:475-81.

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.

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

The final variables for each of the five risk models with associated odds ratios (**Dataset** 1) are shown in the attached Data Dictionary or Code Table 2.2b.

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 SDS factors in the HWR cohort varies across measured entities. The median percentage of Medicaid patients is 14.9% (interquartile range [IQR] 9.8%-22.6 %). The median percentage of African-American patients is 2.2% (IQR 0.0%-9.4%). The median percentage of low SES AHRQ indicator patients is 19.4% (IQR 5.0%-57.3%).

Empirical association with the outcome (bivariate)

The patient-level observed hospital wide readmission rate is higher for Medicaid patients, 19.3%, compared with 14.8% for all other patients. The readmission rate for African-American patients was also higher at 19.2% compared with 15.1% for patients of all other races. Similarly the readmission rate for patients in the lowest SES quartile by AHRQ Index was 16.8% compared with 15.1% for all other patients.

Incremental effect of SDS variables in a multivariable model

We then examined the strength and significance of the SDS variables in the context of a multivariable model. Consistent with the above findings, when we include any of these variables in a multivariate model that includes all of the claims-based clinical variables the effect size of each of these variables is small. We also find that the c-statistic is essentially unchanged with the addition of any of these variables into the model. Furthermore we find that 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 mean median absolute change in hospitals' RSRRs when adding a Medicaid indicator is 0.004% (interquartile range [IQR] -0.017% – 0.024%, minimum -0.309% – maximum 0.135%) with a correlation coefficient between RSRRs for each hospital with and without Medicaid added of 0.998. The median absolute change in hospitals' RSRRs when adding a race indicator is 0.011% (IQR -0.010% – 0.033%, minimum -0.671% – maximum 0.130%) with a correlation coefficient between RSRRs for each hospital with and without race added of 0.998. The median absolute change in hospitals' RSRRs when adding a low SES AHRQ indicator is 0.007% (IQR -0.033% – 0.036%, minimum -0.322% – maximum 0.135%) with a correlation coefficient between RSRRs for each hospital with and without low SES added of 0.997.

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

The patient-level and hospital-level effects were significantly associated with each of the hospital wide readmission models (Medicine, Surgery, Cardiorespiratory, Cardiovascular, and Neurology) in the decomposition analysis. If the dual eligible, race, or low AHRQ SES Index variables are used 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.

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.

HWR Decomposition Analysis

Parameter	Estimate (Standard Error)	P-value
Dual Eligible – Patient-Level – Medicine	0.0599 (0.00433)	<.0001
Dual Eligible – Hospital-Level – Medicine	0.3207 (0.0177)	<.0001
Dual Eligible – Patient-Level – Surgery	0.1483 (0.00794)	<.0001
Dual Eligible – Hospital-Level – Surgery	0.4743 (0.0332)	<.0001
Dual Eligible – Patient-Level – Cardio Respiratory	0.1043 (0.00634)	<.0001
Dual Eligible – Hospital-Level – Cardio Respiratory	0.4148 (0.0269)	<.0001
Dual Eligible – Patient-Level – Cardiovascular	0.1607 (0.0101)	<.0001
Dual Eligible – Hospital-Level – Cardiovascular	0.5318 (0.0418)	<.0001
Dual Eligible – Patient-Level – Neurology	0.0874 (0.0129)	<.0001
Dual Eligible – Hospital-Level – Neurology	0.4997 (0.0526)	<.0001
African American – Patient-Level – Medicine	0.0374 (0.00558)	<.0001
African American – Hospital-Level – Medicine	0.3208 (0.0119)	<.0001
African American – Patient-Level – Surgery	0.0959 (0.0103)	<.0001
African American – Hospital-Level – Surgery	0.4423 (0.0214)	<.0001
African American – Patient-Level – Cardio Respiratory	0.0470 (0.00884)	<.0001
African American – Hospital-Level – Cardio Respiratory	0.3386 (0.0186)	<.0001
African American – Patient-Level – Cardiovascular	0.0763 (0.0131)	<.0001
African American – Hospital-Level – Cardiovascular	0.3501 (0.0269)	<.0001
African American – Patient-Level – Neurology	0.1200 (0.0155)	<.0001
African American – Hospital-Level – Neurology	0.5252 (0.0331)	<.0001
AHRQ SES Index – Patient-Level – Medicine	0.0249 (0.00444)	<.0001
AHRQ SES Index – Hospital-Level – Medicine	0.0788 (0.00653)	<.0001

AHRQ SES Index – Patient-Level – Surgery	0.0349 (0.00689)	<.0001
AHRQ SES Index – Hospital-Level – Surgery	0.1254 (0.0120)	<.0001
AHRQ SES Index – Patient-Level – Cardio Respiratory	0.0376 (0.00661)	<.0001
AHRQ SES Index – Hospital-Level – Cardio Respiratory	0.1105 (0.00910)	<.0001
AHRQ SES Index – Patient-Level – Cardiovascular	0.0307 (0.00943)	0.0011
AHRQ SES Index – Hospital-Level – Cardiovascular	0.1375 (0.0149)	<.0001
AHRQ SES Index – Patient-Level – Neurology	0.0544 (0.0125)	<.0001
AHRQ SES Index – Hospital-Level – Neurology	0.1314 (0.0198)	<.0001

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 (**Dataset 1** & **Dataset 2**)
We tested the performance of the model for **Dataset 1 & Dataset 2** described in section 1.7. We computed three summary statistics for assessing model performance (Harrell and Shih, 2001) for the development and validation cohort: ***Discrimination Statistics*** (1) Area under the receiver operating characteristic (ROC) curve (the c-statistic) is the probability that predicting the outcome is better than chance, which is a measure of how accurately a statistical model is able to distinguish between a patient with and without an outcome. (2) Predictive ability (discrimination in predictive ability measures the ability to distinguish high-risk subjects from low-risk subjects; therefore, we would hope to see a wide range between the lowest decile and highest decile.) ***Calibration Statistics* (Dataset 2)** (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.) References: Harrell FE and Shih YCT. 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*)**:**

Medicine Cohort Model Discrimination

Indices	2007-2008 Development Sample	2007-2008 Validation Sample	2009 Validation Sample	2015 HWR Data for Public Reporting
Number of hospital stays	3,085,962	3,082,357	3,032,518	2,864,028
Number of hospitals	4,954	4,946	4,908	4,713
Unadjusted readmission rate	18.0%	18.0%	18.1%	17.1%
Discrimination - Predictive Ability (lowest decile %, highest decile %)	9 – 34	9 – 34	7 – 36	9-33
Discrimination – c statistic	0.640	0.641	0.663	.643

Surgical Cohort Model Discrimination

Indices	2007-2008 Development Sample	2007-2008 Validation Sample	2009 Validation Sample	2015 HWR Data for Public Reporting
Number of hospital stays	2,208,753	2,208,482	2,109,292	1,695,227
Number of hospitals	4,354	4,353	4,232	4,031
Unadjusted readmission rate	12.6%	12.6%	12.6%	11.1%
Discrimination - Predictive Ability (lowest decile %, highest decile %)	4 – 27	4 – 27	3 – 30	5-27
Discrimination – c statistic	0.675	0.675	0.699	0.675

Cardiorespiratory Cohort Model Discrimination

Indices	2007-2008 Development Sample	2007-2008 Validation Sample	2009 Validation Sample	2015 HWR Data for Public Reporting
Number of hospital stays	1,396,562	1,396,855	1,331,539	1,144,451
Number of hospitals	4,810	4,806	4,718	4,596
Unadjusted readmission rate	21.1%	21.2%	21.4%	19.5%
Discrimination - Predictive Ability (lowest decile %, highest decile %)	11 – 37	11 – 37	9 – 40	10-35
Discrimination – c statistic	0.630	0.631	0.657	0.636

Cardiovascular Cohort Model Discrimination

Indices	2007-2008 Development Sample	2007-2008 Validation Sample	2009 Validation Sample	2015 HWR Data for Public Reporting
Number of hospital stays	860,485	861,925	809,520	707,529

Number of hospitals	4,702	4,703	4,641	4,438
Unadjusted readmission rate	15.2%	15.2%	15.4%	14.4%
Discrimination - Predictive Ability (lowest decile %, highest decile %)	5 – 31	6 – 30	5 – 33	7-31
Discrimination – c statistic	0.657	0.656	0.680	0.658

Neurology Cohort Model Discrimination

Indices	2007-2008 Development Sample	2007-2008 Validation Sample	2009 Validation Sample	2015 HWR Data for Public Reporting
Number of hospital stays	461,225	461,262	452,743	432,573
Number of hospitals	4,699	4,686	4,609	4,426
Unadjusted readmission rate	14.7%	14.7%	14.6%	13.1%
Discrimination - Predictive Ability (lowest decile %, highest decile %)	8 – 27	8 – 26	6 – 29	8-26
Discrimination – c statistic	0.614	0.613	0.646	0.622

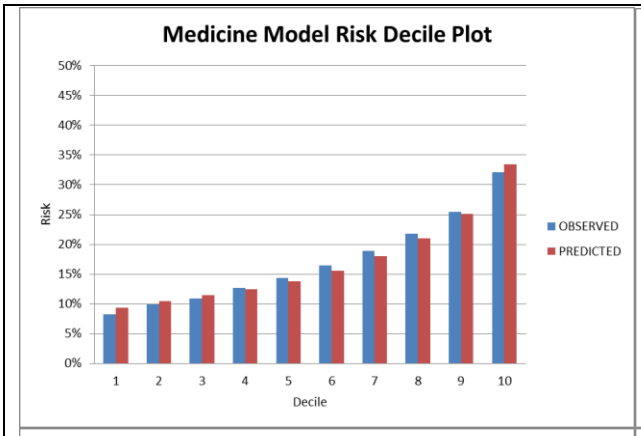
2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

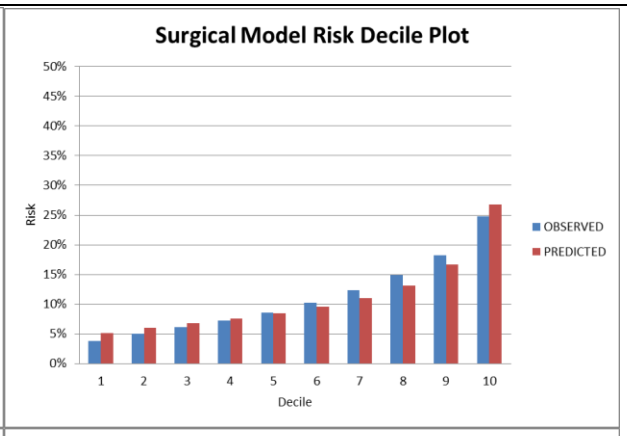
Model Calibration (γ0, γ1) (Dataset 2)

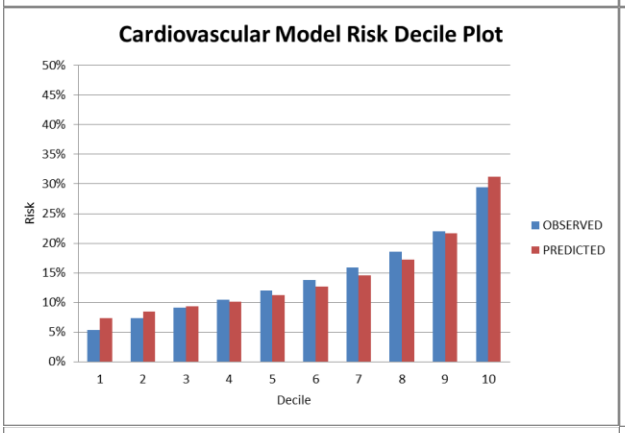
Indices	2007-2008 Development Sample	2007-2008 Validation Sample	2009 Sample
Medicine Cohort	(0, 1)	(0.011, 1.006)	(0.132, 1.118)
Surgical Cohort	(0, 1)	(-0.012, 0.995)	(0.104, 1.076)
Cardiorespiratory	(0, 1)	(0.010, 1.006)	(0.193, 1.184)
Cardiovascular Cohort	(0, 1)	(-0.019, 0.993)	(0.145, 1.109)
Neurology Cohort	(0, 1)	(-0.036, 0.982)	(0.201, 1.163)

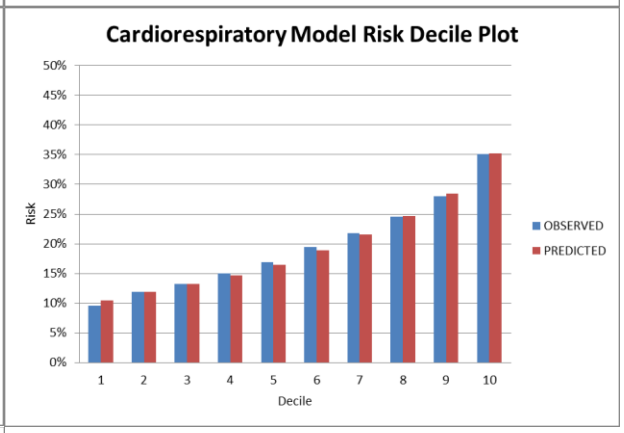
2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

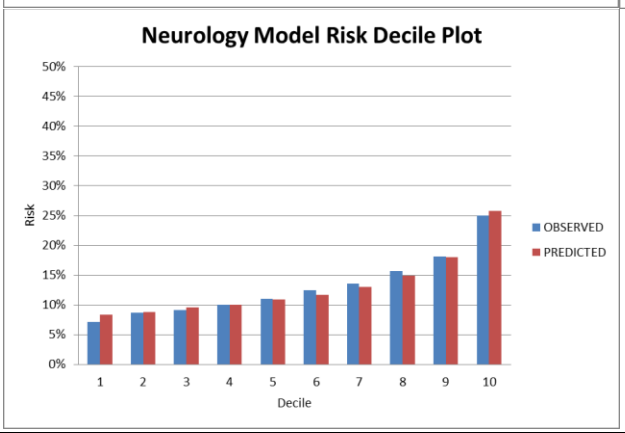
The risk decile plot is a graphical depiction of the deciles calculated to measure predictive ability. Below, we present the risk decile plot showing the distributions for Medicare FFS data from July 2013 to June 2014.











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-statistics indicate fair discrimination for each of the models in **Datasets 1 and 2**. Each of the models 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 one, there is potential evidence of over-fitting. The calibration values close to 0 at one end and close to 1 to the other end indicate good calibration of each of the models.

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 good discrimination of the model and good predictive ability.

Overall Interpretation
Interpreted together, our diagnostic results demonstrate that 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*)

N/A

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 one-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*)

In the 2015 public reporting year (**Dataset 1**), out of 4,772 hospitals in the U.S., 178 performed "better than the U.S. national rate," 4,078 performed "no different from the U.S. national rate," and 337 performed "worse than the U.S. national rate." One hundred and seventy-nine hospitals were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing.

Note that this analysis included index admissions from July 2011 – June 2014 from the 2015 public reported data (**Dataset 1**). We used the planned readmission algorithm version 3.0 for measure calculation in these data. The planned readmission algorithm 4.0 will first be applied in the 2016 publically reported measure results.

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 that differences in the quality of care received across hospitals for the Hospital-Wide All-Cause Unplanned Readmission Measure (HWR) remain, which support continued measurement in order to reduce variation.

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.**

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)
	Public Reporting Hospital Inpatient Quality Reporting (IQR) Program http://cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.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 reporting 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 IPPS non-federal acute care hospitals and Veteran Affairs (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 4,772 hospitals across the U.S. The final index cohort includes 6,843,808 admissions.

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 unplanned, all-cause readmissions. The median 30-day RSRR decreased by 0.7 absolute percentage points from the 2013 public reporting period (median RSRR: 15.9%) to the 2015 public reporing period (median RSRR: 15.2%).

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)**

0171 : Acute Care Hospitalization During the First 60 Days of Home Health
0173 : Emergency Department Use without Hospitalization During the First 60 Days of Home Health
0329 : Risk-Adjusted 30-Day All-Cause Readmission Rate
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
0695 : Hospital 30-Day Risk-Standardized Readmission Rates following Percutaneous Coronary Intervention (PCI)
1551 : Hospital-level 30-day all-cause risk-standardized readmission rate (RSRR) following elective primary total hip arthroplasty (THA) and total knee arthroplasty (TKA)
1768 : Plan All-Cause Readmissions (PCR)
1891 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization

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?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
This measure and the National Committee for Quality Assurance (NCQA) Plan All-Cause Readmissions (PCR) Measure #1768 are related measures, but are not competing because they don't have the same measure focus and same target population. In addition, both have been previously harmonized to the extent possible under the guidance of the National Quality Forum Steering Committee in 2011. Each of these measures has different specifications. NCQA's Measure #1768 counts the number of inpatient stays for patients aged 18 and older during a measurement year that were followed by an acute readmission for any diagnosis to any hospital within 30 days. It contrasts this count with a calculation of the predicted probability of an acute readmission. NCQA's measure is intended for quality monitoring and accountability at the health plan level. This measure estimates the risk-standardized rate of unplanned, all-cause readmissions to a hospital for any eligible condition within 30 days of hospital discharge for patients aged 18 and older. The measure will result in a single summary risk-adjusted readmission rate for conditions or procedures that fall under five specialties: surgery/gynecology, general medicine, cardiorespiratory, cardiovascular, and neurology. This measure is specified for evaluating hospital performance. However, despite these differences in cohort specifications, both measures under NQF guidance have been harmonized to the extent possible through modifications such as exclusion of planned readmissions. We did not include in our list of related measures any non-outcome (e.g., process) measures with the same target population as our measure. Because this is an outcome measure, clinical coherence of the cohort takes precedence over alignment with related non-outcome measures. Furthermore, non-outcome measures are limited due to broader patient exclusions. This is because they typically only include a specific subset of patients who are eligible for that measure (for example, patients who receive a specific medication or undergo a specific procedure).

5b. Competing Measures

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

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
N/A

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** 2015_Measures_Reevaluation_Hospital-Wide_Readmission_AUS_Report_FINAL_508_Compliant_01-29-16_v1.0.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services (CMS)
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.

The working group involved in the initial measure development is detailed in the original technical report available at www.qualitynet.org.

Our measure development team consisted of the following members:
Leora Horwitz, MD, MHS
Chohreh Partovian, MD, PhD
Zhenqiu Lin, PhD
Jeph Herrin, PhD
Jacqueline Grady, MS
Mitchell Conover, BA
Julia Montague, MPH
Chloe Dillaway, BA
Kathleen Bartczak, BA
Lisa Suter, MD, MHS
Joseph Ross, MD, MHS
Susannah Bernheim, MD, MHS
Harlan Krumholz, MD, SM
Elizabeth Drye, MD, SM

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2009
Ad.3 **Month and Year of most recent revision:** 09, 2012
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?** 01, 2017

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

Ad.8 **Additional Information/Comments:** N/A

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 #: 2879
Measure Title: Hybrid Hospital-Wide Readmission Measure with Claims and Electronic Health Record Data
Measure Steward: Centers for Medicare & Medicaid Services (CMS)
Brief Description of Measure: The measure estimates a hospital-level risk-standardized readmission rate (RSRR) of unplanned, all-cause readmission after admission for any eligible condition within 30 days of hospital discharge. The measure reports a single summary RSRR, derived from the volume-weighted results of five different models, one for each of the following specialty cohorts based on groups of discharge condition categories or procedure categories: surgery/gynecology, general medicine, cardiorespiratory, cardiovascular, and neurology, each of which will be described in greater detail below. The measure also indicates the hospital-level standardized readmission ratios (SRR) for each of these five specialty cohorts. The outcome is defined as unplanned readmission for any cause within 30 days of the discharge date for the index admission (the admission included in the measure cohort). A specified set of planned readmissions do not count in the readmission outcome. The target population is Medicare Fee-for-Service beneficiaries who are 65 years or older.
This Hybrid Hospital-Wide Readmission (HWR) measure is a re-engineered version of measure 1789, the Hospital-Wide All-Cause Unplanned Readmission Measure which was developed for patients 65 years and older using Medicare claims and is currently publically reported in the Hospital Inpatient Quality Reporting Program. This reengineered measure uses clinical data elements from patients' electronic health records in addition to claims data for risk adjustment.
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 all cause unplanned readmission rates among Medicare beneficiaries 65 years and older admitted to all non-federal US acute care hospitals. 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 their patient case mix and hospital service mix, and therefore promote hospital quality improvement and better inform consumers about care quality.

Hospital-wide readmission is a priority area for outcomes 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.

This Hybrid HWR measure incorporates both data from claims as well as clinical data elements pulled from the EHR in risk adjustment of the readmission models. Some benefits of including the clinical data elements are:

1. Inclusion of patient-level clinical data related to severity of illness is responsive to providers who continue to express preference for using patient-level clinical data, and provides an opportunity to incorporate clinical data into outcome measures.

2. Hospitals will increasingly use EHR data to assess severity of illness and patients' risk of poor outcomes. This provides an opportunity to align the measure with clinical decision support systems that many providers utilize to alert care teams about patients at increased risk of poor outcomes in real time during the inpatient stay.

3. Collecting a simple core set of clinical data elements that perform well as risk-adjustment variables (for illness severity) across conditions can greatly reduce the cost and effort of future measure development, improve harmonization, and create opportunity for longitudinal assessment of patient status and quality of care across settings.

4. These core clinical data elements will provide measure developers with a standard set of reliable data that can be used as a starting place when building risk-adjustment models for quality measures using clinical data.

Numerator Statement: The outcome for this measure is 30-day readmission. We define readmission as an inpatient admission for any cause, with the exception of certain planned readmissions, within 30 days from the date of discharge from an eligible index admission. If a patient has more than one unplanned admission (for any reason) within 30 days after discharge from the index admission, only one is counted as a readmission. The measure looks for a dichotomous yes or no outcome of whether each admitted patient has an unplanned readmission within 30 days. However, if the first readmission after discharge is considered planned, any subsequent unplanned readmission is not counted as an outcome for that index admission because the unplanned readmission could be related to care provided during the intervening planned readmission rather than during the index admission.
Denominator Statement: The measure includes admissions for Medicare beneficiaries who are 65 years and older and are discharged from all non-federal, acute care inpatient US hospitals (including territories)with a complete claims history for the 12 months prior to admission.

Additional details are provided in S.9 Denominator Details.
Denominator Exclusions: The measure excludes index admissions for patients:

1. Admitted to Prospective Payment System (PPS)-exempt cancer hospitals;
2. Without at least 30 days post-discharge enrollment in FFS Medicare;
3. Discharged against medical advice (AMA);
4. Admitted for primary psychiatric diagnoses;
5. Admitted for rehabilitation; or
6. Admitted for medical treatment of cancer.

Measure Type: Outcome
Data Source: Administrative claims, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Laboratory
Level of Analysis: Facility

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

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

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

De.4. **IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** N/A

New Measure -- Preliminary Analysis

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.

- This hybrid measure estimates a hospital-level risk-standardized readmission rate (RSRR) for unplanned readmission for any eligible condition within 30 days of hospital discharge, using both claims and electronic health record data (EHR).
- As a rationale for measuring this health outcome, the developer suggests that hospitals are able to influence readmission rates through a broad range of clinical activities including communication between providers, prevention of, and response to, complications, patient safety and coordinated transitions to the outpatient

environment.

- The hybrid measure includes data from both claims and clinical data elements from EHR. The developer cites some benefits of using clinical data elements:
    - "Inclusion of patient-level clinical data related to severity of illness is responsive to providers who continue to express preference for using patient-level clinical data, and provides an opportunity to incorporate clinical data into outcome measures."
    - "Hospitals will increasingly use EHR data to assess severity of illness and patients' risk of poor outcomes. This provides an opportunity to align the measure with clinical decision support systems that many providers utilize to alert care teams about patients at increased risk of poor outcomes in real time during the inpatient stay."
    - "Collecting a simple core set of clinical data elements that perform well as risk-adjustment variables (for illness severity) across conditions can greatly reduce the cost and effort of future measure development, improve harmonization, and create opportunity for longitudinal assessment of patient status and quality of care across settings."
    - "These core clinical data elements will provide measure developers with a standard set of reliable data that can be used as a starting place when building risk-adjustment models for quality measures using clinical data."

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

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 provide performance data for that they used for model development purposes only, using Dataset 1, which contains inpatient claims with clinical data elements derived from patients' EHRs.
- This data included 381,980 admissions at 21 hospitals. The mean RSRR was 14.84%, with a minimum of 13.15% and a maximum RSRR of 16.16%.
- The developers explained that, "CMS currently publicly reports a claims-based Hospital-Wide All-Cause Unplanned Readmission Measure (NQF #1789). The results for this measure, as reported in the 2015 update to the Hospital Compare website, are based on RSRRs calculated for admissions among Medicare FFS patients aged 65 and older from July 1, 2012 - June 30, 2013. It includes 4,772 hospitals. The median hospital RSRR was 15.5%, with an interquartile range of 11.0% to 21.4%."

Disparities

- The developers did not perform disparities analyses measure as specified due to the small number of hospitals and lack of diversity with respect to SES in the Kaiser Permanente of Northern California system. They conducted disparities analysis for the claims-only HWR measure (NQF #1789).
- For Measure #1789, the developers provided performance scores for hospitals serving a low proportion of dual eligible patients vs. those serving a high proportion of dual eligible patients, performance scores for hospitals serving a low proportion of African-American patients vs. those serving a high proportion of African-American patients and performance scores for hospitals serving a low proportion of patients with AHRQ SES Index Score index score equal to or below 45.9 vs those serving a high proportion of patients with an AHRQ SES index score equal to or below 45.9.

- By proportion of **Dual Eligible Patients:**

// Low proportion (=9.8%) Dual Eligible patients//Hospitals with a high proportion (=22.6%) Dual Eligible

patients
Number of Measured Hospitals// 1,257 // 1,219
Number of Patients// 2,137,895 patients in low-proportion hospitals // 927,007 in high-proportion hospitals
Maximum// 18.7 // 20.1
90th percentile// 16.2 // 16.8
75th percentile// 15.7 // 16.0
Median (50th percentile)// 15.3 // 15.6
25th percentile// 14.8 // 15.2
10th percentile// 14.3 // 14.9
Minimum // 11.5 // 12.2

- By proportion of **African-American Patients**:

  // Low proportion (=2.2%) African-American patients//Hospitals with a high proportion (=9.4%) African-American patients
  Number of Measured Hospitals// 1,156 // 1,180
  Number of Patients// 222,648 patients in low-proportion hospitals/ 2,294,715 in high-proportion hospitals
  Maximum// 19.1 // 19.9
  90th percentile// 16.0 // 17.1
  75th percentile// 15.6 // 16.3
  Median (50th percentile)// 15.4 // 15.7
  25th percentile// 15.1 // 15.2
  10th percentile// 14.8 // 14.8
  Minimum // 12.9 // 12.2

- By Proportion of Patients with AHRQ SES Index Scores Equal or Below 45.9:

  // Low proportion of patients below AHRQ SES index score of 45.0 (=5.0%)// Hospitals with a high proportion of patients below AHRQ SES index score of 45.0 (=57.1%)

  Number of Measures Hospitals// 1,209 // 1,217
  Number of Patients// 1,651,852 patients in hospitals with low proportion of patients below AHRQ SES index score of 45.0 //795,899 patients in hospitals with high proportion of patients below AHRQ SES index score of 45.0
  Maximum// 19.9 // 20.1
  90th percentile// 16.2 // 16.6
  75th percentile// 15.7 // 16.0
  Median (50th percentile)// 15.3 // 15.5
  25th percentile// 14.9 // 15.2
  10th percentile// 14.5 // 14.8
  Minimum // 11.5 // 13.0

- The developer does not provide interpretation or analysis of these data.

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)

1. Importance to Measure and Report

1a. Evidence to Support Measure Focus

Comments: **Although the measure is based upon a "non hybrid model" using claims data, does the introduction of the CCDE make it less likely to be relevant across multiple provider settings with varying EMRs?

**Rationale for inclusion of clinical data from patient EMR-related to illness severity is to build better risk adjustment models.

**In fact, 2879 is a new "flavor" of measure 1789. 2879 proposes to add clinical data to the administrative claims data already used for 1789.

This measure is designed to identify hospitals that perform better or worse than expected based on patient case mix/hospital service mix. By using clinical data to augment risk-adjustment and severity of illness data, provider communication, complication identification and avoidance techniques and transition of care activities can be isolated at top performing facilities to improve outcomes at all hospitals.

1b. Performance Gap

Comments: **The case was well thought out concerns are based upon KP data only where data in the CCDE is normalized. I did not see disparities specifically addressed and SES was not part of the measure in the original or hybrid state.

**Gap cited parallels that for measure #1789; no disparities analysis were done due to small number of hospitals (21) and lack of diversity in those hospitals (Kaiser Northern CA).

**The developer cited existing performance results from measure 1789. Admissions for Medicare FFS patients between 7/1/12 and 6/30/13 yielded a median RSRR of 15.5%.

The developer also cited disparities data from measure 1789. Data was provided for hospitals with a high share of dual-eligibles, African American patients and AHRQ SES index score.

The developer did not provide interpretation of any of the disparities results. However, each cohort of hospitals with larger proportions of at-risk populations demonstrated higher maximum and minimum RSRRs.

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

Comments: **yes

**n/a

Criteria 2: Scientific Acceptability of Measure Properties
2a. Reliability
2a1. Reliability Specifications

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): Administrative claims, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Laboratory

Specifications:

- This measure calculates 30-day readmissions for patients with an eligible index admission.
- This Hybrid Hospital-Wide Readmission (HWR) measure is a re-engineered version of measure 1789, the Hospital-Wide All-Cause Unplanned Readmission Measure which was developed for patients 65 years and older using Medicare claims and is currently publically reported in the Hospital Inpatient Quality Reporting Program. This reengineered measure uses clinical data elements from patients' electronic health records in addition to claims data for risk adjustment.
- The measure produces a risk-standardized readmission rate (RSRR), which is calculated as the ratio of the number of "predicted" to the number of "expected" readmission at a given hospital, multiplied by the national observed readmission rate.
- The denominator includes Medicare beneficiaries who are 65 years and older and are discharged from all non-federal, acute care inpatient US hospitals (including territories) with a complete claims history for the 12 months prior to admission.
- The numerator includes patients were readmitted to any acute care hospital for any cause within 30 days of the date of discharge of the index admission, excluding planned readmissions
- The denominator population is defined using ICD-9 and ICD-10 codes; a list of applicable codes is included in the submission.

- The measure aggregates the ICD-9 principal diagnosis and all procedure codes of the index admission into clinically coherent groups of conditions and procedures (condition categories or procedure categories) using the Agency for Healthcare Research and Quality (AHRQ) Clinical Classifications System (CCS).
- The data sources for this measure may include Medicare Part A and B claims, the Medicare Enrollment Database and electronic clinical data. The datasets used for testing included Medicare Parts A and B claims inpatient claims, as well as electronically and manually abstracted electronic health record (EHR) data from several health systems.
- The measure's time window is based on one year of data.
- The measure is risk-adjusted using a statistical risk model (see details below).

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?*

eMeasure Technical Advisor review:

Submitted measure is an HQMF compliant eMeasure	The submitted eMeasure specifications follow the industry accepted format for eMeasure (HL7 Health Quality Measures Format (HQMF)). **HQMF specifications** ☒ **Yes** ☐ **No**
Documentation of HQMF or QDM limitations	Submitted eMeasure contains components that cannot be represented due to limitations of HQMF or QDM and the submission explains the work around for these limitations; This is a hybrid measure using a combination of claims and EHR data to identify the measure population. The eMeasure portion of the measure extracts a set of clinical data elements from the measure and uses them to risk adjust a hospital outcome measure.
Value Sets	The submitted eMeasure specifications uses existing value sets when possible and uses new value sets that have been vetted through the VSAC
Measure logic is unambiguous	Submission includes test results from a data set demonstrating the measure logic can be interpreted precisely and unambiguously
Feasibility Testing	The submission contains a feasibility assessment that addresses data element feasibility and follow-up with measure developer indicates that the measure logic is feasible based on assessment by EHR vendors.

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

- Datasets used for testing included Medicare Parts A and B claims and the Medicare Enrollment Database (EDB). Additionally, census data were used to assess socio-demographic factors.
- **Data element reliability**:

- With regard to data element reliability, the developer notes that the measure has been developed to avoid the use of claims data elements that are thought to be coded inconsistently across hospitals or providers, instead using fields that are consequential for payment and which are audited by CMS.
- In addition, the developer compared frequencies and odds ratios of variables from their risk model across three years of data in order to assess the consistency of those variables over time.

- **Performance score reliability:**
  - For test-retest reliability, the developer used the randomly split development and validation samples and calculated the measure for each hospital separately in each sample.
  - Each hospital is measured twice, but each measurement is made using an entirely distinct set of patients. The developers note that 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 the developer calculated the intra-class correlation coefficient (ICC) (Shrout and Fleiss, 1979), and assessed the values according to conventional standards (Landis and Koch, 1977).
  - The developer notes that because reliability of the measure result could only be tested in a small sample of hospitals (n=21) and admissions, testing will be repeated in a larger nationally representative set of hospitals prior to implementation. This testing will depend on implementation of hospital reporting of the core clinical data elements used in the measure's risk models.

Results of reliability testing

- **Data element reliability**:
  - Summarizing the results of this analysis, the developer notes that the the frequency of model variables remained relatively constant between 2010 and 2012, with no model variables increasing or decreasing by more than 2%.
  - The developer notes the stability over time of the odds ratios or variable coefficients and the model variable frequencies and rates of capture for clinical data elements suggests that the underlying data elements are reliable.
- **Clinical Data Element Capture**
  - The developer provided the rate of capture of all of the clinical data elements used in the measures' risk models in the development, validation, and 2012 samples by cohort in table 1 (below).
  - Coefficients for the claims and clinical risk variables for all three samples can be found in data field S.2b.
- **Performance score reliability:**
  - There were a total of 81,589 admissions in the development sample, and 79, 813 in the validation sample. Two risk-standardized readmission rates (RSRR) were calculated for each hospital: one from each of the two separate samples.
    - The agreement between the two RSRRs for each hospital (as measured by an intra-class correlation coefficient (ICC)) was 0.688; the developer states that according to the conventional interpretation, this is considered a "moderate" level of agreement.

Guidance from the Reliability Algorithm

- Question 1. Submitted specifications are precise, unambiguous, and complete. Measure can be consistently implemented.
- Question 2. Empirical reliability testing was conducted using statistical tests with the measure as specified.
- Question 3. Empirical validity testing of patient-level data was conducted.
- Question 4. Reliability testing was conducted with computed performance measure scores for each measured entity.
- Question 5. Random split-half correlation was used to assess the proportion of variability due to real differences among the measured entities.
- Question 6. The ICC was 0.688 which is considered a moderate level of agreement.

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?*

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

2b. Validity

2b1. Validity: Specifications

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

- This measure estimates 30-day readmissions for any cause with the exclusion of certain planned readmissions.
- As a rationale for measuring this health outcome, the developers suggest that hospitals are able to influence readmission rates through a broad range of clinical activities, including ensuring patients are ready for discharge, reducing the risk of infection, reconciling medications, improving communication among providers, promoting disease management, and educating patients.

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.

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 of EHR data elements:**
  - Several critical clinical data elements used in the measure's risk models were derived from patients' electronic medical records. When this measure is implemented, CMS intends to obtain these critical data elements from hospital EHRs and merge the data with claims data to calculate and report measure results.
  - The developer tested the validity of electronic extraction of these critical data elements as part of a more comprehensive evaluation of a larger set of core clinical data elements (CCDEs). The CCDE are a set of 21 EHR data elements that are captured on most adults (plus Troponin, which is a condition-specific CCDE for patients with acute myocardial infarction) admitted to acute care hospitals, are easily extracted from EHRs, and can be used to risk adjust hospital outcome measures for a variety of conditions and procedures. All of the critical data elements used in the Hybrid HWR measure are included in the CCDE.
  - Testing of the CCDE involved three phases: 1) identification of potentially feasible clinical data through qualitative assessment, 2) empirical feasibility testing of several clinical data elements electronically extracted from two large multi-facility health systems, and 3) validity testing of the CCDE at an additional health system.
- Phase 1: Identification of potentially feasible clinical data through qualitative assessment
  - To identify the CCDEs for risk adjustment of hospital outcome measures for adult patients, the developer first conducted a qualitative assessment of the reliable capture, accuracy, and extractability of categories and subcategories of clinical data as defined by the Quality Data Model (QDM) (e.g., vital signs,

laboratory test results).
    - The developer established a set of criteria to assess the consistency of data capture, relevance to hospital quality measures, and extractability from health records that are aligned with those established in the NQF's eMeasure Feasibility Assessment Report as well as the NQF feasibility criteria.
    - The developer convened a technical expert panel (TEP) to apply these criteria to categories and subcategories (data types) of clinical data based on the Quality Data Model (QDM).
    - Data categories and subcategories were rated on each feasibility criterion independently by TEP members. The ratings were tallied and TEP members met to discuss and resolve areas of disagreement. Through this process the TEP identified a list of data subcategories that were potentially feasible for use in hospital outcome measures. The CCDE were derived from only those subcategories for which the TEP reached consensus agreement on feasibility.
    - The TEP identified seven subcategories of EHR data that they considered feasible for adult hospitalized patients. They were: Encounter Performed, Patient Characteristics including birth date and sex, Physical Examination Findings for vital signs only, Diagnostic Study Order, Diagnostic Study Performed, Medication Discharge, and Laboratory Test Result.
    - The developers limited the CCDE to data elements to only four categories: Encounter Performed, Patient Characteristics, Physical Examination Findings for vital signs only, and Laboratory Test Results, which are unlikely to be reflective of care quality and therefore are thought to be both feasible to extract and appropriate for risk adjustment.
- Phase 2: Empirical feasibility testing using a large multi-site database
    - The developer examined all admissions in Dataset 1 between 2010 and 2011 and analyzed clinical data elements to determine whether they were captured in a numerical field, the consistency and timing of capture, and the accuracy of the data elements.
    - The developer tested several data elements that met the feasibility criteria in models predicting 30-day mortality following admission for several common medical conditions. The complete list of 21 (plus Troponin) CCDE were derived from these analyses.
    - The consistency of data capture of the critical data elements included in the Hybrid HWR measure for all adult hospitalized patients in two health systems with different EHR environments (EPIC and Cerner) are shown in tables 2 and 3 (below).
- Phase 3: Validity testing at two hospital sites the CCDE (including critical data elements for the Hybrid HWR measure)
    - The developer developed electronic specifications (e-specifications) using the Measure Authoring Tool (MAT), and analyzed extracted data from EHRs and assessed the ability of hospitals to use the e-specifications to query and electronically extract CCDEs from the EHR, for all adult inpatient admissions occurring over the course of one year.
    - Validity testing assessed the accuracy of the electronically extracted CCDEs compared to the same CCDEs gathered through manual abstraction (from the EHR) in a subset of 368 charts identified in the data query in Dataset 2, and 391 charts identified in the data query in Dataset 3.
    - Chart Abstraction: The developer calculated the number of admissions that needed to be randomly sampled from the EHR dataset and manually abstracted to yield a statistical margin of error (MOE) of 5% and a confidence level of 95% for the match rates between the two data sources. Sites then used an Access-based manual abstraction tool provided (along with training) to manually abstract the CCDEs from the random samples of the medical records identified through the EHR data query.
    - Validity Testing: The developer was only interested in in the case where the electronic abstraction value exactly matched the manual abstraction value. The developer counted only exact matches in the data value as well as the time and date stamp associated with that value when we calculated the match rate. The 95% confidence level was established based on the sample size and reflects the exact match rate using these criteria.
    - Table 5 demonstrates the comparison between electronic and manual abstraction of data in the two health systems.
- Validation Against Other Risk Models and Registry Data
    - The hybrid model uses a combination of claims data (demographics, comorbidities, and patient medical history) and electronic clinical data (laboratory results and vital signs).
    - The developer compared the Hybrid risk model to the harmonized claims-only risk model used in the publicly reported Hospital-Wide All Cause Unplanned Readmission Measure.

- Measure validity was tested through comparison of this Hybrid risk adjustment model with claims-only risk-adjustment model, and through use of established measure development guidelines.
- The developer estimated hospital-level RSRRs using the corresponding hierarchical logistic regression for each of the models in the linked patient sample.
- The developer then examined the linear relationship between the estimates using regression techniques and weighting by the total number of cases in each hospital.
- The Pearson correlation coefficient of the standardized rates from the claims-only risk-adjustment model and the Hybrid risk-adjustment model in the Development Sample of Dataset 1 is **0.9902.**
- Validity Indicated by Established Measure Development Guidelines: the developer notes that this measure was developed in consultation with national guidelines for publicly reported outcomes measures, with outside experts, and with the public.
- Validity as Assessed by External Groups: the developer solicited public comments on the measure through the CMS site. The resulting input was taken into consideration during the final stages of development and contributed to minor modifications to the measure.

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:

- Patients in the following categories are excluded from the measure:
  - Admitted to Prospective Payment System (PPS)-exempt cancer hospitals;
  - Without at least 30 days post-discharge enrollment in FFS Medicare;
  - Discharged against medical advice (AMA);
  - Admitted for primary psychiatric diagnoses;
  - Admitted for rehabilitation; or
  - Admitted for medical treatment of cancer.
- To determine the impact of exclusions, the developer examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion.
- The number and percentage of patients excluded for each criterion are as follows:
  - Discharged against medical advice (AMA); **679 (0.27%)**
  - Admitted for cancer treatment; **6,356 (2.53%)**
  - Admitted for primary psychiatric diagnoses; **593 (0.24%)**
  - Admitted for rehabilitation; 885 (.35%)

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *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** ☒ **No**

Risk adjustment summary

- The measure employs a hierarchical logistic regression model (a form of hierarchical generalized linear model [HGLM]) to create a hospital-level 30-day risk-standardized readmission rate (RSRR).
- The developer notes that this approach simultaneously models data at the patient and hospital levels to account for the variance in patient outcomes both within and between hospitals.
- Admissions are assigned to one of five mutually exclusive specialty cohort groups consisting of related conditions or procedures. For each specialty cohort group, the standardized readmission ratio (SRR) is calculated as the ratio of the number of "predicted" readmissions to the number of "expected" readmissions at a given hospital. For each hospital, the numerator of the ratio is the number of readmissions within 30 days predicted based on the hospital's performance with its observed case mix and service mix, and the denominator is the number of readmissions expected based on the nation's performance with that hospital's case mix and service mix. This approach is analogous to a ratio of "observed" to "expected" used in other types of statistical analyses.
- To select candidate variables for the Hybrid risk model, the developer began with the list of all administrative claims-based risk-adjustment variables included in the currently publicly reported Hospital-Wide All Cause Unplanned Readmission Measure.
- The developer also used the core clinical data elements CCDE, the EHR-derived data elements used in the measure.
- To adjust for service mix, the measure uses an indicator variable for the discharge condition category in addition to risk variables for comorbid conditions. The models include a condition-specific indicator for all condition categories with sufficient volume (defined as those with more than 1,000 admissions nationally in a given year for Medicare FFS data) as well as a single indicator for conditions with insufficient volume in each model.
- Although the 5 risk models use a common set of claims variables, the CCDE variables are not the same across specialty cohort models. Only those data elements that are statistically significant in each individual model are included.
- **Conceptual analysis of the need for SDS adjustment:**
    - There are at least four potential pathways for SDS factors to affect 30-day readmission rates:
        - One potential pathway is the relationship to health status at the time of admission. SDS factors 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. The developers note that this pathway should be largely accounted for by their clinical risk-adjustment model.
        - The next potential path way is that patients with low income and African-American patient are more likely to be seen in lower quality hospitals, which can contribute to increased risk of readmission.
        - The third major pathway is that a patient's race or SDS status cause them to experience differential, lower quality care or may not receive the differentiated care they require.
        - Finally, some SES risk factors may affect the likelihood of readmission without directly affecting health status at admission or the quality of care received during the hospitalization. Patients may have worse outcomes due to competing economic priorities or a lack of access to care outside the hospital.
    - Based on this model, the developers considered the following SES and race variables:
        - Dual eligible status
        - African American race
        - AHRQ SES index
- **Empirical analysis of SDS factors:**
    - The developers considered African-American race, dual-eligible status-i.e. enrolled in both Medicare and Medicaid, and AHRQ SES index score. Using the data from the Hospital-Wide All-Cause Readmission Measure for the 2015 reporting year the developers assessed the relationship between the SES variables

and race with the outcome and examined the incremental effect in a multivariable mode.

- The developers also examined the extent to which the addition of any one of these variables improved model performance or changed hospital results.
- The developer stated that they examined all patient-level indicators of both SES and race/ethnicity that are reliably available for all Medicare beneficiaries and linkable to claims data and selected those that are most valid.
- The developer noted that the AHRQ-validated SES index score is a widely-used variable that describes the average socioeconomic status of people living in defined geographic areas. The developer notes that its value as a proxy for patient-level SDS is depend on having the most granular level data.
    - These variables are linked to patients by zip code and census block; however, the data are only linked at a 5-digit zip code level—nine-digit zip code data, which may provide a more granular view of patient sociodemographic status, were not available.
    - However, the developers note they are currently performing analyses at the census block level (the most granular level possible in this dataset) and hope to present the results of this analysis to the committee.
- The developer assessed the relationship between the SDS variables and the 30-day readmission rate and examined the incremental effect of SDS in a multivariable model, evaluating the extent to which the addition of any one of these variables improved model performance or changed hospital results.
- The developer notes that 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. Therefore, the developers performed a decomposition analysis to assess the independent effects of the SES and race variables at the patient level and the hospital level.
- The developers' analysis found that the prevalence of SDS factors in the hospital-wide readmission cohort does vary across measured entities.
- With regard to the empirical association of each SDS variable with the outcome (bivariate), the analysis found that patient-level observed hospital-wide readmission rate for dual eligible patient patients was higher, at 19.28% compared with 14.83% for all other patients. The readmission rate for African-American patients was also higher at 19.16% compared with 15.1% for patients of all other races. Similarly the readmission rate for patients in the lowest SES quartile by AHRQ index was 16.81% compared with 15.05% for all other patients.
- With regard to the strength and significance of the SDS variables in the context of a multivariable model, the developers' analysis found that the effect size of each of these variables is small, the c-statistic (i.e., predictive value) is unchanged with the addition of any of these variables into the model, and the addition of any of these variables into the model has little to no effect on hospital performance.
    - The median absolute change in hospitals' RSRRs when adding a dual eligiblity indicator is 0.004% (interquartile range [IQR] -0.017% – 0.024%, minimum -0.309% – maximum 0.135%) with a correlation coefficient between RSRRs for each hospital with and without Medicaid added of 0.99836.
    - The median absolute change in hospitals' RSRRs when adding a race indicator is 0.011% (IQR -0.010% – 0.033%; minimum -0.671% – maximum 0.130%) with a correlation coefficient between RSRRs for each hospital with and without race added of 0.99814.
    - The median absolute change in hospitals' RSRRs when adding a low SES AHRQ indicator is 0.007% (IQR -0.033% – 0.036%; minimum -0.322% – maximum 0.135%) with a correlation coefficient between RSRRs for each hospital with and without low SES added of 0.99691.
- The developers state that patient-level and hospital-level dual eligible, race, and low AHRQ SES Index

effects were significantly associated with each of the hospital wide readmission models (Medicine, Surgery, Cardiorespiratory, Cardiovascular, and Neurology) in the decomposition analysis. The developers note that if the dual eligible, race, or low AHRQ SES Index variables are 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 developers state that given these findings and complex pathways that could explain any relationship between SDS and readmission, they did not incorporate SDS variables into the measure.
- **Risk Model Diagnostics:**
    - To assess the overall performance of their risk-adjustment model, the developers computed three summary statistics, including:
        - *Area under the receiver operating characteristic (ROC) curve* (also known as a c-statistic, which measures the probability that the model's prediction of the outcome is better than chance)
        - *Predictive ability* (the model's ability to distinguish high-risk subjects from low-risk subjects)
        - *Over-fitting indices (model calibration)* (to ensure that the model is not only describing the relationship between predictive variables and outcome in the development dataset but also providing valid predictions in new patients)
    - For Hybrid HWR Measure Development Sample, the findings from this analysis are as follows:
        - C-statistic:
            - Medicine cohort: 0.651
            - Surgery/Gynecology cohort: 0.802
            - Cardiorespiratory cohort: 0.668
            - Cardiovascular cohort: 0.731
            - Neurology cohort: 0.708
            - The developers state the c-statistics indicate good to excellent model discrimination across the specialty cohort models.
        - Predictive ability (lowest decile %, highest decile %):
            - Medicine cohort: 8%-35%
            - Surgery/Gynecology cohort: 0%-35%
            - Cardiorespiratory cohort: 9%-39%
            - Cardiovascular cohort: 2%-29%
            - Neurology cohort: 4%-33%
            - The developers state higher deciles of the predicted outcomes are associated with higher observed outcomes, which show a good calibration of the model. This plot indicates good discrimination of the model and good predictive ability.
        - Overfitting indices (model calibration) [presented as (γ0, γ1)]:
            - The developer states that if the γ0 in the validation samples are substantially far from zero and the γ1 is substantially far from one, there is potential evidence of over-fitting. The calibration value of close to 0 at one end and close to 1 to the other end indicates good calibration of the model.
                - Medicine cohort: (0.000, 1.000)
                - Surgery/Gynecology cohort: (0.000, 1.000)
                - Cardiorespiratory cohort: (0.000, 1.000)
                - Cardiovascular cohort: (0.000, 1.000)
                - Neurology cohort: (0.000, 1.000)
            - The developer state that the calibration value of close to 0 at one end and close to 1 to the other end indicates good calibration of the model. The risk decile plot shows excellent discrimination of the model and good predictive ability.
        - The developer's overall interpretation of the results of their analysis is that the findings demonstrate the risk-adjustment model adequately controls for differences in patient characteristics (case mix).

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 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 method for discriminating hospital performance has not been determined.
- For public reporting of measures of hospital outcomes developed with similar methodology, 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.

Question for the Committee:

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

2b6. Comparability of data sources/methods:

Not applicable

2b7. Missing Data

- For the EHR data elements used in the measure's risk models, the developers anticipate that there will be some missing data.
- The developers note that testing for the rates of capture is above 90% for the data elements included in the risk models.
- To reduce the chance of bias due to missing data the developers set missing values to the median value in all measure risk models and included a dummy variable whenever a data element was missing in 5% or more of the admissions in each specialty cohort.
- To reduce the effect of the spurious outliers, we transformed extreme values by replacing them with a value at the outer limit of a designated range by a process called Winsorization.
- All continuous variables with values less than 1st percentile or higher than the 99th percentile were Winsorized (i.e., values less than the 1st percentile were assigned to the value of the 1st percentile, and values greater than the 99th percentile were assigned to the value of the 99th percentile). Missing data values were set to the median value for the cohort. In addition, dummy variables for missing data were included in the statistical models.

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)

2. Scientific Acceptability of Measure Properties

2a1. & 2b1. Specifications

Comments: **none indicated

**Rationale for use as quality measure (i.e. mutable by QI efforts) provided (same as measure #1789)

**No logic or calculation steps were not clear or poorly documented. I would have the same concerns about implementation that I had with 1789-- that is that hospitals without sophisticated analytics capabilities cannot calculate results independently of CMS,

which could impact improvement efforts. There were no inconsistencies b/t evidence and specification.

2a2. Reliability Testing

Comments: **The developer "considered all measure testing as preliminary due to the small sample of hospitals in the KPNC database, and the lack of patient socio demographic diversity within the integrated network of KPNC hospitals. Confirming the validity and reliability of the measure requires data from a larger, more diverse set of hospitals and EMR systems"
diverse set of hospitals and more than one EHR system." However the conclusions from the non hybrid and hybrid seemed consistent.

**High capture rates for EMR based clinical data elements for risk models noted (see T1 p. 46-48).

Agreement between two RSRRs for each of two patient samples for hospitals using new risk models was ICCagreement = 0.688.

Cronbach's alpha for elements of risk models (?) for standardized risk ratios was high for original and hybrid models (a=0.837,0.833).

**Yes, the measure was reliably tested with adequate scope.

The developer summarized the rate of capture for all clinical data elements within the measure documentation.

For performance scores, two RSRR's were calculated for each hospital. The intra-class correlation coefficient was .688, which is considered to be a moderate level of agreement.

2b2. Validity Testing

Comments: **The developer "considered all measure testing as preliminary due to the small sample of hospitals in the KPNC database, and the lack of patient socio demographic diversity within the integrated network of KPNC hospitals. Confirming the validity and reliability of the measure requires data from a larger, more diverse set of hospitals and EMR systems"
diverse set of hospitals and more than one EHR system." However the conclusions from the non hybrid and hybrid seemed consistent.

**Developer compared electronically versus manually abstracted data elements on 368 charts. Agreement for these cases was greater than 90% for all but 2 of 17 variables (weight, bicarbonate mEq/l). A second set of 391 charts showed lower levels of agreement greater than 90% for only 6 of 17 variables (see Table 5). The relatively small number of charts in each sample and the variation in results suggest further validity testing is needed.

Correlation between RSRRs from claims only versus hybrid risk model adjusted rates were high overall (r=0.99) and high for the 5 specialty cohorts (Table 3.10).

C-statistics were comparable for the hybrid model in developmental versus validation sample across all 5 specialty cohorts (Table 3.7). However, data in Table 3.5 (p. 27 of manuscript provided), do not suggest meaningful differences between "HWR" and "HWR+CCDE" risk models.

**Consistent with the already approved version of the measure, the score from this measure does demonstrate that it is an indicator of quality.

One consideration is that the validity of the abstracted/extracted measures we tested in 3 EMRs at 21 hospitals. As this measure is rolled out, adoption may be impacted by the (well documented) challenges that CMS has encountered with other electronic clinical quality measures.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **The variability or inconsistency of data across systems and EHRs may constitute a gap. I would like to hear from the developer if they will continue to use Winsorized approach, but am curious about gender fields and am curious if this presents a problem.

**2b3. Exclusions affect only a small proportion of patients.

2b4. The same approach to and empirical results by testing SDS variables as for measure #1789 were presented.

2b5. See measure #1789

2b6. N/A

2b7. Developers appeared to impute missing data using median values for the cohort compared. Outliers were winsorized using

outer limits of range.
**Exclusions and risk adjustment methodology seemed appropriate. I saw no threats to validity based on missing data.

Criterion 3. Feasibility

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 based on administrative claims data (e.g., DRG, ICD-9/10), which the developers note are routinely generated and collected as part of hospitals' billing processes and electronic clinical data, which will be collected from hospitals using MAT output and value sets to inform data queries and electronic reporting requirements.
- The developer indicates that all data elements are in defined fields in a combination of electronic sources.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*
- *If an eMeasure, does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?*

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

Committee pre-evaluation comments
Criteria 3: Feasibility

3. Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **Because the study used only kaiser data sources, I would like to hear from the developer any feasibility testing for other non EPIC, non closed systems like Kaiser. #2 "standard definition" across settings may not be equivalent outside of KP. Testing was done against 4 EMRS, using a limited data set. No comments were made about practice or implementation differences.
**Claims and EMR derived data appear to be routinely collected. Data collection strategy appears to be appropriate. Further evidence regarding generalizability of strategy across EMR vendors would be helpful.
**The feasibility of data element collection was explored at length within the measure documentation. The developer took care to limit the extracted data elements to only those that were routinely documented in a structured format for a high percentage of encounter-types by convening a technical expert panel (TEP) and only selecting elements that achieved consensus agreement on feasibility by the TEP.

Criterion 4: Usability and Use

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**
OR
Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

- The developer notes that, "CMS intends to implement this measure in the Hospital Inpatient Quality Reporting (HIQR) Program once the clinical data elements required for this measure have been reported by hospitals for one year. This measure requires one year of data for calculation. The exact timeline therefore depends on the implementation of a reporting mechanism for these data elements. Once this new measure is implemented, it may replace the claims-only other Hospital-Wide All Cause Unplanned Readmission Measure."

Improvement results

- Since this measure is not in use, there are no performance results to assess improvement at this time.
- The developer states that they expect that "there will be improvement in measure scores over time since publicly reported measure scores can reduce adverse patient outcomes associated with days spent in acute care for heart failure by capturing and making acute care utilization following the index hospitalization more visible to providers and patients."

Potential harms

- The developer noted that there were no unintended consequences during development, testing or re-specification. They are committed to ongoing monitoring of potential unintended consequences, such as the inappropriate shifting of care, increased patient morbidity and mortality, and other negative intended consequences over time.

Feedback :

- During the 2014-2015 MAP review, MAP encouraged further development of this e-Measure version of the endorsed Hospital-Wide All-Cause Readmissions Measure (HWR) for inclusion in the IQR program. MAP expressed caution that this measure should contain proper risk adjustment and was supportive of using clinical data to improve the risk adjustment model performance. Further, the MAP noted that CMS should review for the empirical and conceptual relationship between SDS factors and hospital-wide readmissions, and seek endorsement on this version of the measure by the relevant NQF standing committee. MAP noted that after review and endorsement by the NQF Standing Committee, this measure should be brought back to the MAP for further discussion.

***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

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **The benefit of the measure is: despite the potential for variability in CCDE across health systems and EMR systems, the inclusion of EMR data in this measure presents a very low risk in potential negative care outcomes, or care decisions and introduces EMR data into measurement. This is a first of many. I look forward to hearing from the measure developer, but for the benefit outlined above, would recommend approval of this measure

**Since the measure as proposed is not currently in use, usability is difficult to evaluate. Developers cited no unintended consequences in the developmental phase of the measure and intend to continue to monitor for same.
**No unintended consequences outweigh the benefit of implementing this measure. As previously mentioned, this measure would be subject to the larger, macro challenges that regulators have had in implementing electronic measures at scale.

Criterion 5: Related and Competing Measures

Related or competing measures

- Related to #1768: Plan All-Cause Readmissions (PCR) and #1789: Hospital-Wide All-Cause Unplanned Readmission Measure (HWR).

Harmonization

- Since there are no competing measures, harmonization is not required. The developer notes that once the clinical data elements required for this measure calculation are completed, #2879, the hybrid measure may replace #1789: Hospital-Wide All-Cause Unplanned Readmission Measure (HWR).

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Title: Hybrid Hospital-Wide Readmission Measure with Claims and Electronic Health Record Data

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: 1/29/2016

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

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

Subcriterion 1a. Evidence to Support the Measure Focus

The measure focus is a health outcome or is evidence-based, demonstrated as follows:

- Health outcome:[3] a rationale supports the relationship of the health outcome to processes or structures of care.
- Intermediate clinical outcome, Process,[4] or Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence[5] that the measure focus leads to a desired health outcome.
- Patient experience with care: evidence that the measured aspects of care are those valued by patients and for which the patient is the best and/or only source of information OR that patient experience with care is correlated with desired outcomes.
- Efficiency:[6] evidence for the quality component as noted above.

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. 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.
5. 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.
6. Measures of efficiency combine the concepts of resource use and quality (NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1. This is a measure of:

Outcome

☒ Health outcome: 30-day all-cause readmission

Health outcome includes patient-reported outcomes (PRO, i.e., HRQoL/functional status, symptom/burden, experience with care, health-related behaviors)

☐ Intermediate clinical outcome: Click here to name the intermediate outcome

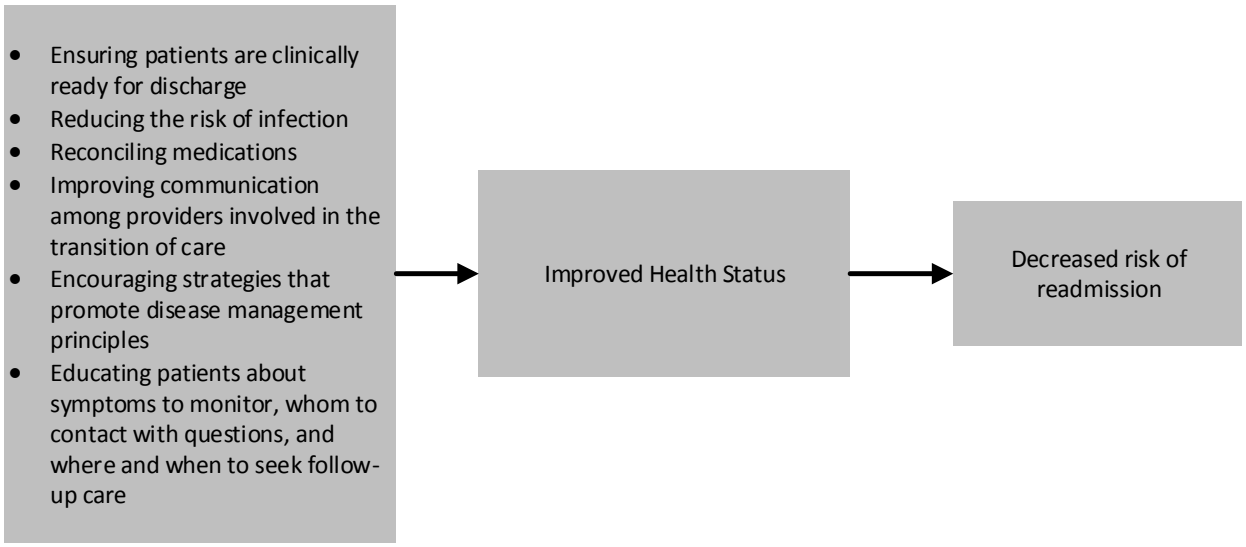
☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME PERFORMANCE MEASURE *If not a health outcome, skip to 1a.3*

1a.2. Briefly state or diagram the linkage between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.



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 of unplanned, all-cause readmission after admission for any eligible condition within 30 days of hospital discharge. 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.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) and at least one healthcare structure, process, intervention, or service.

Note: For health outcome performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

The diagram above indicates some of the many care processes that can influence readmission risk. In general, randomized controlled trials have shown that improvement in the following areas can directly reduce readmission rates: quality of care during the initial admission; improvement in communication with patients, their caregivers, and their clinicians; patient education; pre-discharge assessment; and coordination of care after discharge. Evidence that hospitals have been able to reduce readmission rates through these quality-of-care initiatives illustrates the degree to which hospital practices can affect readmission rates. Successful randomized trials have reduced 30-day readmission rates by 20-40% [1-11]. Since 2008, 14 Medicare Quality Improvement Organizations have been funded to focus on care transitions, applying lessons learned from clinical trials. Several have been notably successful in reducing readmissions. The strongest evidence supporting the efficacy of improved discharge processes and enhanced care at transitions is a randomized controlled trial by the Project RED (Re-Engineered Discharge) intervention, in which a nurse was assigned to each patient as a discharge advocate, responsible for patient education, follow-up, medication reconciliation, and preparing individualized discharge instructions sent to the patient's primary care provider and there was a follow-up phone call from a pharmacist within four days of discharge, which demonstrated a 30% reduction in 30-day readmissions [1]. Hospital processes that reflect the quality of inpatient and outpatient care such as discharge planning, medication reconciliation, and coordination of outpatient care have been shown to reduce readmission rates [12]. Although readmission rates are also influenced by hospital system characteristics, such as the bed capacity of the local health care system, these hospital characteristics should not influence quality of care [13]. Therefore, this measure does not risk adjust for such hospital characteristics.

Studies have estimated the rate of preventable readmissions to be as low as 12% and as high as 76% [14, 15]. Given that studies have shown readmissions to be related to quality of care, and that interventions have been able to reduce 30-day readmission rates, it is reasonable to consider an all-condition readmission rate as a quality measure.

The hospital-wide risk-standardized readmission rate (RSRR) measure is thus intended to inform quality-of-care improvement efforts, as individual process-based performance measures cannot encompass all the complex and critical aspects of care within a hospital that contribute to patient outcomes. As a result, many stakeholders, including patient organizations, are interested in outcome measures that allow patients and providers to assess relative outcomes performance for hospitals.

References:

1. Jack BW, Chetty VK, Anthony D, Greenwald JL, Sanchez GM, Johnson AE, et al. A reengineered hospital discharge program to decrease rehospitalization: a randomized trial. Ann Intern Med 2009;150(3):178-87.

2. Coleman EA, Smith JD, Frank JC, Min SJ, Parry C, Kramer AM. Preparing patients and caregivers to participate in care delivered across settings: the Care Transitions Intervention. J Am Geriatr Soc 2004;52(11):1817-25.

3. 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 2009;57(3):395-402.

4. 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.

5. Koehler BE, Richter KM, Youngblood L, Cohen BA, Prengler ID, Cheng D, 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. J Hosp Med 2009;4(4):211-218.

6. Mistiaen P, Francke AL, Poot E. Interventions aimed at reducing problems in adult patients discharged from hospital to home: a systematic metareview. BMC Health Serv Res 2007;7:47.

7. 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 1994;120(12):999-1006.

8. Naylor MD, Brooten D, Campbell R, Jacobsen BS, Mezey MD, Pauly MV, et al. Comprehensive discharge planning and home follow-up of hospitalized elders: a randomized clinical trial. Jama 1999;281(7):613-20.

9. 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 2002;17(3):186-92.

10. Weiss M, Yakusheva O, Bobay K. Nurse and patient perceptions of discharge readiness in relation to postdischarge utilization. Med Care 2010;48(5):482-6.

11. 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.

12. Nelson EA, Maruish ME, Axler JL. Effects of Discharge Planning and Compliance With Outpatient Appointments on Readmission Rates. Psychiatr Serv. July 1 2000;51(7):885-889.

13. Fisher ES, Wennberg JE, Stukel TA, Sharp SM. Hospital Readmission Rates for Cohorts of Medicare Beneficiaries in Boston and New Haven. New England Journal of Medicine. 1994;331(15):989-995.

14. Benbassat J, Taragin M. Hospital readmissions as a measure of quality of health care: advantages and limitations. Archives of Internal Medicine 2000;160(8):1074-81.

15. Medicare Payment Advisory Commission (U.S.). Report to the Congress promoting greater efficiency in Medicare. Washington, DC: Medicare Payment Advisory Commission, 2007.

Note: For health outcome performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the linkages between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

N/A. This measure is not an intermediate outcome, process, or structure performance measure.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

N/A This measure does not correspond with a United States Preventive Services Task Force Recommendation

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

N/A

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

N/A

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range:

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of 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
Hybrid_HWR_NQF_Evidence_Attachment_01-29-16_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 all cause unplanned readmission rates among Medicare beneficiaries 65 years and older admitted to all non-federal US acute care hospitals. 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 their patient case mix and hospital service mix, and therefore promote hospital quality improvement and better inform consumers about care quality.

Hospital-wide readmission is a priority area for outcomes 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.

This Hybrid HWR measure incorporates both data from claims as well as clinical data elements pulled from the EHR in risk adjustment of the readmission models. Some benefits of including the clinical data elements are:

1. Inclusion of patient-level clinical data related to severity of illness is responsive to providers who continue to express preference for using patient-level clinical data, and provides an opportunity to incorporate clinical data into outcome measures.

2. Hospitals will increasingly use EHR data to assess severity of illness and patients' risk of poor outcomes. This provides an opportunity to align the measure with clinical decision support systems that many providers utilize to alert care teams about patients at increased risk of poor outcomes in real time during the inpatient stay.

3. Collecting a simple core set of clinical data elements that perform well as risk-adjustment variables (for illness severity) across conditions can greatly reduce the cost and effort of future measure development, improve harmonization, and create opportunity for longitudinal assessment of patient status and quality of care across settings.

4. These core clinical data elements will provide measure developers with a standard set of reliable data that can be used as a starting place when building risk-adjustment models for quality measures using clinical data.

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.*
For model development purposes only, we used Dataset 1, which contained merged inpatient claims with clinical data elements derived from patients' EHRs. Our cohort included 381,980 admissions at 21 hospitals.

Overall Measure score – Development Sample, Dataset 1
Mean RSRR (%) 14.84
Min RSRR (%) 13.15
Median RSRR (%) 15.04
Max RSRR (%) 16.16
Results above reflect performance of a small number of hospitals (21) from a single health system, Dataset 1.

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.**
CMS currently publicly reports a claims-based Hospital-Wide All-Cause Unplanned Readmission Measure (NQF #1789). The results for this measure, as reported in the 2015 update to the Hospital Compare website, are based on RSRRs calculated for admissions among Medicare FFS patients aged 65 and older from July 1, 2012 - June 30, 2013. It includes 4,772 hospitals. The median hospital RSRR was 15.5%, with an interquartile range of 11.0% to 21.4%.

Randomized controlled trials have shown that improvement in the following areas can directly reduce readmission rates; quality of care during the initial admission; improvement in communication with patients, their caregivers, and their clinicians; patient education; pre-discharge assessment; and coordination of care after discharge1,2,3,4,5,6,7,8,9,10,11,12,13,14. Successful randomized trials have reduced 30-day readmission rates by 20-40%. Widespread application of these clinical trial interventions to general practice has also been encouraging. Since 2008, CMS has funded 14 Medicare Quality Improvement Organizations (QIOs) to focus on care transitions and to apply lessons learned from clinical trials. Several QIOs have been notably successful in reducing readmissions within 30 days15. Evidence that hospitals have been able to reduce readmission rates through these quality-of-care initiatives illustrates the degree to which hospital practices can affect readmission rates.

Resources:

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. Krumholz HM, Amatruda J, Smith GL, et al. Randomized trial of an education and support intervention to prevent readmission of patients with heart failure. Journal of the American College of Cardiology. Jan 2 2002;39(1):83-89.

3. Van Walraven C, Seth R, Austin PC, Laupacis A. Effect of discharge summary availability during post-discharge visits on hospital readmission. Journal of General Internal Medicine. Mar 2002;17(3):186-192.

4. Conley RR, Kelly DL, Love RC, McMahon RP. Rehospitalization risk with second-generation and depot antipsychotics. Annals of Clinical Psychiatry. Mar 2003;15(1):23-31.

5. 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. Journal of the American Geriatrics Society. Nov 2004;52(11):1817-1825.

6. 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.

7. 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.

8. 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.

9. 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 Services Research. 2007;7:47.

10. 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. Journal of the American Geriatrics Society. Mar 2009;57(3):395-402.

11. 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.

12. 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.

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

14. 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.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.

15. (CFMC) CFfMC. Care Transitions QIOSC. 2010; http://www.cfmc.org/caretransitions/Hospital-wide Readmission Measure 68 July 2012, 2011.

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.*
Due to the small number of hospitals (21) and relative lack of diversity with respect to socioeconomic status (SES) in the Kaiser Permanente of Northern California system, we did not perform disparities analyses for this measure as specified. However, we have conducted disparities analysis for the claims-only HWR measure (NQF #1789) which uses the same exact specifications except for the additional clinical data elements in the measure's risk models. We present that data in 1b.5 below.

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.
The most informative data on potential disparities for hospital-wide readmission come from analysis of 30-day readmission rates for the HWR measure (NQF #1789) using 2013-2014 Medicare data.

Distribution of HWR RSRRs by Proportion of Dual Eligible Patients:
Dates of Data: July 2013 through June 2014
Data Source: Medicare FFS claims

Characteristic//Hospitals with a low proportion (=9.8%) Dual Eligible patients//Hospitals with a high proportion (=22.6%) Dual Eligible patients
Number of Measured Hospitals// 1,257 // 1,219
Number of Patients// 2,137,895 patients in low-proportion hospitals // 927,007 in high-proportion hospitals
Maximum// 18.7 // 20.1
90th percentile// 16.2 // 16.8
75th percentile// 15.7 // 16.0
Median (50th percentile)// 15.3 // 15.6
25th percentile// 14.8 // 15.2
10th percentile// 14.3 // 14.9
Minimum // 11.5 // 12.2

Distribution of HWR RSRRs by Proportion of African-American Patients:
Dates of Data: July 2013 through June 2014
Data Source: Medicare FFS claims

Characteristic// Hospitals with a low proportion (=2.2%) African-American patients//Hospitals with a high proportion (=9.4%) African-American patients
Number of Measured Hospitals// 1,156 // 1,180
Number of Patients// 222,648 patients in low-proportion hospitals/ 2,294,715 in high-proportion hospitals
Maximum// 19.1 // 19.9

90th percentile// 16.0 // 17.1
75th percentile// 15.6 // 16.3
Median (50th percentile)// 15.4 // 15.7
25th percentile// 15.1 // 15.2
10th percentile// 14.8 // 14.8
Minimum // 12.9 // 12.2

Distribution of HWR RSRRs by Proportion of Patients with AHRQ SES Index Scores Below 45.0:
Dates of Data: July 2013 through June 2014
Data Source: Medicare FFS claims and the American Community Survey (2008-2012) data

Characteristic//Hospitals with a low proportion of patients below AHRQ SES index score of 45.0 (=5.0%)// Hospitals with a high proportion of patients below AHRQ SES index score of 45.0 (=57.1%)
Number of Measures Hospitals// 1,209 // 1,217
Number of Patients// 1,651,852 patients in hospitals with low proportion of patients below AHRQ SES index score of 45.0 //795,899 patients in hospitals with high proportion of patients below AHRQ SES index score of 45.0
Maximum// 19.9 // 20.1
90th percentile// 16.2 // 16.6
75th percentile// 15.7 // 16.0
Median (50th percentile)// 15.3 // 15.5
25th percentile// 14.9 // 15.2
10th percentile// 14.5 // 14.8
Minimum // 11.5 // 13.0

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.**
Readmission following hospitalization is a costly and often preventable event. Between July 2011 and June 2012, almost one-sixth of Medicare beneficiaries – more than 1.1 million patients – were readmitted within 30 days of discharge from an acute care hospital1. Medicare reported that readmissions cost Medicare more than $17 billion annually2. In a 2013 report to the Congress, the Medicare Payment Advisory Commission (MedPAC) estimated that in 2011, more than 76% of Medicare admissions were followed by potentially preventable readmissions3. They report that these potentially preventable readmissions cost Medicare roughly $10 billion per year.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
1. 2013 measure updates and specifications report: Hospital-wide all-cause unplanned readmission measure (Version 2.0). Centers for Medicare & Medicaid Services, March 2013.
2. National Medicare readmission findings: Recent data and trends. Centers for Medicare & Medicaid Services presentation to AcademyHealth, Slide 3. June 24, 2012.
3. Medicare Payment Advisory Commission (U.S.). Report to the Congress: Medicare and the health care delivery system. Washington, DC: Medicare Payment Advisory Commission, 2013.

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

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 : Acute Myocardial Infarction, Cardiovascular : Atrial Fibrillation, Cardiovascular : Congestive Heart Failure, Cardiovascular : Hyperlipidemia, Cardiovascular : Hypertension, Cardiovascular : Ischemic Heart Disease, Coronary Artery Disease, Cardiovascular : Percutaneous Coronary Intervention (PCI), Cardiovascular : Screening, Endocrine, Endocrine : Diabetes, Endocrine : Screening, Endocrine : Thyroid Disorders, Gastrointestinal (GI), Gastrointestinal (GI) : Appendicitis, Gastrointestinal (GI) : Cirrhosis, Gastrointestinal (GI) : Gall Bladder Disease, Gastrointestinal (GI) : Gastroenteritis, Gastrointestinal (GI) : Gastro-Esophageal Reflux Disease (GERD), Gastrointestinal (GI) : GI Bleeding, Gastrointestinal (GI) : Peptic Ulcer, Gastrointestinal (GI) : Polyps, Gastrointestinal (GI) : Screening, GU/GYN, GU/GYN : Incontinence, GU/GYN : Screening, Infectious Diseases, Infectious Diseases : Hepatitis, Infectious Diseases : Human Immunodeficiency Virus/Acquired Immune Deficiency Syndrome (HIV/AIDS), Infectious Diseases : Respiratory, Infectious Diseases : Sexually Transmitted, Infectious Diseases : Tuberculosis, Musculoskeletal : Hip/Pelvic Fracture, Musculoskeletal : Low Back Pain, Musculoskeletal : Osteoporosis, Neurology, Neurology : Brain Injury, Neurology : Cognitive Impairment/Dementia, Neurology : Delirium, Neurology : Stroke/Transient Ischemic Attack (TIA), Pulmonary/Critical Care, Pulmonary/Critical Care : Asthma, Pulmonary/Critical Care : Chronic Obstructive Pulmonary Disease (COPD), Pulmonary/Critical Care : Critical Care, Pulmonary/Critical Care : Dyspnea, Pulmonary/Critical Care : Pneumonia, Pulmonary/Critical Care : Sleep Apnea, Renal, Renal : Chronic Kidney Disease (CKD), Renal : End Stage Renal Disease (ESRD), Surgery, Surgery : Cardiac Surgery, Surgery : General Surgery, Surgery : Perioperative, Surgery : Thoracic Surgery, Surgery : Vascular Surgery

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Care Coordination : Readmissions, Safety, Safety : Complications, Safety : Healthcare Associated Infections, Safety : Medication Safety, Safety : Readmissions

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 an eMeasure **Attachment:** CCDE_v4_Artifacts_01-29-16_v1.0.zip

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:** Hybrid_HWR_NQF_Data_Dictionary_01-29-16_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 readmission. We define readmission as an inpatient admission for any cause, with the exception of certain planned readmissions, within 30 days from the date of discharge from an eligible index admission. If a patient has more than one unplanned admission (for any reason) within 30 days after discharge from the index admission, only one is

counted as a readmission. The measure looks for a dichotomous yes or no outcome of whether each admitted patient has an unplanned readmission within 30 days. However, if the first readmission after discharge is considered planned, any subsequent unplanned readmission is not counted as an outcome for that index admission because the unplanned readmission could be related to care provided during the intervening planned readmission rather than during the index admission.

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 admission.

Denominator Time Window: This measure was developed with 12 months of data and is currently publicly reported with one year of data.

Numerator time window: Unplanned readmission from any cause within 30 days from the discharge date for the index admission.

The time period for public reporting has not been determined, however, the publicly reported HWR measure (NQF #1789) that this measure was based on uses one year of data.

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.
The measure counts readmissions to any acute care hospital for any cause within 30 days of the date of discharge of the index admission, excluding planned readmissions as defined below.

Planned Readmission Algorithm (Version 3.0)
The Planned Readmission Algorithm is a set of criteria for classifying readmissions as planned among the general Medicare population using Medicare administrative claims data. The algorithm identifies admissions that are typically planned and may occur within 30 days of discharge from the hospital.

The Planned Readmission Algorithm has three fundamental principles:

1. A few specific, limited types of care are always considered planned (obstetric delivery, transplant surgery, maintenance chemotherapy/immunotherapy, rehabilitation);
2. Otherwise, a planned readmission is defined as a non-acute readmission for a scheduled procedure; and,
3. Admissions for acute illness or for complications of care are never planned.

The algorithm was developed in 2011 as part of the Hospital-Wide Readmission measure. In 2013, CMS applied the algorithm to its other readmission measures.

The Planned Readmission Algorithm and associated code tables are attached in data field S.2b (Data Dictionary or Code Table).

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The measure includes admissions for Medicare beneficiaries who are 65 years and older and are discharged from all non-federal, acute care inpatient US hospitals (including territories)with a complete claims history for the 12 months prior to admission.

Additional details are provided in S.9 Denominator Details.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

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)*
To be included in the measure cohort, patients must be:

• Enrolled in Medicare Fee-for-Service (FFS) Part A for the 12 months prior to the date of admission and during the index admission;
• Aged 65 or over;
• Discharged alive from a non-federal short-term acute care hospital; and,
• Not transferred to another acute care facility.

The measure aggregates the ICD-9 principal diagnosis and all procedure codes of the index admission into clinically coherent groups of conditions and procedures (condition categories or procedure categories) using the Agency for Healthcare Research and Quality (AHRQ) Clinical Classifications System (CCS). There are a total of 285 mutually exclusive AHRQ condition categories, most of which are single, homogenous diseases such as pneumonia or acute myocardial infarction. Some are aggregates of conditions, such as "other bacterial infections." There are a total of 231 mutually exclusive procedure categories. Using the AHRQ CCS procedure and condition categories, the measure assigns each index hospitalization to one of five mutually exclusive specialty cohorts: surgery/gynecology, cardiorespiratory, cardiovascular, neurology, and medicine. The rationale behind this organization is that conditions typically cared for by the same team of clinicians are expected to experience similar added (or reduced) levels of readmission risk.

The measure first assigns admissions with qualifying AHRQ procedure categories to the Surgery/Gynecology Cohort. This cohort includes admissions likely cared for by surgical or gynecological teams.

The measure then sorts admissions into one of the four remaining specialty cohorts based on the AHRQ diagnosis category of the principal discharge diagnosis:

The Cardiorespiratory Cohort includes several condition categories with very high readmission rates such as pneumonia, chronic obstructive pulmonary disease, and heart failure. These admissions are combined into a single cohort because they are often clinically indistinguishable and patients are often simultaneously treated for several of these diagnoses.

The Cardiovascular Cohort includes condition categories such as acute myocardial infarction that in large hospitals might be cared for by a separate cardiac or cardiovascular team.

The Neurology Cohort includes neurologic condition categories such as stroke that in large hospitals might be cared for by a separate neurology team.

The Medicine Cohort includes all non-surgical patients who were not assigned to any of the other cohorts.

The full list of the specific diagnosis and procedure AHRQ CCS categories used to define the specialty cohorts are attached in Excel Data Dictionary data field S.2b.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
The measure excludes index admissions for patients:

1. Admitted to Prospective Payment System (PPS)-exempt cancer hospitals;
2. Without at least 30 days post-discharge enrollment in FFS Medicare;
3. Discharged against medical advice (AMA);
4. Admitted for primary psychiatric diagnoses;
5. Admitted for rehabilitation; or
6. Admitted for medical treatment of cancer.

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)*
1. Admitted to a PPS-exempt cancer hospital, identified by the Medicare provider ID.

2. Admissions without at least 30 days post-discharge enrollment in FFS Medicare are determined using data captured in the Medicare Enrollment Database (EDB).

3. Discharges against medical advice (AMA) are identified using the discharge disposition indicator in claims data.

4. Admitted for primary psychiatric disease, identified by a principal diagnosis in one of the specific AHRQ CCS categories listed in the attached in Excel Data Dictionary data field S.2b.

5. Admitted for rehabilitation care, identified by the specific ICD-9 diagnosis codes included in CCS 254 (Rehabilitation care; fitting of proestheses; and adjustment of devices).

6. Admitted for medical treatment of cancer, identified by the specific AHRQ CCS categories listed in the attached data dictionary.

The full list of the specific diagnosis and procedure CCS categories excluded from the specialty cohorts are attached in Excel Data Dictionary data field S.2b.

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 employs a hierarchical logistic regression model to create a hospital-level 30-day RSRR. In brief, the approach simultaneously models data at the patient and hospital levels to account for the variance in patient outcomes within and between hospitals (Normand & Shahian, 2007). At the patient level, the model adjusts the log-odds of readmission within 30 days of discharge for age and selected clinical covariates. At the hospital level, the approach models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of readmission at the hospital, after accounting for patient risk. If there were no differences among hospitals, then after adjusting for patient risk, the hospital intercepts should be identical across all hospitals.

Candidate and Final Risk-adjustment Variables:
This measure uses risk variables from both claims data and from electronic health records (EHR). Candidate variables were patient-level risk-adjusters that were expected to be predictive of readmission, based on empirical analysis, prior literature, and clinical judgment, including age, indicators of comorbidity, and disease severity. For risk variables derived from claims data, only those variables in the current publicly reported claims-based Hospital-Wide All Cause Unplanned Readmission Measure were considered as candidate variables. For each patient, risk variables were obtained from claims extending 12 months prior to and including the index admission and, for the clinical data elements from the electronic health record (EHR), only those captured during the index admission. These risk-adjusters are identified using inpatient Medicare FFS claims data.

We use a fixed, common set of claims-based variables in all our models for simplicity and ease of data collection and analysis. However, we estimate a hierarchical logistic regression model for each specialty cohort separately, and the coefficients associated with each variable may vary across specialty cohorts. The model adjusts for casemix differences based on the clinical status of patients at the time of admission. For the claims data, we use condition categories (CCs), which are clinically meaningful groupings of more than 15,000 ICD-9-CM diagnosis codes (Pope et al., 2000). A file that contains a list of the ICD-9-CM codes and their groupings into CCs is attached in the Excel Data Dictionary data field S.2b. In addition, only comorbidities that convey information about the patient at admission or in the 12 months prior, and not complications that arise during the course of the index hospitalization, are included in the risk adjustment. Hence, we do not risk adjust for CCs that may represent adverse events of care when they are only recorded in the index admission. The models also include a condition-specific indicator for all AHRQ CCS categories with sufficient volume (defined as those with more than 1,000 admissions nationally each year for Medicare FFS data) as well as a single indicator for conditions with insufficient volume in each model.

In addition to the claims-derived candidate variables, we include clinical data elements derived from patients' electronic medical records as candidate variables. Unlike the uniform set of claims-variables used in the risk models, each of the five risk models includes a different set of clinical data elements because some variables were predictive of the readmission outcome some but not all of the specialty cohorts. The clinical data elements include the first vital signs captured within two hours of the start of the encounter and the results of several laboratory tests captured within 24 hours of the start of the encounter (complete blood count and basic chemistry profile). The final set of risk adjustment variables for each cohort are listed in the Excel Data Dictionary data field S.2b and attached Hybrid Hospital-Wide Readmission Measure with Claims and Electronic Health Record Data Technical Report. Some clinical data elements were also transformed into squared data values due to the non-linear relationship between the raw values and the readmission outcome.

Demographics (Common to all risk models)
Age-65 (years, continuous) for patients aged 65 and over cohorts

Clinical Variables (Listed by risk model):

Surgery Cohort:
Systolic Blood Pressure
Heart Rate
Respiratory Rate
Temperature
Weight

Cardiorespiratory Cohort:
Bicarbonate
Creatinine
Glucose
Hematocrit
Sodium
Systolic Blood Pressure
Heart Rate
Oxygen Saturation
WBC Count
Temperature

Cardiovascular Cohort:
Bicarbonate
Creatinine
Hematocrit
Potassium
Sodium
WBC Count
Systolic Blood Pressure
Heart Rate
Oxygen Saturation

Neurology Cohort:
Creatinine
Hematocrit
Sodium
WBC Count
Systolic Blood Pressure
Heart Rate
Oxygen Saturation
Respiratory Rate

Medicine Cohort:
Bicarbonate

Creatinine
Glucose
Hematocrit
Potassium
Sodium
WBC Count
Systolic Blood Pressure
Heart Rate
Respiratory Rate
Temperature

Comorbidities (Common to each of the five risk models)
Metastatic cancer or acute leukemia (CC 7)
Severe cancer (CC 8-9)
Other cancers (CC 10-12)
Severe hematological disorders (CC 44)
Coagulation defects and other specified hematological disorders (CC 46)
Iron deficiency or other unspecified anemias and blood disease (CC 47)
End-stage liver disease (CC 25-26)
Pancreatic disease (CC 32)
Dialysis status (CC 130)
Renal failure (CC 131)
Transplants (CC 128, 174)
Severe infection (CC 1, 3-5)
Other infectious diseases and pneumonias (CC 6, 111-113)
Septicemia/shock (CC 2)
Congestive heart failure (CC 80)
Coronary atherosclerosis or angina, cerebrovascular disease (CC 81-84, 89, 98-99, 103-106)
Specified arrhythmias and other heart rhythm disorders (CC 92-93)
Cardio-respiratory failure or shock (CC 79)
Chronic obstructive pulmonary disease (COPD) (CC 108)
Fibrosis of lung or other chronic lung disorders (CC 109)
Protein-calorie malnutrition (CC 21)
Disorders of fluid/electrolyte/acid-base (CC 22-23)
Rheumatoid arthritis and inflammatory connective tissue disease (CC 38)
Diabetes mellitus (DM) or DM complications (CC 15-20, 119-120)
Decubitus ulcer or chronic skin ulcer (CC 148-149)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)
Seizure disorders and convulsions (CC 74)
Respirator dependence/tracheostomy status (CC 77)
Drug/alcohol psychosis or dependence (CC 51-52)
Psychiatric comorbidity (CC 54-56, 58, 60)
Hip fracture/dislocation (CC 158)

Principal Diagnoses
Refer to the attached Hybrid Hospital-Wide Readmission Measure with Claims and Electronic Health Record Data Technical Report for the full lists of principal diagnosis AHRQ CCS categories included in each specialty cohort risk adjustment model.

References:
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.

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

Pope GC, et al. 2000. Principal Inpatient Diagnostic Cost Group Models for Medicare Risk Adjustment. Health Care Financing Review

21(3): 93-118.

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*)

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.)*
The measure estimates hospital-level 30-day all-cause RSRRs using hierarchical logistic regression models. In brief, the approach simultaneously models data at the patient and hospital levels to account for variance in patient outcomes within and between hospitals. At the patient level, it models the log-odds of hospital readmission within 30 days of discharge using age, selected clinical covariates, and a hospital-specific effect. At the hospital level, the approach models the hospital-specific effects as arising from a normal distribution. The hospital effect represents the underlying risk of a readmission at the hospital, after accounting for patient risk. The hospital-specific effects are given a distribution 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 effects should be identical across all hospitals.

Admissions are assigned to one of five mutually exclusive specialty cohort groups consisting of related conditions or procedures. For each specialty cohort group, the standardized readmission ratio (SRR) is calculated as the ratio of the number of "predicted" readmissions to the number of "expected" readmissions at a given hospital. For each hospital, the numerator of the ratio is the number of readmissions within 30 days predicted based on the hospital's performance with its observed case mix and service mix, and the denominator is the number of readmissions expected based on the nation's performance with that hospital's case mix and service mix. This approach is analogous to a ratio of "observed" to "expected" used in other types of statistical analyses. It conceptually allows a particular hospital's performance, given its case mix and service mix, to be compared to an average hospital's performance with the same case mix and service mix. Thus, a lower ratio indicates lower-than-expected readmission rates or better quality, while a higher ratio indicates higher-than-expected readmission rates or worse quality.

For each specialty cohort, the "predicted" number of readmissions (the numerator) is calculated by using the coefficients estimated by regressing the risk factors (found in Table D.9) and the hospital-specific effect on the risk of readmission. The estimated hospital-specific effect for each cohort is added to the sum of the estimated regression coefficients multiplied by patient characteristics. The results are log transformed and summed over all patients attributed to a hospital to get a predicted value. The "expected" number of readmissions (the denominator) is obtained in the same manner, but a common effect using all hospitals in our sample is added in place of the hospital-specific effect. The results are log transformed and summed over all patients in the hospital to get an expected value. To assess hospital performance for each reporting period, we re-estimate the model coefficients using the data in that period.

The specialty cohort SRRs are then pooled for each hospital using a volume-weighted geometric mean to create a hospital-wide composite SRR. The composite SRR is multiplied by the national observed readmission rate to produce the RSRR. The statistical modeling approach is described fully in the attached Hybrid Hospital-Wide Readmission Measure with Claims and Electronic Health Record Data Technical Report.

References:

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

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)*
No diagram provided

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

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

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.

Electronic clinical data

When this measure is implemented the clinical data elements will be derived from hospital EHRs. We have empirically tested the feasibility of each of these data elements and have shown them to be consistently captured for nearly all adults hospitalized for acute care and extractable from hospital EHRs. In the instances where these data elements were missing from patients' medical records, we use multiple imputation to generate a range of plausible values for all missing data and estimate values for missing data.

In multiple imputation, missing variable values are predicted using other patient variables available. The predicted values are substituted for the missing values, which results in a full data set without any missing variables (the imputed data set). By repeating this process multiple times, we get multiple imputed data sets. We then conduct analyses on and obtain results for each imputed data set. The results based on multiple data sets are combined to produce the overall final results. Because we do not rely on one particular plausible version of the value, we have multiple versions of the plausible values. In general, imputed values are not intended to be "guesses" of what any particular missing value might be; instead, multiple imputation is used to preserve the important characteristics of the underlying data set and the inherent relationships among the variables in the data set. The multiple imputation represents a random sample of the missing values according to the association of the non-missing values of all the variables considered. The resulting inferences of multiple imputation are statistically valid and reflective of the uncertainty due to missing values [He & Belin, 2014; Carpenter & Kenward; Rubin, 1987].

Five copies of imputation datasets were produced for the analyses, and then the results based on these data separately were aggregated according to the standard statistical methods for presentation and for the measure score calculation. The approach to handling missing variables will be updated for implementation.

References:
He R, Belin T. Multiple imputation for high-dimensional mixed incomplete continuous and binary data. Stat. Med. 2014;33:2251–2262.

Carpenter J, Kenward M. Wiley: Multiple Imputation and its Application - [Internet]. [cited 2015 May 18];Available from: http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470740523.html

Rubin DB. Frontmatter [Internet]. In: Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons, Inc.; 1987 [cited 2015 May 15]. p. i–xxix.Available from: http://onlinelibrary.wiley.com/doi/10.1002/9780470316696.fmatter/summary

S.23. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Laboratory

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:

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

2. 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 and following discharge from index admission.

3. Patients' electronic health records: The clinical data elements used in the risk models for this measure will be derived from patients EHRs. The measure was developed and tested using data from EHRs.

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/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

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
Hybrid_HWR_NQF_Testing_Attachment_01-29-16_v1.0.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)
Last Updated 1/19/16

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Hybrid Hospital-Wide Readmission Measure with Claims and Electronic Health Record Data
Date of Submission: 1/29/2016
Type of Measure:

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

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 decision making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient

preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [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 that influence the measured outcome (but not factors related to disparities in care or the quality of care) 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. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.

16. 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: Electronically abstracted from EHRs

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 inpatient claims, as well as electronically and manually abstracted electronic health record (EHR) data from several health systems. Data set varies by testing type; see Section 1.7 for details.

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

The dates used vary by testing type; see Section 1.7 for details.

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

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

☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

The number of measured entities varies by testing type: see Section 1.7 for details.

1.6. How many and which patients 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*)

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.

Please note this model was developed using electronically extracted EHR data and merged inpatient claims data.

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

Measure Development and Testing:

For measure development and testing, we used a three-year dataset (**Dataset 1**) provided by Kaiser Permanente of Northern California. The dataset contained merged inpatient claims with clinical data elements derived from patients' electronic health records (EHRs). This health system uses an Epic EHR system. The merged data were provided for all patients discharged from any of their 21 acute care hospitals from January 1, 2010 through December 31, 2012. We randomly split the first 2-years of this dataset (January 1, 2010 – December 31, 2011) into a "development sample" (used to develop a risk model) and a "validation sample" (used to re-test the model); the random split was stratified by hospital and the measure's five specialty cohorts used to calculate the measure score. We re-tested the five risk models that make up the measure in the third year of data, from January 1, 2012 through December 31, 2012. This "2012 sample," was used to look for temporal stability in the models' performance.

In **Dataset 1**:

Number of admissions = 381,980

Number of hospitals = 21

Patient Descriptive Characteristics: mean age = 58 years; standard deviation = 21 years

%female = 62.6

The number of index admissions is listed below by specialty cohort.

Surgery:

- Development sample: 23,201 admissions
- Validation sample: 23,490 admissions
- 2012 sample: 25,471 admissions

Cardiorespiratory

- Development sample: 9,261 admissions
- Validation sample: 9,364 admissions
- 2012 sample: 9,070 admissions

Cardiovascular

- Development sample: 8,108 admissions
- Validation sample: 8,037 admissions
- 2012 sample: 8,338 admissions

Neurology

- Development sample: 4,400
- Validation sample: 4,348
- 2012 sample: 4,487

Medicine

- Development sample: 34,619
- Validation sample: 34,574
- 2012 sample: 35,747

For testing data Element and Measure Reliability Testing (Section 2a2)

Dataset 1

Validity Testing (Section 2b2)

Dataset 1 was used for measure validity testing.

Three datasets were used to assess the feasibility or validity of the clinical data elements used in the measure's risk models:

Dataset 1: (data element feasibility testing)

Dataset 2: (data element feasibility and validity testing)

Electronically extracted clinical data from three hospitals that used Cerner as their clinical EHR.

- Feasibility testing: 3 hospitals with 25,829; 56,812; and 29,586 admissions
- Validity testing: 1 hospital with data abstracted from 368 admissions (subset of admissions above)

Dataset 3: (data element validity testing)

Data were electronically extracted from 1 hospital that used GE Centricity as their clinical EHR

- Validity testing: 1 hospital with data abstracted from 391 admissions

For testing of measure exclusions (Section 2b3)

Dataset 1

For testing of measure risk adjustment (Section 2b4)

Dataset 1

For testing to identify meaningful differences in performance (Section 2b5)

Dataset 1

For testing of socioeconomic status (SES) factors and race in risk models (Section 2b4)

Dataset 4 and **Dataset 5** (Section 2b4)

The impact of socioeconomic factors was not directly tested in the Hybrid HWR measure due to lack of availability of EHR data from a nationally representative set of hospitals with patients who represent the full spectrum of socioeconomic status. Instead, we report results of testing done in the claims-only HWR measure.

Dataset 4: (2015 public reporting cohort version 4.0): Medicare Part A Inpatient Claims and Medicare Enrollment Database

Dates of Data: July 1, 2013 – June 30, 2014

Number of index admissions: 6,843,808

Number of hospitals: 4,772

Average age of patients: 78.3

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 SES.

Dataset 5: The American Community Survey (2008-2012)

We also used the Agency for Healthcare Research and Quality (AHRQ) SES index score derived from the American Community Survey (2008-2012) 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 (the American Community Survey [2008-2012]) (**Dataset 5**)

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).

The impact of socioeconomic factors was not directly tested in the Hybrid HWR measure due to lack of availability of EHR data from a nationally representative set of hospitals with patients who represent the full spectrum of socioeconomic status. Instead, we report results of testing done in the claims-only HWR measure (**Datasets 4** and **5**).

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 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 across multiple conditions is more limited, available studies suggest a consistent association between SES/race variables and risk of readmission. (Aseltine et al., 2015; Gu et al., 2014; Arbaje et al., 2008). 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 or 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 5-digit zip code and census block levels only. The data are not currently available at the 9-digit zip code level. In this submission, we present analysis using the 5-digit level. However, we are currently performing analysis at the census block level, the most granular level possible. We hope to present the results of the census block-level analysis to the committee.

References:

Arbaje AI, Wolff JL, Yu Q, Powe NR, Anderson GF, Boult C. Post discharge environmental and socioeconomic factors and the likelihood of early hospital readmission among community-dwelling Medicare beneficiaries. The Gerontologist. 2008;48(4):495-504.

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.

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.

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.

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.

Gu Q, Koenig L, Faerberg J, Steinberg CR, Vaz C, Wheatley MP. The Medicare Hospital Readmissions Reduction Program: potential unintended consequences for hospitals serving vulnerable populations. Health services research. 2014;49(3):818-837.

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.

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.

We also assess the reliability of the claims data elements by comparing model variable frequencies and odds ratios from logistic regression models in the development and validation samples (January 1, 2010-December 31, 2012, **Dataset 1**). We assessed the reliability of the clinical data elements by comparing rate of capture, and coefficients associated with each variable in the development and validation samples' risk models.

Measure Score Reliability

For test-retest reliability, we use the randomly split development and validation samples (**Dataset 1**) and calculated the measure for each hospital separately in each sample. 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 risk-standardized readmission rate (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.

Because reliability of the measure result could only be tested in a small sample of hospitals (n=21) and admissions in **Dataset 1,** testing will be repeated in a larger nationally representative set of hospitals prior to implementation. This testing will depend on implementation of hospital reporting of the core clinical data elements used in the measure's risk models.

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 intra rater, interrater and test–retest reliability of continuous measurements. Statistics in Medicine 2002; 21:3431-3446.

Shrout P, Fleiss J. Intra class 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

For the claims-derived data elements (**Dataset 1**), the frequency of model variables remained relatively constant between 2010 and 2012, with no model variables increasing or decreasing by more than 2%.

Clinical Data Element Capture

Table 1 shows the rate of capture of all of the clinical data elements used in the measures' risk models in the development, validation, and 2012 samples by cohort. Coefficients for the claims and clinical risk variables for all three samples can be found in data field S.2b (Data Dictionary or Code Table).

Table 1 Rate of Capture of all of the Clinical Data Elements (Dataset 1)

	Development Sample	Validation Sample	2012 Sample
Heart rate (% captured)			
Surgery/Gynecology	95.0	95.2	96.6
Cardiorespiratory	98.7	98.4	99.1
Cardiovascular	97.7	97.9	98.5
Neurology	97.7	98.1	98.6

Medicine	98.1	98.1	98.7
Systolic BP (% captured)			
Surgery/Gynecology	94.5	94.6	96.0
Cardiorespiratory	98.5	98.1	98.8
Cardiovascular	97.6	97.8	97.9
Neurology	97.7	98.1	98.5
Medicine	97.9	97.9	98.5
Respiratory Rate (% captured)			
Surgery/Gynecology	94.4	94.4	96.1
Cardiorespiratory	97.8	97.7	98.1
Cardiovascular	96.8	97.3	97.3
Neurology	97.0	97.3	97.6
Medicine	97.1	97.2	97.6
Temperature (% captured)			
Surgery/Gynecology	93.7	94.0	95.7
Cardiorespiratory	95.0	94.5	95.2
Cardiovascular	93.6	93.8	94.3
Neurology	93.1	94.0	94.5
Medicine	95.1	95.0	96.0
Weight (% captured)			
Surgery/Gynecology	94.1	94.1	95.7
Cardiorespiratory	93.7	93.6	94.9
Cardiovascular	94.3	94.7	95.2
Neurology	91.0	91.6	92.4
Medicine	91.1	91.2	92.3
Oxygen Saturation (% captured)			
Surgery/Gynecology	93.3	93.5	95.8
Cardiorespiratory	97.6	97.3	98.4
Cardiovascular	96.1	96.3	97.4
Neurology	96.2	96.6	97.4
Medicine	96.0	95.9	97.3
Hematocrit (% captured)			
Surgery/Gynecology	83.3	83.8	82.0
Cardiorespiratory	98.5	98.5	99.0
Cardiovascular	95.4	95.5	94.9
Neurology	97.8	97.9	98.0
Medicine	97.6	97.6	98.0
WBC Count (% captured)			
Surgery/Gynecology	79.4	80.1	78.6
Cardiorespiratory	98.5	98.4	98.9

Cardiovascular	95.3	95.3	94.9
Neurology	97.8	97.8	97.9
Medicine	97.4	97.4	97.8
Potassium (% captured)			
Surgery/Gynecology	70.6	71.1	70.0
Cardiorespiratory	96.8	96.5	97.1
Cardiovascular	93.6	93.6	93.5
Neurology	96.1	95.9	95.8
Medicine	95.6	95.6	95.8
Sodium (% captured)			
Surgery/Gynecology	71.8	72.3	71.1
Cardiorespiratory	98.7	98.5	99.1
Cardiovascular	95.0	95.2	94.8
Neurology	98.0	98.0	98.3
Medicine	97.4	97.4	97.9
Bicarbonate (% captured)			
Surgery/Gynecology	71.3	71.7	70.8
Cardiorespiratory	98.8	98.5	99.1
Cardiovascular	95.0	95.3	94.8
Neurology	98.0	97.9	98.2
Medicine	97.4	97.4	97.8
Creatinine (% captured)			
Surgery/Gynecology	72.0	72.2	71.5
Cardiorespiratory	98.7	98.5	99.1
Cardiovascular	95.2	95.3	94.8
Neurology	98.1	98.0	98.3
Medicine	97.4	97.4	97.9
Glucose (% captured)			
Surgery/Gynecology	71.1	71.4	70.5
Cardiorespiratory	98.6	98.4	99.0
Cardiovascular	94.9	95.1	94.6
Neurology	98.0	97.9	98.2
Medicine	97.3	97.3	97.8

Measure Score Reliability Results

In **Dataset 1**, there were 81,589 in the development sample and 79,813 in the validation sample. The agreement between the two RSRRs for each hospital was 0.688, which according to the conventional interpretation is "moderate" (Landis & Koch, 1977).

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 over time of the odds ratios or variable coefficients and the model variable frequencies and rates of capture for clinical data elements suggests that the underlying data elements are reliable.

Measure Testing

For the hospital event rate based on the patient binomial outcomes like readmission (Yes/No), an ICC value of 0-0.2 indicates poor agreement; 0.3-0.4 indicates fair agreement; 0.5-0.6 indicates moderate agreement; 0.7-0.8 indicates strong agreement; and >0.8 indicates almost perfect agreement. The ICC of 0.688 is moderate to strong in **Dataset 1.**

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)*

Validity of EHR Data Elements

Several critical clinical data elements used in the measure's risk models were derived from patients' electronic medical records. When this measure is implemented, CMS intends to obtain these critical data elements from hospital EHRs and merge the data with claims data to calculate and report measure results. We tested the validity of electronic extraction of these critical data elements as part of a more comprehensive evaluation of a larger set of core clinical data elements (CCDEs). The CCDE are a set of 21 EHR data elements that are captured on most adults (plus Troponin, which is a condition-specific CCDE for patients with acute myocardial infarction) admitted to acute care hospitals, are easily extracted from EHRs, and can be used to risk adjust hospital outcome measures for a variety of conditions and procedures. All of the critical data elements used in the Hybrid HWR measure are included in the CCDE. Testing of the CCDE involved three phases: 1) identification of potentially feasible clinical data through qualitative assessment, 2) empirical feasibility testing of several clinical data elements electronically extracted from two large multi-facility health systems, and 3) validity testing of the CCDE at an additional health system.

Phase 1: Identification of potentially feasible clinical data through qualitative assessment

In order to identify the CCDEs for risk adjustment of hospital outcome measures for adult patients, we first conducted a qualitative assessment of the reliable capture, accuracy, and extractability of categories and subcategories of clinical data as defined by the Quality Data Model (QDM) (e.g., vital signs, laboratory test results). We established a set of criteria to assess the consistency of data capture, relevance to hospital quality

measures, and extractability from health records.

Data Capture Criteria:

- Obtained consistently under current practice. Routinely collected for patients admitted to the hospital under current clinical practice and EHR workflows.
- Captured with a standard definition. Consistent conceptual understanding, method of collection, and units of measurement.
- Entered in a structured field. Captured in numerical, pseudo-numerical, or list format.
- Data Extraction Criteria:
- Encoded consistently. Can be linked to a standard and uniform coding structure such as ICD-9 or Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT).
- Extractable from the EHR. Can be readily and consistently identified and exported from current EHR databases.
- Exported with metadata. Additional information such as time stamps and reference values that are needed for interpretation are consistently available.

These criteria are aligned with those established in the NQF's eMeasure Feasibility Assessment Report as well as the NQF feasibility criteria (see included Data Element Feasibility Scorecard). The NQF report emphasized four key aspects of feasibility. First, data should be structured or easily converted to a structured and interpretable format. Second, data should be accurate. Third, data should be easily associated with a standard set of codes to ensure consistent extraction across EHR environments. Finally, data should not require changes to current clinical practice or workflows.

We then convened a Technical Expert Panel (TEP) to apply these criteria to categories and subcategories (data types) of clinical data based on the Quality Data Model (QDM). We asked TEP members to consider only the context of adult hospitalized patients when making their assessments. Data categories and subcategories were rated on each feasibility criterion independently by TEP members. The ratings were tallied and TEP members met to discuss and resolve areas of disagreement. Through this process the TEP identified a list of data subcategories that were potentially feasible for use in hospital outcome measures. The CCDE were derived from only those subcategories for which the TEP reached consensus agreement on feasibility.

Phase 2: Empirical feasibility testing using a large multi-site database

In **Dataset 1**, we next directly examined the feasibility of clinical data elements from the subcategories identified by the TEP as feasible (for all adult inpatient admissions). We examined all admissions in **Dataset 1** between 2010 and 2011. We analyzed clinical data elements to determine whether they were captured in a numerical field, the consistency and timing of capture, and the accuracy of the data elements. We examined the data elements across conditions, hospitals, and point of hospital entry. We tested several data elements that met the feasibility criteria in models predicting 30-day mortality following admission for several common medical conditions. The complete list of 21 (plus Troponin) CCDE were derived from these analyses.

To verify that the findings from our analysis of **Dataset 1** were generalizable to other hospitals and electronic health systems, we partnered with Premier Inc., a collaborative healthcare alliance of approximately 2,900 U.S. community hospitals focused on measuring and improving their members' quality outcomes and safely reducing healthcare costs. We administered a survey to four of their member hospital systems that used a variety of EHR systems to confirm the availability of the clinical data elements. Additionally, we assessed the rate and timing of capture of the data elements identified in **Dataset 1** in ERH abstracted from a second health system consisting

of three hospitals (**Dataset 2**).

Phase 3: Validity testing at two hospital sites the CCDE (including critical data elements for the Hybrid HWR measure)

In Phase 3, we developed electronic specifications (e-specifications) using the Measure Authoring Tool (MAT), and analyzed extracted data from EHRs. We assessed the ability of hospitals to use the e-specifications to query and electronically extract CCDEs from the EHR, for all adult inpatient admissions occurring over the course of one year. Validity testing assessed the accuracy of the electronically extracted CCDEs compared to the same CCDEs gathered through manual abstraction (from the EHR) in a subset of 368 charts identified in the data query in **Dataset 2**, and 391 charts identified in the data query in **Dataset 3**.

Chart Abstraction: We calculated the number of admissions that needed to be randomly sampled from the EHR dataset and manually abstracted to yield a statistical margin of error (MOE) of 5% and a confidence level of 95% for the match rates between the two data sources. Sites then used an Access-based manual abstraction tool provided (along with training) to manually abstract the CCDEs from the random samples of the medical records identified through the EHR data query. The manual chart abstraction data is considered the "gold standard" for the purpose of this analysis.

Validity Testing: We conducted validity testing on the critical EHR data elements in the Hybrid HWR measure. For each continuous data element, we were only interested in the case where the electronic abstraction value exactly matched the manual abstraction value. We therefore only calculated the raw agreement rate between data from electronic and manual chart abstraction. For simple data values, we believe taking this approach, as compared to reporting statistical tests of accuracy, better reflects the concept of matching exact data values rather than calculated measure results. Therefore, we do not report statistical testing of the accuracy of the EHR derived data value as compared with the abstracted value. Instead, we counted only exact matches in the data value as well as the time and date stamp associated with that value when we calculated the match rate. The 95% confidence level was established based on the sample size and reflects the exact match rate using these criteria.

<u>Validation Against Other Risk Models and Registry Data</u>

The Hybrid model we developed uses a combination of claims data (demographics, comorbidities, and patient medical history) and electronic clinical data (laboratory results and vital signs).

We compared the Hybrid risk model to the harmonized claims-only risk model used in the publicly reported Hospital-Wide All Cause Unplanned Readmission Measure. Both models use inpatient administrative claims data to derive the cohort, to derive risk variables, and to assess the unplanned readmission outcome.

Measure validity was tested through comparison of this Hybrid risk adjustment model with claims-only risk-adjustment model, and through use of established measure development guidelines.

For the derivation of both risk models, we used **Dataset 1**. Both the Hybrid and claims-only risk models used the same inclusion/exclusion criteria and a risk-adjustment (statistical modeling) strategy and only differed with respect to the risk variables used. We compared the model discrimination and the correlation in hospital performance results for the two models.

<u>Validity Indicated by Established Measure Development Guidelines</u>

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 (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" (Krumholz, Brindis, et al., 2006).

Validity as Assessed by External Groups

Following completion of the preliminary model, we solicited public comment on the measure through the CMS site link: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/CallforPublicComment.html. The public comments were then posted publicly for 30 days. The resulting input was taken into consideration during the final stages of measure development and contributed to minor modifications to the measure.

References:

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

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

Krumholz HM, Brindis RG,Brush JE, et al. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. *Circulation.* January 24, 2006 2006;113(3):456-462.

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

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

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

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

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

ICD-9 to ICD-10 Conversion

Statement of Intent

[X] Goal was to convert this measure to a new code set, fully consistent with the intent of the original measure.

[ ] Goal was to take advantage of the more specific code set to form a new version of the measure, but fully consistent with the original intent.

[ ] The intent of the measure has changed.

Process of Conversion

ICD-10 codes were initially identified using 2015 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. An ICD-9 to ICD-10 crosswalk is attached in field S.2b. (Data Dictionary or Code Table).

We have also examined the updated ICD-9 Map to AHRQ Conditions Categories (CCS) crosswalk to the ICD-10 CCS map provided by AHRQ in preparation for the inclusion of ICD 10 data in this measure. Please refer to the ICD-10 CCS map on the AHRQ website.

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

Validity of EHR Data Elements

Phase 1: TEP Survey Results

The TEP identified seven subcategories of EHR data that they considered feasible for adult hospitalized patients. They were: Encounter Performed, Patient Characteristics including birth date and sex, Physical Examination Findings for vital signs only, Diagnostic Study Order, Diagnostic Study Performed, Medication Discharge, and Laboratory Test Result. We limited the CCDE to data elements to only four categories: Encounter Performed, Patient Characteristics, Physical Examination Findings for vital signs only, and Laboratory Test Results, which are unlikely to be reflective of care quality and therefore are thought to be both feasible to extract and appropriate for risk adjustment.

Phase 2: Feasibility Testing Results

Datasets 1 and 2: The consistency of data capture of the critical data elements included in the Hybrid HWR measure for all adult hospitalized patients in **two health systems** with different EHR environments (EPIC and Cerner). Results presented in Table 2 and Table 3 are from EPIC system; results presented in Table 4 are from Cerner system. These tables show consistent capture of all the clinical data elements used in the risk models.

In addition, the four Premier member hospitals all reported that the CCDE were: captured in the inpatient EHR; captured in the emergency department EHR; recorded in a structured format; extracted for reporting; extracted for other purposes; and time and date stamps capture.

Table 2. Proportion of Episodes with Captured Vital Signs at Various Time points (Dataset 1)

Vital Sign Finding – Full Cohort	Total with Finding and Timestamp %	Within 2 Hours %	Within 6 Hours %	Within 12 Hours %
Basic vital signs				
Heart rate	99.7	96.8	99.4	99.6
Systolic blood pressure	99.7	96.7	99.3	99.6
Diastolic blood pressure	99.7	96.7	99.3	99.6
Respiratory rate	99.7	95.8	99.1	99.6
Temperature	99.7	93.7	98.5	99.5
Oxygen saturation	98.2	86.0	92.6	95.4
Weight	92.5	80.2	85.2	88.8

Table 3. Proportion of Admissions with Laboratory Results at Various Time Points (Dataset 1)

Lab Test Result – Full Cohort	Total with Result and Timestamp (%)	Within 2 Hours (%)	Within 6 Hours (%)	Within 12 Hours (%)	Within 24 Hours (%)
Hemoglobin	92.7	61.2	72.7	77.3	90.6
Hematocrit	92.8	61.6	73.8	78.0	90.8
Platelets	92.0	61.1	72.4	76.5	89.8
WBC count	92.0	61.1	72.4	76.5	89.8
Potassium	71.3	49.3	57.2	60.2	69.4
Sodium	71.6	49.3	57.3	60.3	69.6
Chloride	71.1	49.3	56.1	59.4	69.0
Bicarbonate	71.2	49.2	56.8	59.8	69.2
Glucose	72.0	49.7	57.6	60.6	70.0
Troponin	32.2	25.6	28.7	30.0	30.6

Table 4. Percent Captured per Data Element per Hospital (Dataset 2)

Data Element/ CCDE	% Captured Hospital 1	% Captured Hospital 2	% Captured Hospital 3

Data Element/ CCDE	% Captured Hospital 1	% Captured Hospital 2	% Captured Hospital 3
Heart Rate (BPM)	94.27	80.54	84.48
Systolic Blood Pressure (mmHG)	94.42	80.88	84.07
Diastolic Blood Pressure (mmHG)	94.29	80.81	83.99
Respiratory Rate (BPM)	93.58	87.69	84.26
Temperature (C)	92.43	86.48	83.77
Oxygen Saturation (%)	83.22	78.80	83.61
Weight (Kg)	99.01	98.66	98.61
Hemoglobin	96.19	96.29	91.47
Hematocrit	95.77	96.29	91.25
Platelet	89.40	96.52	92.26
WBC Count	89.12	96.25	91.11
Potassium	78.44	81.08	82.03
Sodium	78.29	81.03	81.78
Chloride	78.28	81.03	81.78
Bicarbonate	11.53	11.23	12.29
Glucose	77.64	80.93	81.32
Troponin	29.67	32.10	15.71

Phase 3: Further Feasibility and Validity Testing Results

Chart abstraction for validity testing was done in **Dataset 2** and **Dataset 3.** Table 5 demonstrates the comparison between electronic and manual abstraction of data in the two health systems.

Table 5. Percent Agreement and Confidence Internals (Dataset 2 and 3)

Data Element/ CCDE	% Agreement Between Datasets (Number Matching/ Total Records With A Data Value)	95% Confidence Internal for Agreement	% Present in Electronic Extraction, Missing in Manual Abstraction (N)	% Present in Manual Abstraction, Missing in Electronic Extraction (N)	% Missing in Both Electronic Extraction and Manual Abstraction (N)
Dataset 2 (n=368)					
Heart rate (BPM)	95.55 (322/337)	92.76 - 97.49	0 (0.00)	8.42 (31)	0 (0.00)
Syst Blood Pressure	94.67 (320/338)	91.71 - 96.81	0 (0.00)	8.15 (30)	0 (0.00)

Data Element/ CCDE	% Agreement Between Datasets (Number Matching/ Total Records With A Data Value)	95% Confidence Internal for Agreement	% Present in Electronic Extraction, Missing in Manual Abstraction (N)	% Present in Manual Abstraction, Missing in Electronic Extraction (N)	% Missing in Both Electronic Extraction and Manual Abstraction (N)
(mmHG)					
Diast Blood Pressure (mmHG)	94.38 (319/338)	91.36 - 96.58	0 (0.00)	8.15 (30)	0 (0.00)
Respiratory Rate (BPM)	94.89 (316/333)	91.95 - 97.00	0 (0.00)	9.51 (35)	0 (0.00)
Temperature (C)	95.41 (312/327)	92.55 - 97.41	0 (0.00)	10.87 (40)	0.27 (1)
Oxygen Saturation (%)	94.68 (285/301)	91.51 - 96.93	0.27 (1)	14.40 (53)	3.53 (13)
Weight (Kg)*	14.66 (51/348)	11.11 - 18.81	1.09 (4)	3.53 (13)	0.82 (3)
Hemoglobin (g/dL)	96.50 (331/343)	93.97 - 98.18	0.82 (3)	3.80 (14)	2.17 (8)
Hematocrit (%)	96.19 (328/341)	93.57 - 97.95	0.82 (3)	3.80 (14)	2.72 (10)
Platelet (x10(9)/L)	96.88 (310/320)	94.33 - 98.49	0.82 (3)	4.62 (17)	7.61 (28)
WBC Count (x10(9)/L)	96.56 (309/320)	93.93 - 98.27	0.82 (3)	4.62 (17)	7.61 (28)
Potassium (meq/L)	97.22 (280/288)	94.60 - 98.79	0 (0.00)	5.16 (19)	16.58 (61)
Sodium (meq/L)	97.21 (279/287)	94.58 - 98.79	0 (0.00)	5.43 (20)	16.58 (61)
Chloride (meq/L)	97.21 (279/287)	94.58 - 98.79	0 (0.00)	5.43 (20)	16.58 (61)
Bicarbonate (meq/L)	14.81 (8/54)	6.62 - 27.12	0.27 (1)	68.21 (251)	16.85 (62)
Glucose (mg/dL)	96.14 (274/285)	93.20 - 98.06	0 (0.00)	5.43 (20)	17.12 (63)
Troponin (ng/mL)	93.33 (98/105)	86.75 - 97.28	4.08 (15)	0.54 (2)	66.85 (246)
Dataset 3 (n=391)					
Heart rate (BPM)	57.45 (135/235)	50.85 - 63.85	0 (0.00)	39.39 (154)	0.51 (2)
Syst Blood Pressure (mmHG)	60.26 (138/235)	53.61 - 66.65	0 (0.00)	40.92 (160)	0.51 (2)
Diast Blood Pressure (mmHG)	60.09 (137/228)	53.41 - 66.50	0.26 (1)	40.92 (160)	0.51 (2)
Respiratory Rate (BPM)	70.14 (155/221)	63.63 - 76.09	0 (0.00)	42.71 (167)	0.77 (3)
Temperature (C)	79.09 (174/220)	73.11 - 84.27	0 (0.00)	42.46 (166)	1.28 (5)
Oxygen Saturation (%)	56.65 (115/203)	49.53 - 63.57	0.26 (1)	46.80 (183)	1.02 (4)
Weight (Kg)	84.41 (157/186)	78.38 - 89.30	0.26 (1)	48.34 (189)	3.84 (15)
Hemoglobin	88.78 (87/98)	80.80 - 94.26	0.26 (1)	58.06 (227)	16.62 (65)

Data Element/ CCDE	% Agreement Between Datasets (Number Matching/ Total Records With A Data Value)	95% Confidence Internal for Agreement	% Present in Electronic Extraction, Missing in Manual Abstraction (N)	% Present in Manual Abstraction, Missing in Electronic Extraction (N)	% Missing in Both Electronic Extraction and Manual Abstraction (N)
(g/dL)					
Hematocrit (%)	91.67 (88/96)	84.24 - 96.33	0.26 (1)	58.31 (228)	16.88 (66)
Platelet (x10(9)/L)	94.68 (89/94)	88.02 - 98.25	0.26 (1)	59.08 (231)	16.62 (65)
WBC Count (x10(9)/L)	94.62 (88/93)	87.90 - 98.23	0.26 (1)	59.34 (232)	16.62 (65)
Potassium (meq/L)	86.75 (72/83)	77.52 - 93.19	0 (0.00)	62.66 (245)	16.11 (63)
Sodium (meq/L)	91.46 (75/82)	83.20 - 96.50	0 (0.00)	61.64 (241)	17.39 (68)
Chloride (meq/L)	97.56 (80/82)	91.47 - 99.70	0 (0.00)	60.61 (237)	18.41 (72)
Bicarbonate (meq/L)	29.41 (5/17)	10.31 - 55.96	0.26 (1)	77.49 (303)	17.90 (70)
Glucose (mg/dL)	95.12 (78/82)	87.98 - 98.66	0 (0.00)	63.17 (247)	15.86 (62)
Troponin (ng/mL)	82.61 (19/23)	61.22 - 95.05	0 (0.00)	13.30 (52)	80.82 (316)

A post-validation review of the code used by the hospital in **Dataset 3**, revealed that the hospital experienced a number of errors. The most significant of which was extracting data only within an incorrect two-hour window for laboratory test results (the correct window was 24 hours). Additionally, physical exam (vital signs) data were extracted based on the date/time that results were documented rather than the date/time the physical exams were performed, driving down the accuracy of these data. However, post-validation review of the code used by the hospital in **Dataset 2** showed no such errors in the query executed. As a result the match rate was much higher.

Validation against Claims-Only Risk Model **(Dataset 1)**

We estimated hospital-level RSRRs using the corresponding hierarchical logistic regression for each of the models in the linked patient sample. We then examined the linear relationship between the estimates using regression techniques and weighting by the total number of cases in each hospital. The Pearson correlation coefficient of the standardized rates from the claims-only risk-adjustment model and the Hybrid risk-adjustment model in the Development Sample of **Dataset 1** is 0.9902.

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 of EHR Data Elements

Feasibility Testing (Phases 1-3)

The critical data elements were demonstrated to be feasible through consensus of the TEP and direct examination of EHR data establishing consistent capture of the CCDE among adult hospitalized patients. In addition, we established the validity of electronic extraction of the

CCDE demonstrated by the high match rate when comparing EHR extracted and manual medical record abstracted CCDE values.

Measure Validity (**Dataset 1**)

The results between the Hybrid HWR model and the claims-only risk model were nearly identical. In addition, the high correlation among the RSRRs calculated from all models shows that each model provides a similar or consistent measure result for hospitals.

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*)

All exclusions were determined by careful clinical review and have been made based on clinically relevant decisions and to ensure accurate calculation of the measure. To ascertain impact of exclusions on the cohort, we examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion (**Dataset 1**). These exclusions are consistent with similar NQF-endorsed outcome measures. Rationales for the exclusions are detailed in data field S.10 (Denominator Exclusions).

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. This data is from the original initial cohort (n= 251,006) of index admissions.

Exclusion	N	%
Discharged Against Medical Advice	679	0.27%
Cancer Treatment	6,356	2.53%
Psychiatric Treatment	593	0.24%
Rehabilitation	885	0.35%

These exclusions represent only 3.88% of the initial cohort. We do not report frequency of distribution of exclusions across measured entities due to the minimal impact of the exclusions on the measure cohort. The final cohort was 242,515 index admissions.

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*)

1. Patients discharged against medical advice (AMA) account for 0.27% of all index admissions excluded from the initial index cohort. This exclusion is needed for acceptability of the measure to hospitals, who do not have the opportunity to adequately deliver full care and prepare the patient for discharge.
2. Patients admitted for primary psychiatric diagnoses account for 0.24% of all index admissions excluded from the initial cohort. This exclusion is needed because these patients are typically cared for in separate psychiatric or rehabilitation centers which are not comparable to acute care hospitals.
3. Patients admitted for rehabilitation account for 0.35% of all index admissions excluded from the initial cohort. This exclusion is needed because patients admitted for rehabilitation are not admitted for treatment of acute illness and the care provided in rehabilitation centers is not comparable to care provided in acute care hospitals.
4. Patients admitted for medical treatment of cancer account for 2.53% of all index admissions excluded from the initial cohort. Admissions for treatment of cancer are associated with a very different mortality and readmission risk compared with admissions to other Inpatient Prospective Payment Systems (IPPS) hospitals for treatment of other diseases. Additionally, outcomes for these admissions do not correlate well with outcomes for other types of admissions. (Patients with cancer who are admitted for other diagnoses or for surgical treatment of their cancer remain in 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** Click here to enter number of factors **risk factors**

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

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

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (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 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 and not related to disparities*)

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 estimates hospital-level 30-day all-cause RSRRs using hierarchical logistic regression models. In brief, the approach simultaneously models data at the patient and hospital levels to account for variance in patient outcomes within and between hospitals (Normand S-LT, Shahian DM, 2007). At the patient level, it models the log-odds of hospital readmission within 30 days of discharge using age, selected clinical covariates, and a hospital-specific intercept. At the hospital level, the approach models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of a readmission at the hospital, after accounting for patient risk. The hospital-specific intercepts are given a distribution to account for the clustering (non-independence) of patients within the same hospital (Normand S-LT, Shahian DM., 2007). If there were no differences among hospitals, then after adjusting for patient risk, the hospital intercepts should be identical across all hospitals.

Admissions are assigned to one of five mutually exclusive specialty cohort groups consisting of related conditions or procedures. For each specialty cohort group, the standardized readmission ratio (SRR) is calculated as the ratio of the number of "predicted" readmissions to the number of "expected" readmissions at a given hospital. For each hospital, the numerator of the ratio is the number of readmissions within 30 days predicted based on the hospital's performance with its observed case mix and service mix, and the denominator is the number of readmissions expected based on the nation's performance with that hospital's case mix and service mix. This approach is analogous to a ratio of "observed" to "expected" used in other types of statistical analyses. It conceptually allows a particular hospital's performance, given its case mix and service mix, to be compared to an average hospital's performance with the same case mix and service mix. Thus, a lower ratio indicates lower-than-expected readmission rates or better quality, while a higher ratio indicates higher-than-expected readmission rates or worse quality.

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

The specialty cohort SRRs are then pooled for each hospital using a volume-weighted geometric mean to create a hospital-wide composite SRR. The composite SRR is multiplied by the national observed readmission rate to produce the RSRR.

Data Source

To select candidate variables for the Hybrid risk model, we began with the list of all administrative claims-based risk-adjustment variables included in the currently publicly reported Hospital-Wide All Cause Unplanned Readmission Measure. These candidate variables were derived from: the index admission, with comorbidities identified from the index admission secondary diagnoses (excluding potential complications) and 12-month pre-index inpatient data (for any condition). In identifying these variables for the current publically reported HWR measure, we sought to develop a model that was parsimonious, using a grouper that is in the public domain for the 15,000+ ICD-9-CM codes we started with the 189 diagnostic groups included in the Hierarchical Condition Category (HCC) clinical classification system (Pope et al., 2000). The HCC clinical classification system was developed for CMS in preparation for all-encounter risk adjustment for Medicare Advantage (managed care) plans and represented a refinement of an earlier risk-adjustment method based solely on principal inpatient diagnosis. The HCC model makes use of all physician and hospital encounter diagnoses and was designed to predict a beneficiary's expenditures based on the total clinical profile represented by all of his/her assigned HCCs. Under the HCC algorithm, the 15,000+ ICD-9-CM diagnosis codes are first assigned to one of 804 mutually exclusive groupings ("DxGroups") and then subsequently aggregated into 189 condition categories (CCs). During development, we used the April 2008 version of the ICD-9-CM to CC assignment map, which is maintained annually by CMS and posted at www.qualitynet.org. We do not use the hierarchy and therefore refer to the CCs rather than HCCs. The HWR risk-adjustment models use only inpatient claims data (history and current) in order to make it feasible to implement with Medicare data, and to make it applicable to all-payer data, which are typically restricted to inpatient claims.

We also used the core clinical data elements CCDE, the EHR-derived data elements used in the measure. The CCDE include the first vital signs captured within 2 hours of the start of the encounter and the first set of results for several basic laboratory tests captured within 24 hours of the start of the encounter (for example, complete blood count and basic chemistry panel). A file that contains a list of the ICD-9-CM codes and their groupings into CCs as well as a list of the CCDE is attached in data field S.2b (Data Dictionary and Code Table).

Complications occurring during hospitalization are not comorbid illnesses, may reflect hospital quality of care, and therefore should not be used for risk adjustment. Hence, conditions that may represent adverse outcomes due to care received during the index hospital stay are not included in the risk-adjusted model.

Service mix adjustment: The measure includes many different discharge condition categories that differ in their baseline readmission risks. In addition, hospitals differ in their relative distribution of these condition categories (service mix). To adjust for service mix, the measure uses an indicator variable for the discharge condition category in addition to risk variables for comorbid conditions. The models include a condition-specific indicator for all condition categories with sufficient volume (defined as those with more than 1,000 admissions nationally in a given year for Medicare FFS data) as well as a single indicator for conditions with insufficient volume in each model.

Although the 5 risk models use a common set of claims variables, the CCDE variables are not the same across specialty cohort models. Only those data elements that are statistically significant in each individual model are included. We estimate a hierarchical logistic regression model for each specialty cohort separately, and the coefficients associated with each variable may vary across specialty cohorts.

The final set of risk-adjustment variables is listed on the Submission form in item S.14 and in the Data Dictionary and Code Table attached in data filed S.2b

References:

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.

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

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

Socioeconomic Status (SES) Factors and Race

We selected variables representing 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 HF Readmission

To examine the relationship between SES and race variables and hospital 30-day, hospital-wide, all-cause, unplanned readmission following hospitalization, 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 readmission across multiple conditions. One hundred and sixty nine articles were initially reviewed, and 155 studies were excluded from full-text review based on the above criteria. Studies indicate that SES and race variables were associated with increased risk of readmission across multiple

major illnesses and conditions (Aseltine RH, et al., 2015; Mitchell SE, et al., 2012; Odonkor CA, et al., 2015; Herrin J, et al., 2015; Gu Q, et al., 2014, Kim H, et al., 2010; Kangovi S, et al., 2012; Iloabuchi TC, 2014; Beck AF, et al., 2012; Arbaje AI, et al., 2008; Hu J, 2014; Nagasako EM, et al., 2014; Joynt, K. E., et al., 2013), though there may not be a significant effect on hospital-level profiling (Blum et al., 2014).

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 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

Using the data from the Hospital-Wide All-Cause Readmission Measure for the 2015 reporting year (**Dataset 4** and **Dataset 5** [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:

Arbaje AI, Wolff JL, Yu Q, Powe NR, Anderson GF, Boult C. Postdischarge environmental and socioeconomic factors and the likelihood of early hospital readmission among community-dwelling Medicare beneficiaries. *The Gerontologist.* 2008;48(4):495-504.

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.

Beck AF, Simmons JM, Huang B, Kahn RS. Geomedicine: area-based socioeconomic measures for assessing risk of hospital reutilization among children admitted for asthma. *American journal of public health.* 2012;102(12):2308-2314.

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.

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.

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.

Gu Q, Koenig L, Faerberg J, Steinberg CR, Vaz C, Wheatley MP. The Medicare Hospital Readmissions Reduction Program: potential unintended consequences for hospitals serving vulnerable populations. *Health services research.* 2014;49(3):818-837.

Herrin J, St Andre J, Kenward K, Joshi MS, Audet AM, Hines SC. Community factors and hospital readmission rates. *Health services research.* 2015;50(1):20-39.

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.

Iloabuchi TC, Mi D, Tu W, Counsell SR. Risk factors for early hospital readmission in low-income elderly adults. *Journal of the American Geriatrics Society.* 2014;62(3):489-494.

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.

Kangovi S, Grande D, Meehan P, Mitra N, Shannon R, Long JA. Perceptions of readmitted patients on the transition from hospital to home. *Journal of hospital medicine.* 2012;7(9):709-712.

Kim H, Ross JS, Melkus GD, Zhao Z, Boockvar K. Scheduled and unscheduled hospital readmissions among patients with diabetes. *The American journal of managed care.* 2010;16(10):760-767.

Mitchell SE, Sadikova E, Jack BW, Paasche-Orlow MK. Health literacy and 30-day postdischarge hospital utilization. *Journal of health communication.* 2012;17 Suppl 3:325-338.

Nagasako EM, Reidhead M, Waterman B, Dunagan WC. Adding socioeconomic data to hospital readmissions calculations may produce more useful results. *Health affairs (Project Hope).* 2014;33(5):786-791.

Odonkor CA, Hurst PV, Kondo N, Makary MA, Pronovost PJ. Beyond the Hospital Gates: Elucidating the Interactive Association of Social Support, Depressive Symptoms, and Physical Function with 30-Day Readmissions. *American journal of physical medicine & rehabilitation / Association of Academic Physiatrists.* 2015;94(7):555-567.

Reames BN, Birkmeyer NJ, Dimick JB, Ghaferi AA. Socioeconomic disparities in mortality after cancer surgery: failure to rescue. JAMA surgery 2014; 149:475-81.

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.

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

Candidate and final model variables, with a corresponding CC to ICD-9 code map, are listed in the accompanying Excel Data Dictionary. The model variables from the original HWR measure are forced into the final model to align with that measure. The CCDE variables included in the model use logistic regression models with stepwise selection method (P=0.05).

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 SDS factors in the claims-only HWR cohort varies across measured entities. The median percentage of dual eligible patients is 14.9% (interquartile range [IQR] 9.8%-22.6

%). The median percentage of African-American patients is 2.2% (IQR 0.0%-9.4%). The median percentage of low SES AHRQ indicator patients is 19.4% (IQR 5.0%-57.3%).

Empirical association with the outcome (bivariate)

The patient-level observed hospital-wide readmission rate is higher for dual eligible patients, 19.28%, compared with 14.83% for all other patients. The readmission rate for African-American patients was also higher at 19.16% compared with 15.10% for patients of all other races. Similarly, the readmission rate for patients in the lowest SES quartile by AHRQ Index was 16.81% compared with 15.05% for all other patients.

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. 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 small. The c-statistic is essentially unchanged with the addition of any of these variables into the model. Furthermore, we find that 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 0.004% (interquartile range [IQR] -0.017% – 0.024%; minimum -0.309% – maximum 0.135%) with a correlation coefficient between RSRRs for each hospital with and without dual eligibility added of 0.99836. The median absolute change in hospitals' RSRRs when adding a race indicator is 0.011% (IQR -0.010% – 0.033%; minimum -0.671% – maximum 0.130%) with a correlation coefficient between RSRRs for each hospital with and without race added of 0.99814. The median absolute change in hospitals' RSRRs when adding a low SES AHRQ indicator is 0.007% (IQR -0.033% – 0.036%; minimum -0.322% – maximum 0.135%) with a correlation coefficient between RSRRs for each hospital with and without low SES added of 0.99691.

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, race, and low AHRQ SES Index effects were significantly associated with each of the hospital wide readmission models (Medicine, Surgery, Cardiorespiratory, Cardiovascular, and Neurology) in the decomposition analysis. If the dual eligible, race, or low AHRQ SES Index variables are used 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.

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.

Table 6. HWR Decomposition Analysis

Parameter	Estimate (Standard Error)	P-value
Dual Eligible – Patient-Level – Medicine	0.0599 (0.00433)	<.0001
Dual Eligible – Hospital-Level – Medicine	0.3207 (0.0177)	<.0001
Dual Eligible – Patient-Level – Surgery	0.1483 (0.00794)	<.0001
Dual Eligible – Hospital-Level – Surgery	0.4743 (0.0332)	<.0001
Dual Eligible – Patient-Level – Cardio Respiratory	0.1043 (0.00634)	<.0001
Dual Eligible – Hospital-Level – Cardio Respiratory	0.4148 (0.0269)	<.0001
Dual Eligible – Patient-Level – Cardiovascular	0.1607 (0.0101)	<.0001
Dual Eligible – Hospital-Level – Cardiovascular	0.5318 (0.0418)	<.0001
Dual Eligible – Patient-Level – Neurology	0.0874 (0.0129)	<.0001
Dual Eligible – Hospital-Level – Neurology	0.4997 (0.0526)	<.0001
African American – Patient-Level – Medicine	0.0374 (0.00558)	<.0001
African American – Hospital-Level – Medicine	0.3208 (0.0119)	<.0001
African American – Patient-Level – Surgery	0.0959 (0.0103)	<.0001
African American – Hospital-Level – Surgery	0.4423 (0.0214)	<.0001
African American – Patient-Level – Cardio Respiratory	0.0470 (0.00884)	<.0001
African American – Hospital-Level – Cardio Respiratory	0.3386 (0.0186)	<.0001
African American – Patient-Level – Cardiovascular	0.0763 (0.0131)	<.0001
African American – Hospital-Level – Cardiovascular	0.3501 (0.0269)	<.0001
African American – Patient-Level – Neurology	0.1200 (0.0155)	<.0001
African American – Hospital-Level – Neurology	0.5252 (0.0331)	<.0001
AHRQ SES Index – Patient-Level – Medicine	0.0249 (0.00444)	<.0001
AHRQ SES Index – Hospital-Level – Medicine	0.0788 (0.00653)	<.0001
AHRQ SES Index – Patient-Level – Surgery	0.0349 (0.00689)	<.0001
AHRQ SES Index – Hospital-Level – Surgery	0.1254 (0.0120)	<.0001

AHRQ SES Index – Patient-Level – Cardio Respiratory	0.0376 (0.00661)	<.0001
AHRQ SES Index – Hospital-Level – Cardio Respiratory	0.1105 (0.00910)	<.0001
AHRQ SES Index – Patient-Level – Cardiovascular	0.0307 (0.00943)	0.0011
AHRQ SES Index – Hospital-Level – Cardiovascular	0.1375 (0.0149)	<.0001
AHRQ SES Index – Patient-Level – Neurology	0.0544 (0.0125)	<.0001
AHRQ SES Index – Hospital-Level – Neurology	0.1314 (0.0198)	<.0001

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*)

We tested the performance of the model across the development, validation, and 2012 Samples (**Dataset 1**). We computed three summary statistics for assessing model performance (Harrell and Shih, 2001):

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.)

References:

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*)**:**

C-statistics and discrimination in the Development, Validation, and 2012 Sample in **Dataset 1** by specialty cohort:

Table 7. C-statistics and discrimination in the Development, Validation, and 2012 Sample (Dataset 1)

	Hybrid HWR Measure Development Sample	Hybrid HWR Measure Validation Sample	Hybrid HWR Measure 2012 Sample
c-statistics			
Surgery/Gynecology	0.802	0.799	0.800
Cardiorespiratory	0.668	0.673	0.666
Cardiovascular	0.731	0.717	0.726
Neurology	0.708	0.697	0.693
Medicine	0.651	0.656	0.665
Discrimination-Predictive Ability (lowest decile %, highest decile %)			
Surgery/Gynecology	0-35	0-36	0-31
Cardiorespiratory	9-39	7-41	6-36
Cardiovascular	2-29	2-32	2-24
Neurology	4-33	5-37	5-34
Medicine	8-35	7-35	6-34

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

Calibration in the Development, Validation, and 2012 Sample in **Dataset 1** by specialty cohort:

Table 8. Calibration in the Development, Validation, and 2012 Sample (Dataset 1)

	Hybrid HWR Measure Development Sample	Hybrid HWR Measure Validation Sample	Hybrid HWR Measure 2012 Sample
Calibration (γ0, γ1)			
Surgery/Gynecology	(0.000, 1.000)	(-0.049,0.948)	(-0.192,0.971)
Cardiorespiratory	(0.000, 1.000)	(-0.004,0.995)	(-0.111,0.931)
Cardiovascular	(0.000, 1.000)	(0.067,1.007)	(-0.333,0.854)
Neurology	(0.000, 1.000)	(-0.129,0.920)	(-0.464,0.781)
Medicine	(0.000, 1.000)	(-0.047,0.977)	(0.077,1.108)

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

See Calibration Curves in the Excel Data Dictionary attached in data field S.2b.

2b4.9. Results of Risk Stratification Analysis:

N/A This measure is not stratified.

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*)

In **Dataset 1**:

Discrimination Statistics

The range of c-statistics from 0.651 to 0.802 showing good to excellent discrimination across the specialty cohort models.

Calibration Statistics

The calibration value of close to 0 at one end and close to 1 to the other end indicates good calibration of the model. The risk decile plot shows excellent discrimination of the model and good predictive ability.

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 good 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.

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)

The method for discriminating hospital performance has not been determined. For public reporting of measures of hospital outcomes developed with similar methodology, 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 mortality 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 RSSR 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. However, the measure is not currently publicly reported, and decisions about the approach to discriminating hospital performance have not been made.

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*)

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?*)

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 criterion is directed 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).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of 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*)

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*)

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of 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*)

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*)

For the EHR data elements used in the measure's risk models, we anticipate that there will be some missing data. We examined the rates of data capture, and missing data, in the development and testing samples (**Dataset 1**) as well as in the EHR data element feasibility and validity datasets (**Datasets 2** and **3**). We also examined the distribution of the CCDE data values in **Dataset 1** to determine what proportion were out of physiological range and might represent data errors. We found that most values fell within physiological range and that there were few apparent errors in the data entry.

As was shown in 2a.2.3 and 2b.2.3, in all datasets used for testing the rates of capture are above 90% for the data elements included the risk models. Because missing values were rare in the development and testing datasets, it was not necessary to do tests of bias in measure results. However, in order to reduce any small chance of bias due to missing data, we set missing values to the median value in all measure risk models and included a dummy variable whenever a data element was missing in 5% or more of the admissions in each specialty cohort.

To reduce the effect of the spurious outliers, we transformed extreme values by replacing them with a value at the outer limit of a designated range by a process called Winsorization[1,2]. All continuous variables with values less than 1st percentile or higher than the 99th percentile were Winsorized (i.e., values less than the 1st percentile were assigned to the value of the 1st percentile, and values greater than the 99th percentile were assigned to the value of the 99th percentile). Missing data values were set to the median value for the cohort. In addition, dummy variables for missing data were included in the statistical models.

References:

1. Altenburg HP. Estimation of Radioimmunoassay Data Using Robust Nonlinear Regression Methods. In: Dodge Y, Whittaker J, eds. Computational Statistics: Physica-Verlag HD; 1992:367-372.

2. Dixon WJ, Yuen KK. Trimming and winsorization: A review. Statistische Hefte. 1974/06/01 1974;15(2-3):157-170

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*)

We report the capture rate of all EHR data elements in each dataset in section 2a.2.3, and 2b.2.3.

The range of missing data across hospitals in **Dataset 1** can be found in the Data Dictionary attached in data field S.2b

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*)

The rate of missing values was low in all of the datasets and for all hospitals used for testing and therefore not likely to introduce bias. However, we did account for potential outlier values as well as missing values in our risk models to reduce any small possibility of bias. However, approaches to handling missing clinical data in measure calculation will be reassessed during implementation.

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)*

ALL data elements are in defined fields in a combination of electronic sources

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

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.**

Attachment **Attachment:** NQF_2879_Hybrid_HWR_Feasibility_Scorecard_01-29-16_v1.1.docx

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.

Electronic clinical data will be collected from hospitals using MAT output and value sets to inform data queries and electronic reporting requirements.

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)
Public Reporting Not in use	

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

N/A

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?*)

This is a new measure.

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.*)

CMS intends to implement this measure in the Hospital Inpatient Quality Reporting (HIQR) Program once the clinical data elements required for this measure have been reported by hospitals for one year. This measure requires one year of data for calculation. The exact timeline therefore depends on the implementation of a reporting mechanism for these data elements. Once this new measure is implemented, it may replace the claims-only other Hospital-Wide All Cause Unplanned Readmission Measure.

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**

This is a new measure and there is no information available on performance improvement.

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.

In order for CMS to implement this measure in HIQR, there must be a requirement for Inpatient Prospective Payment Systems (IPPS) hospitals to submit the clinical data elements required for measure calculation. This requirement is not yet in place and there is no current timetable for implementation. However, once the core clinical data elements are collected, this hybrid measure may replace the claims-only measure. The hybrid measure has improved credibility and face validity among stakeholders. It also aligns with CMS's goal to incorporate electronic clinical data into quality measures wherever possible.

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 preliminary measure development or model testing. 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)**
0329 : Risk-Adjusted 30-Day All-Cause Readmission Rate
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
0695 : Hospital 30-Day Risk-Standardized Readmission Rates following Percutaneous Coronary Intervention (PCI)
1551 : Hospital-level 30-day all-cause risk-standardized readmission rate (RSRR) following elective primary total hip arthroplasty (THA) and total knee arthroplasty (TKA)
1768 : Plan All-Cause Readmissions (PCR)
1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)
1891 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization

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.**
N/A

5b. Competing Measures
The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
We did not include in our list of related measures any non-outcome measures, such as process measures, with the same target population as our measure. Because this is an outcome measure, clinical coherence of the cohort takes precedence over alignment with related non-outcome measures. Furthermore, non-outcome measures are limited due to broader patient exclusions. This is because they typically only include a specific subset of patients who are eligible for that measure (for example, patients who receive a specific medication or undergo a specific procedure).

The proposed Hybrid HWR measure is a reengineered version of the HWR measure (NQF #1789) in that the proposed measure uses clinical data elements collected from EHR in addition to claims data for risk adjustment. The measure listed above uses only claims data for risk adjustment. In order for CMS to implement this measure in HIQR, there must be a requirement for IPPS hospitals to submit the clinical data elements required for measure calculation. This requirement is not yet in place and there is no current timetable for implementation. However, once the CCDE are collected, this Hybrid measure may replace the claims-only measure.

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:** eHWR_Tech_Report_01-29-16_v1.0.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services (CMS)
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.

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

Ad.6 **Copyright statement:**
Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

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 #: 2880
De.2. **Measure Title:** Excess days in acute care (EDAC) after hospitalization for heart failure
Co.1.1. **Measure Steward:** Centers for Medicare & Medicaid Services (CMS)
De.3. **Brief Description of Measure:** This measure assesses days spent in acute care within 30 days of discharge from an inpatient hospitalization for heart failure to provide a patient-centered assessment of the post-discharge period. This measure is intended to capture the quality of care transitions provided to discharged patients hospitalized with heart failure by collectively measuring a set of adverse acute care outcomes that can occur post-discharge: emergency department (ED) visits, observation stays, and unplanned readmissions at any time during the 30 days post-discharge. In order to aggregate all three events, we measure each in terms of days. In 2016, CMS will begin annual reporting of the measure for patients who are 65 years or older, are enrolled in fee-for-service (FFS) Medicare, and are hospitalized in non-federal hospitals.
1b.1. **Developer Rationale:** The goal of this measure is to improve patient outcomes by providing patients, physicians, and hospitals with information about hospital-level, risk-standardized outcomes following hospitalization for heart failure. Measurement of patient outcomes allows for a broad view of quality of care that cannot be captured entirely by individual process-of-care measures. Safely transitioning patients from hospital to home requires a complex series of tasks which would be cumbersome to capture individually as process measures: timely and effective communication between providers, prevention of and response to complications, patient education about post-discharge care and self-management, timely follow-up, and more. Suboptimal transitions contribute to a variety of adverse events post-discharge, including ED evaluation, need for observation, and readmission. Measures of unplanned readmission already exist, but there are no current measures for ED and observation stay utilization. It is thus difficult for providers and consumers to gain a complete picture of post-discharge outcomes. Moreover, separately reporting each of these outcomes encourages "gaming," such as re-categorizing readmission stays as observation stays to avoid a readmission outcome. By capturing a range of acute care events that are important to patients, we can produce a more complete picture of post-discharge outcomes that better informs consumers about care quality and incentivizes global improvement in transitional care.

S.4. **Numerator Statement:** The outcome of the measure is a count of the number of days the patient spends in acute care within 30 days of discharge. We define days in acute care as days spent in an ED, admitted to an observation unit, or admitted as an unplanned readmission for any cause within 30 days from the date of discharge from the index heart failure hospitalization. Each ED treat-and-release visit is counted as one half-day (0.5 days). Observation stays are recorded in terms of hours and are rounded up to the nearest half-day. Each readmission day is counted as one full-day (1 day). We count all eligible outcomes occurring in the 30-day period, even if they are repeat occurrences.
S.7. **Denominator Statement:** The target population for this measure is Medicare FFS beneficiaries aged 65 years and older hospitalized at non-Federal acute care hospitals for heart failure.

The cohort includes admissions for patients discharged from the hospital with a principal discharge diagnosis of heart failure (see codes below in S.9) and with continuous 12 months Medicare enrollment prior to admission. The

measure will be publicly reported by CMS for those patients 65 years and older who are Medicare FFS beneficiaries admitted to non-federal hospitals.

Additional details are provided in S.9 Denominator Details.
S.10. **Denominator Exclusions:** The measure excludes index admissions for patients:
1. Without at least 30 days post-discharge enrollment in FFS Medicare.
2. Discharged against medical advice (AMA);
3. Admitted within 30 days of a prior index discharge.

For 2016 public reporting, the measure will also exclude:
4. Admissions with a procedure code for left ventricular assist device (LVAD) implantation or heart transplantation either during the index admission or in the 12 months prior to the index admission. Patients with these procedures are a highly selected group of patients with different risk of the outcome. This exclusion will be added to the heart failure EDAC measure so that it remains fully harmonized with the CMS 30-day heart failure readmission measure. We did not exclude patients with LVAD or heart transplantation from the cohort of admissions used in the analyses for measure development and testing presented here.

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Administrative claims
S.26. **Level of Analysis:** Facility

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

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 any measure; however, it is harmonized with a measure of hospital-level, all-cause, 30-day, risk-standardized readmission following heart failure hospitalization.

New Measure -- Preliminary Analysis

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.

Summary of evidence provided by the developer:

- This is a new health outcome measure and the level of analysis is facility.
- This measure calculates excess days in acute care (EDAC) for patients with heart failure.
- As a rationale for measuring this health outcome, the developer suggests that hospitals are able to influence readmission rates through a broad range of clinical activities including communication between providers, prevention of, and response to, complications, patient safety and coordinated transitions to the outpatient environment.
- The developer cites that "in the context of the Centers for Medicare and Medicaid Services' (CMS's) publicly

reported readmission measures, the increasing use of ED visits and observation stays has raised concerns that current readmission measures do not capture the full range of unplanned acute care in the post-discharge period (Vashi et al., 2013; Rising et al., 2012; Feng et al., 2012). Observation stays can occur in many different parts of the hospital, including dedicated treatment rooms, the ED, or inpatient units. In particular, there is concern that high use of observation stays could in some cases replace readmissions, and that hospitals with high rates of observation stays in the post-discharge period may therefore have low readmission rates that do not accurately reflect the quality of care (Vashi et al., 2013)."

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

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 developer provides performance data from one measurement period from 2010-2013, covering a total of 575,672 discharges.
- The data show that during the measurement period of 2010-2013, heart failure readmission rates ranged from a minimum of -67.02% to a maximum of 196.31%, with the 10th percentile at -29%, the 50th percentile at 3.62%, and the 90th percentile at 44.4%.

Disparities

- To help in assessment of potential disparities, the developers also provide performance scores for hospitals serving a low proportion of dual eligible patients vs. those serving a high proportion of dual eligible patients and performance scores for hospitals serving a low proportion of African-American patients vs. those serving a high proportion of African-American patients.

- By proportion of **Dual Eligible Patients:**

  // Low proportion (=7.69%) dual-eligible patients)//Hospitals with a high proportion (= 23.08% dual-eligible patients)
  Number of Measured Hospitals//1,137//1,139
  Number of Patients//131,204 patients in low-proportion hospitals/ 88,954 in high-proportion hospitals
  Maximum//223.58 //140.48
  90th percentile//37.54 //55.54
  75th percentile//15.46 //28.50
  Median (50th percentile)//-1.95 //2.03
  25th percentile//-18.36//-14.85
  10th percentile//-30.28 //-28.23
  Minimum //-63.47 //-60.35

- By proportion of **African-American Patients:**

  //Low proportion (=0.0%) African-American patients//Hospitals with a high proportion (=11.56%) African-American patients
  Number of Measured Hospitals//1,851//1,155
  Number of Patients//72590 patients in low-proportion hospitals/199,085 in high-proportion hospitals

Maximum//223.58 //144.04
90th percentile//32.52 //51.70
75th percentile//7.98 //31.35
Median (50th percentile)//-7.15// 9.45
25th percentile//-22.81//-7.99
10th percentile//-34.45//-22.63
Minimum//-69.99//-54.39

- The developer explains that: "low-proportion hospitals are those hospitals whose population of dual-eligible patients or of African-American patients is small enough to place them at or below the 25th percentile among all hospitals; and high proportion are those hospitals whose population of dual eligible patients or African-American patients is large enough to place them at or above the 75th percentile among all hospitals."

Questions for the Committee:

- *Specific question on information provided for gap in care.*
- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

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)

1. Importance to Measure and Report

1a. Evidence to Support Measure Focus

Comments: **Relationship between outcome and healthcare action exists; many interventions have been shown to reduce readmissions (and presumably by proxy this outcome) in heart failure patients.

**Evidence to support readmission measures already accepted. This measure addresses concern that high use of observation stays could in some cases replace readmissions, and that hospitals with high rates of observation stays in the post-discharge period may therefore have low readmission rates that do not accurately reflect the quality of care.

1b. Performance Gap

Comments: **I don't think the provided information on this measure worksheet is correct - it reports readmission rates of negative 67%. Not sure if these are supposed to be the betas associated with each characteristic, though that doesn't square with what's in the measure sheet from the developer.

**Measure developers report significant performance gap:

minimum of -67.0 to a maximum of 196.31, with the 10th percentile at -29, the 50th percentile at 3.62, and the 90th percentile at 44.4

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

Comments: **Yes

**NA

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

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): Administrative claims

- This measure calculates the number of days spent in acute care within 30 days of discharge from an inpatient hospitalization for heart failure to provide a patient-centered assessment of the post-discharge period.
- The outcome of the measure is a count of the number of days the patient spends in acute care within 30 days of discharge. The measure defines days in acute care as days spent in an ED, admitted to an observation unit, or admitted as an unplanned readmission for any cause within 30 days from the date of discharge from the index heart failure hospitalization. Each ED treat-and-release visit is counted as one half-day (0.5 days). Observation stays are recorded in terms of hours and are rounded up to the nearest half-day. Each readmission day is counted as one full-day (1 day). The measure counts all eligible outcomes occurring in the 30-day period, even if they are repeat occurrences.
- The Numerator is the count of the number of days the patient spends in acute care within 30 days of discharge. We define days in acute care as days spent in an ED, admitted to an observation unit, or admitted as an unplanned readmission for any cause within 30 days from the date of discharge from the index heart failure hospitalization.
- The Denominator is the Medicare FFS beneficiaries aged 65 years and older hospitalized at non-Federal acute care hospitals for heart failure.
- The denominator population is defined using ICD-9 and ICD-10 codes; a list of applicable codes is included in the submission.
- The data sources for this measure are Medicare Part A inpatient, Part B hospital outpatient claims and physician Carrier claims, and the Medicare Enrollment Database (EDB).
- The measure's time window is three years.
- The measure is risk-adjusted using a statistical risk model (see details below).

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

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**

- The developer has assessed reliability at both the data element and the performance score levels.
- The datasets used for testing included Medicare Parts A and B claims, the Medicare Enrollment Database (EDB), and the Chronic Condition Data Warehouse (CCW) 100% condition-specific dataset to capture emergency department (ED) visits and observation stays. Additionally, census data were used to assess socio-demographic factors.
    - **Data element reliability**:

- With regard to data element reliability, the developer notes that the measure has been developed to avoid the use of claims data elements that are thought to be coded inconsistently across hospitals or providers, instead using fields that are consequential for payment and which are audited by CMS. Additionally, the developer used the final risk-adjustment variables in the existing, NQF-endorsed measure of hospital-level risk-standardized readmission rates following AMI (NQF #0505).
- Additionally, the developer compared variable frequencies between the development and validation samples.

- **Performance score reliability**:
  - The developer defines performance score reliability as the degree to which repeated measurements of the same entity agree with each other.
  - In line with this thinking, the developer's approach to assessing score-level reliability was to consider the extent to which assessments of a hospital using different but randomly-selected subsets of patients produce similar measures of hospital performance. The developers refer to this as a "test-retest" approach; it may also be called a "split-half" method.
  - For test-retest reliability, the developer calculated the EDAC for each hospital using first the development sample, then the validation sample. Thus, each hospital twice was measured twice, each time using an entirely distinct set of patients. The developer states that the extent to which the calculated measures of these two subsets agree is evidence that the measure is assessing an attribute of the hospital, not of the patients. As a metric of agreement, the developer calculated the intra-class correlation coefficient (ICC) as defined by ICC[2,1] by Shrout and Fleiss (1979) and assessed the values according to conventional standards (Landis and Koch, 1977).
  - A total of 1,180,895 admissions were examined, with 590,448 in one sample and 590,447 in the other.
  - The agreement between the two EDAC values for each hospital (as measured by an intra-class correlation coefficient (ICC)) was **0.73**; the developer states that according to the conventional interpretation, this is considered a "substantial" level of agreement.
  - The developer notes that this analysis was limited to hospitals with to hospitals with at least 12 discharges in both samples to approximate the set of hospitals that would have at least 24 discharges over three years and are thus likely to be included in public reporting. [*Note: It is unclear whether the measure itself is limited to hospitals with 12 or more cases and if three years of data are needed to calculate the measure; if it is not, then testing was not conducted with the measure as specified*.]
  - The developer expects that the correlation coefficient would be higher using a full three-year sample since it would include more patients. To correct this problem, the developer used the Spearman-Brown prophecy formula (Spearman 1910, Brown 1910) to adjust the ICC[2,1] to represent three years of data.

- The developer's overall interpretation of reliability testing results is that the compared to the development sample, the mean age of patients and the frequencies of the risk-adjustment variables were very similar in the validation sample; this indicates that the data elements are reliable and that the ICC score from performance score analysis demonstrates moderate agreement across samples. The developer notes that the ICC [2,1] score of 0.54, estimated for three years of data, demonstrates moderate agreement between samples across the full range of measure values. We interpret this to mean that when used with a full three years of data, the measure will be reliable by the standards of hospital measurement.

Guidance from the Reliability Algorithm

- Question 1. Submitted specifications are precise, unambiguous, and complete. Measure can be consistently

implemented.

- Question 2. Empirical reliability testing was conducted using statistical tests with the measure as specified.
- Question 3. Empirical validity testing of patient-level data was conducted.
- Question 4. Reliability testing was conducted with computed performance measure scores for each measured entity.
- Question 5. Random split-half correlation was used to assess the proportion of variability due to real differences among the measured entities.
- Question 6. The ICC was 0.73 which is considered a substantial level of agreement.

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?*
- *Does the measure need three years of data to achieve this level of reliability?*

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

2b. Validity

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**

- This measure calculates the number of days spent in acute care within 30 days of discharge from an inpatient hospitalization for heart failure to provide a patient-centered assessment of the post-discharge period.
- As a rationale for measuring this health outcome, the developer suggests that hospitals are able to influence readmission rates through a broad range of clinical activities including communication between providers, prevention of, and response to, complications, patient safety and coordinated transitions to the outpatient environment.

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.

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 demonstrated measure validity through prior validity testing done on their claims-based measures, through use of established measure development guidelines, and by systematic assessment of measure face validity by a Technical Expert Panel (TEP).
    - Empirical Validity Testing
        - The developer notes this measure is closely related in design to the existing, NQF-endorsed readmission measure for patients with HF. While this measure includes additional endpoints and measures them in a different metric (days rather than rates), the developer expects that hospitals would have similar – though not identical – performance rankings on the two measures. Therefore as one assessment of validity, they compared the rankings of all hospitals using the two measures to assess the consistency of hospital performance on closely related outcomes. The developer calculated the Pearson correlation, and graphed the readmission measure against the EDAC measure to determine if there were outliers.
        - Comparison of the new measure with the existing CMS 30-day HF readmission measure found a Pearson correlation of 0.714 ($P < 0.0001$).
    - Validity of Claims-Based Measures:
        - The developer states that they have demonstrated for a number of other readmission measures the validity of claims-based measures by comparing either the measure result or the individual data elements against medical records.
        - Claims model validation was conducted by building comparable models using abstracted medical chart data for risk adjustment. 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.
    - Validity Indicated by Established Measure Development Guidelines
        - The developer states that this measure was developed in consultation with national guidelines for publicly reported outcomes measures, with outside experts, and with the public.
    - Validity as Assessed by External Groups:
        - Input was obtained through regular discussions with an advisory working group, a TEP, and a 30-day public comment period.
    - Face Validity as Determined by TEP:
        - The developer asked members of the TEP to note their agreement with the statement "The risk-standardized acute care days obtained from the measures as specified can be used to distinguish between better and worse quality hospitals."
        - Of the TEP members who responded, 91.7% agreed (83.3% moderately or strongly agreed) that the measure will provide an accurate reflection of quality.
        - The developer interpreted this as a moderate level of agreement.

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?*
- *Is the claims model validation sufficiently similar to the measure?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- Patients in the following categories are excluded from the measure:
  - Discharged patients without at least 30 days post-discharge information
  - Discharges against medical advice (AMA)
  - Admissions within 30 days of a prior index admission
- The developer notes that all exclusions were determined by careful clinical review and have been made based on clinically relevant decisions and to ensure accurate calculation of the measure
- To determine the impact of exclusions, the developer examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion.
- The number and percentage of patients excluded for each criterion are as follows:
  - Without at least 30 days post-discharge enrollment in FFS Medicare for index admissions: **6,107 (0.47%)**
  - Discharged against medical advice (AMA): **5,092 (.39%)**
  - Admissions within 30 days of a prior index admission: **104,470 (8.06%)**
- The developer also provides the distribution across hospitals for each exclusion criterion.
- The developer notes that the first exclusion criterion, is needed since the outcome cannot be assess in this group since claims data are used to determine whether a patient returned to the hospital for an ED visit, was placed under observation care, or was readmitted.
- The developer states that the second exclusion criterion is needed for acceptability of the measure to hospitals, who do not have the opportunity to adequately deliver full care and prepare the patient for discharge.
- The developer notes that exclusion criterion 3 is needed to prevent admissions from being counted as both an index admission and a readmission, consistent with the approach taken in the HF readmission measure.

Questions for the Committee:

- *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: **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

- For this measure the develop adopted the risk factors from the existing NQF-endorsed CMS 30-day AMI readmission measure. These risk factors are comprised of age, sex, and condition categories (CCs) for prior 12-month and current claims.
- The developer notes these risk factors had been systematically chosen as predictors of any readmission for the same patient cohort as the current measure; the outcome of this measure is dominated by the number of days of a readmission, so they judged it unlikely that repeating the original analysis would produce different results.
- The developer confirmed that there were no additional risk factors to consider by comparing the model estimated using the a priori set of risk factors to a model which included all additional CCs.
- The measure employs a hierarchical generalized linear model [HGLM]) that consists of two parts, a logit model

and a truncated Poisson model. The two-part logit/Poisson model (often called a "hurdle" model) assumes that the outcome results from two related processes: an initial dichotomous event – that a patient has at least one acute care event – which is modeled as the logit of the probability of the event, and for patients with an event (those which clear the "hurdle"), the number of days, which is modeled as a Poisson process. The outcome, number of days, is a half-integer count variable (because ED visits count as 0.5 days).

- There are two random effects for each hospital, one for the logit model and one for the truncated Poisson model, as well as a covariance between the two random effects. The developer suggests that the random effects allows for within-hospital correlation of the observed outcome and accommodates the assumption that underlying differences in quality across hospitals lead to systematic differences in outcomes.
- The final set of 37 risk-adjustment variables is included in the testing attachment; the odds ratio associated with each variable is also provided.
- The developers also considered a number of variables related to sociodemographic status (SDS) for potential inclusion in the risk-adjustment model. Candidate SDS variables were selected for examination based on a review of literature and national data sources.
- **Conceptual analysis of the need for SDS adjustment:**
    - The developers note there is a large body of literature linking various SES factors and African-American race to worse health status and higher readmission risk with income, education, and occupational level being the most commonly examined variables. The developers state that the literature directly examining how SES factors or race might influence the likelihood of older, insured, Medicare patient of being readmitted within 30 days of an admission for heart failure is more limited.
    - The developers state that few studies directly address causal pathways for SDS factors to affect 30-day readmission rates or examine the role of the hospital in these pathways.
    - There are at least four potential pathways for SDS factors to affect 30-day readmission rates:
        - One potential pathway is the relationship to health status at the time of admission. SDS factors 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. The developers note that this pathway should be largely accounted for by their clinical risk-adjustment model.
        - The next potential path way is that patients with low income and African-American patient are more likely to be seen in lower quality hospitals, which can contribute to increased risk of readmission.
        - The third major pathway is that a patient's race or SDS status cause them to experience differential, lower quality care or may not receive the differentiated care they require.
        - Finally, some SES risk factors may affect the likelihood of readmission without directly affecting health status at admission or the quality of care received during the hospitalization. Patients may have worse outcomes due to competing economic priorities or a lack of access to care outside the hospital.
- **Empirical analysis of SDS factors:**
    - The developers considered African-American race, and dual-eligible status-i.e. enrolled in both Medicare and Medicaid. The developers assessed the relationship between the SES variables and race with the outcome and examined the incremental effect in a multivariable mode.
    - The developer assessed the relationship between the SDS variables and the days in acute care and

examined the incremental effect of SDS in a multivariable model, evaluating the extent to which the addition of any one of these variables improved model performance or changed hospital results.
    - The developer notes that 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. Therefore, the developers performed a decomposition analysis to assess the independent effects of the SES and race variables at the patient level and the hospital level.
    - The developers' analysis found that the prevalence of SDS factors in the HF cohort does vary across measured entities.
    - With regard to the empirical association of each SDS variable with the outcome (univariate), the analysis found that patient-level observed days in acute care for dual-eligible patients was higher, at 172.70 per 100 discharges compared with 142.84 days in acute care per 100 discharges for all other patients. The mean observed days in acute care for African-American patients was also higher at 174.06 days per 100 discharges compared with 143.52 days per 100 discharges for patients of all other races
    - With regard to the strength and significance of the SDS variables in the context of a multivariable model, the developers' analysis found that the effect size of each of these variables is small. The developers also found that the c-statistics (i.e., predictive value) for the logit part of the model and the deviance R2 values for the Poisson part of the model are similar with and without the addition of either of these variables into the model. Additionally the developers found the addition of these variables has little to no effect on hospital performance.
        - The median absolute change in hospitals' EDAC when adding a dual-eligibility indicator is 0.43 EDAC per 100 discharges (interquartile range [IQR] 0.20-0.75; minimum 0.00-maximum 7.16), with a Spearman correlation coefficient between EDAC for each hospital with and without dual eligibility added of 0.9996.
        - The median absolute change in hospitals' EDAC when adding a race indicator is 0.42 EDAC per 100 discharges (IQR 0.19-0.78; minimum 0.00-maximum 7.91), with a Spearman correlation coefficient between EDAC for each hospital with and without race added of 0.9958.
    - The developers state that both the patient-level and hospital-level dual eligible and race effects were significant in the logistic part of the HF EDAC model, but only the hospital-level effect was significant in the Poisson part of the model. This indicates that a) both the patient- and hospital-level dual eligible and race effects are associated with an increased risk of acute care but b) only the hospital-level effect is associated with the expected duration of that care. The developers note that if the dual eligible or race are 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 developers state that given these findings and complex pathways that could explain any relationship between SDS and readmission, they did not incorporate SDS variables into the measure.
- **Risk Model Diagnostics:**
    - To assess model discrimination the developers computed two different statistics: one for the logit part of the model and one for the Poisson part.
        - For the logit model of zero versus non-zero days, which includes all patients in the cohort, the developers calculated the c-statistic.
            - C-statistic for logit part of model: 0.587
        - For the Poisson model of non-zero days, which includes only patients with some acute care, the

developers calculated the deviance R2. The deviance R2 is computed from the difference in the log-likelihoods between the final model and an empty model (no covariates) attributed to each observation, averaged over all observations.
        - Deviance R2 for truncated Poisson part of model: 0.026 (2.6%)
    - The developers interpret these results as good model calibration.
- In a generalization of the calibration statistics for logistic models, the developers calculated the linear prediction Z = XB and W = XC using the coefficients B and C from the development sample and data X from the validation sample. The developers then estimated a model using the same functional form but only two independent variables, Z for the truncated Poisson part and W for the logit part. The intercepts and coefficients of Z and W in these second models are reported as (γ_0, γ_1), , the calibration statistics for each part of the model. The closer they are to (0, 1), the better the model calibration
    - Calibration Statistics (y0, y1):
        - Logit part of model: (0.03, 1.00)
        - Poisson part of model: (-0.06, 0.97)

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 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)*:

- To categorize hospital performance, the developers estimated each hospital's EDAC and the corresponding 95% credible interval (CI).
- The developers then assigned hospitals to a performance category by comparing each hospital's EDAC interval estimate to zero. Comparative performance for hospitals with 25 or more eligible cases was classified as follows:
    - "Lower than expected" if the entire 95% CI surrounding the hospital's days is below zero.
    - "No different than expected" if the 95% CI surrounding the hospital's days includes zero.
    - "Higher than expected" if the entire 95% CI surrounding the hospital's days is above zero.
- Hospitals with fewer than 25 eligible cases were assigned to a separate category: "The number of cases is too small (fewer than 25) to reliably assess the hospital's EDAC."
- Of 4,654 hospitals in the study cohort (data from July 1, 2010 through June 30, 2013), 532 had EDACs "lower than expected," 2,501 were "no different than expected," and 915 had EDACs "higher than expected." 706 were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing. The mean EDAC per 100 discharges for hospitals in the top decile of performance is -29.0, compared to 196.3 for hospitals in the bottom decile.
- The developer states that the variation in hospital-level EDAC suggests there are meaningful differences in the quality of care received across hospitals for the HF EDAC measure.

Question for the Committee:

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

2b6. Comparability of data sources/methods:

NA

2b7. Missing Data

NA

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)

2. Scientific Acceptability of Measure Properties

2a1. & 2b1. Specifications

Comments: **None

**To evaluate test-retest reliability, the developer calculated the EDAC for each hospital using first the development sample, then the validation sample:

1,180,895 admissions were examined, with 590,448 in one sample and 590,447 in the other. ICCC= 0.73 == substantial agreement

2a2. Reliability Testing

Comments: **Yes

**Yes.

The datasets used for testing included Medicare Parts A and B claims, the Medicare Enrollment Database (EDB), and the Chronic Condition Data Warehouse (CCW) 100% condition-specific dataset to capture emergency department (ED) visits and observation stays.

2b2. Validity Testing

Comments: **Yes, adequate testing was completed, assuming the three-year issue (as noted in the worksheet) was appropriately communicated.

**Data elements are the same as those used in the existing, NQF-endorsed measure of hospital-level risk-standardized readmission rates following AMI (NQF #0505). Prior efforts have established that the publicly reported CMS 30-day heart failure readmission measure risk model variables derived from claims data are consistent with those based on medical chart review.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **No

**The amount of missing data is not a threat to the validity of the measure:

- Without at least 30 days post-discharge enrollment in FFS Medicare for index admissions: 6,107 (0.47%)
- Discharged against medical advice (AMA): 5,092 (.39%)
- Admissions within 30 days of a prior index admission: 104,470 (8.06%)
  - needed to prevent admissions from being counted as both an index admission and a readmission, consistent with the approach taken in the HF readmission measure

Criterion 3. Feasibility

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 based on administrative claims data (e.g., DRG, ICD-9/10), which the developers note are routinely generated and collected as part of hospitals' billing processes.
- The developer indicates that all data elements are in defined fields in electronic claims.

Questions for the Committee:

- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3. Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **Measure is feasible

**Highly feasible because based on widely available administrative data.

Criterion 4: Usability and Use

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**

OR

Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

- This measure has been finalized for use in CMS's Hospital Inpatient Quality Reporting (IQR) program starting in Fiscal Year (FY) 2018 (80 FR 49690).

Improvement results

- Since this measure is not in use, there are no performance results to assess improvement at this time.
- The developer states that they expect that "there will be improvement in measure scores over time since publicly reported measure scores can reduce adverse patient outcomes associated with days spent in acute care for heart failure by capturing and making acute care utilization following the index hospitalization more visible to providers and patients."

Potential harms

- The developer noted that there were no unintended consequences during development, testing or re-specification. They are committed to ongoing monitoring of potential unintended consequences, such as the inappropriate shifting of care, increased patient morbidity and mortality, and other negative intended consequences over time.

Feedback :

- MAP reviewed this measure during its 2014-2015 pre-rulemaking deliberations for consideration in the IQR program. MAP was conditionally supportive of this measure on the condition that this measure is reviewed by NQF and endorsed. In particular, members noted that the measure should be considered for SDS adjustment in the upcoming NQF trial period, reviewed for the empirical and conceptual relationship between SDS factors and risk-standardized days following acute care, and endorsed with appropriate consideration of SDS factors as determined by NQF standing committees. Some MAP members noted this measure could help address concerns about the growing use of observation stays.

***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

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **This measure identifies many more outliers than the current RSRR (at least for public reporting), which is a good thing if you believe that public reporting works (though the evidence suggests it does not). Unclear whether that approach is of any salience given that the RSRR is used for the HRRP without any consideration of statistical difference from the mean. Also - adding here since no box available for 2b4 response: risk adjustment is a concern given low c-statistic and reliance on same risk adjusters as prior measures. However, I am not qualified to comment on the "hurdle" approach or the additive value of the deviance R2, but it would be worth a review by someone who is. As for SES issue: dual is clearly a risk factor - if the developers wish not to adjust for it for philosophical reasons it should be stated as such.

**Not able to evaluate usability and use since not currently in use.

Criterion 5: Related and Competing Measures

Related or competing measures

- 0330: Hospital 30-day All-Cause RSRR Following Heart Failure Hospitalization

Harmonization

- The developers note that both measures are harmonized.

***Questions for the Committee**:*

- *Are #2880 and #0330 sufficiently different and harmonized to endorse both?*

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: Excess days in acute care (EDAC) after hospitalization for heart failure

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: 1/29/2016

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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: Single measure: quality outcome measure

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

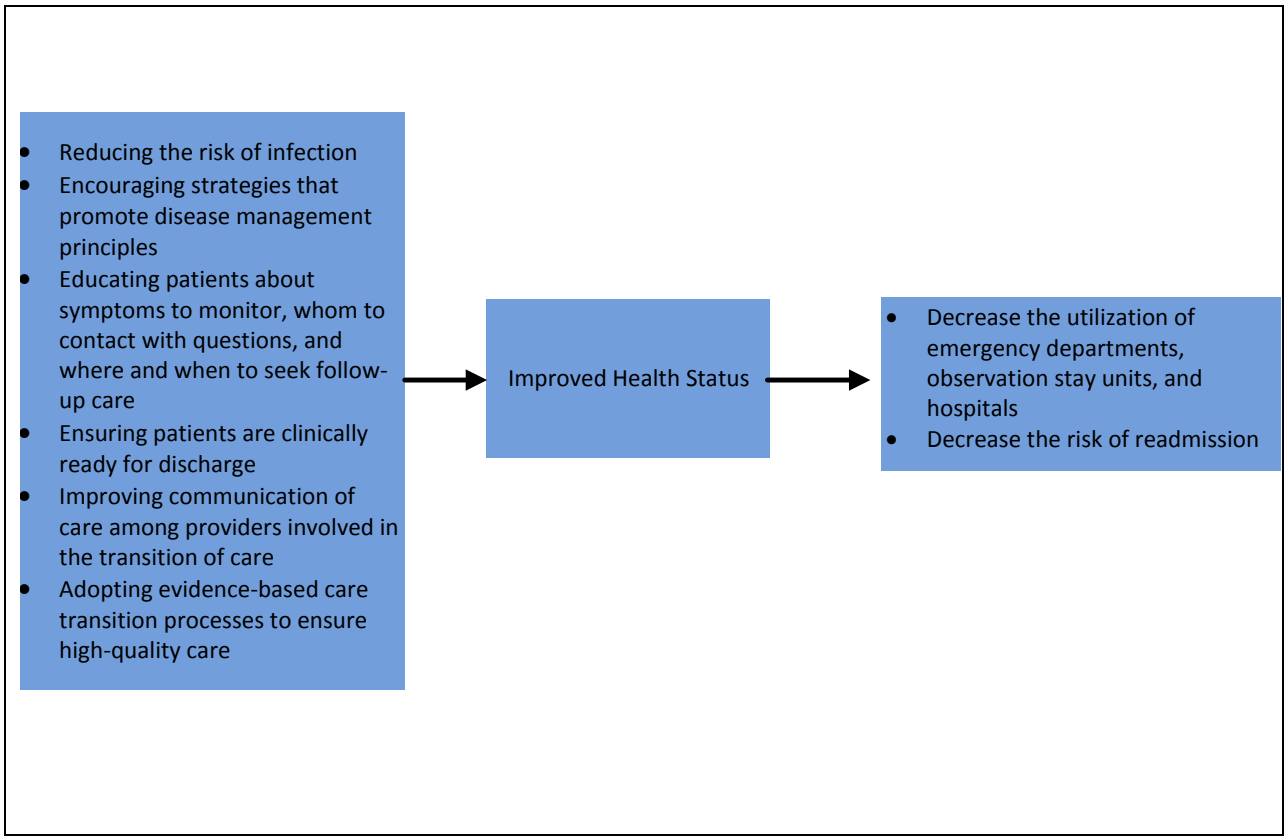
☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.



1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

The diagram above indicates some of the many care processes that can influence post-discharge acute care utilization. These complex and critical aspects of care – such as communication between providers, patient education, patient safety, and coordinated transitions to the outpatient environment – all contribute to patient outcomes but are difficult to measure by individual process measures. Numerous studies have shown that improvement in the following areas can favorably impact utilization rates: communication with patients, caregivers, and other providers; patient education; and quality of care during the initial inpatient admission.

Interventions during and after a hospitalization can be effective in reducing readmission rates in geriatric populations (Benbassat et al., 2000; Naylor et al., 1999; Coleman et al., 2006; Courtney et al., 2009; Koehler et al., 2009) and, particularly, for older patients with heart failure (Phillips

et al., 2004; Naylor et al., 2004; Koelling et al., 2005; Krumholz et al., 2002; Nohria et al, 2002). Several randomized trials have reduced 30-day readmission rates by 20-40% (Jack et al., 2009; Coleman et al., 2004; Courtney et al., 2009; Garasen et al., 2007; Koehler et al., 2009; Mistiaen et al., 2007; Naylor et al., 1994; Naylor et al., 1999; van Walraven et al., 2002; Weiss et al., 2010; Krumholz et al., 2012; Balaban et al., 2008). These types of interventions have also been demonstrated to be cost-saving (Naylor et al., 1999; Naylor et al., 2004; Koelling, 2005; Krumholz et al., 2002; Stauffer et al., 2011). Outside the randomized controlled trial setting, there is also increasing evidence that hospitals and health plans have been able to reduce readmission rates through more generalizable quality improvement initiatives (Gerhardt et al., 2012; Stauffer et al., 2011; Graham et al., 2012; Harrison et al., 2011; Hernandez et al., 2010).

Improvements in transitional care do not only benefit readmissions. Some studies of patients with heart failure involving patient education, telephone monitoring, and improved communication among providers and patients have reduced post-discharge emergency department (ED) utilization (Domingues et al., 2011; Harrison et al., 2002). Studies also have reported significant reductions in ED visit rates in patients with other conditions after implementation of interventions that focused on the inpatient and outpatient settings (Bondestam et al., 1995).

The current process-based performance measures cannot capture all the ways that care within the hospital might influence outcomes. As a result, many stakeholders, including patient organizations, are interested in outcomes measures that allow patients and providers to assess relative outcomes performance among hospitals (Bratzler et al., 2007).

In the context of the CMS's publicly reported readmission measures, the increasing use of ED visits and observation stays has raised concerns that current readmission measures do not capture the full range of unplanned acute care in the post-discharge period (Vashi et al., 2013; Rising et al., 2012; Feng et al., 2012). Observation stays can occur in many different parts of the hospital, including dedicated treatment rooms, the ED, or inpatient units. In particular, there is concern that high use of observation stays could in some cases replace readmissions, and that hospitals with high rates of observation stays in the post-discharge period may therefore have low readmission rates that do not accurately reflect the quality of care (Carlson et al., 2013).

Citations:

Balaban RB, Weissman JS, Samuel PA, Woolhandler S. Redefining and redesigning hospital discharge to

enhance patient care: a randomized controlled study. *J Gen Intern Med.* 2008;23(8):1228-1233.

Benbassat, J., and M. Taragin. Hospital readmissions as a measure of quality of health care:

advantages and limitations. *Arch Intern Med.* 2000;160 (8):1074-81.

Bondestam E, Breikss A, Hartford M. Effects of early rehabilitation on consumption of medical care

during the first year after acute myocardial infarction in patients > or = 65 years of age. *American*

Journal of Cardiology. 1995;75(12):767-771.

Bratzler, DW, Nsa W, Houck PM. Performance measures for pneumonia: are they valuable, and are

process measures adequate. *Current Opinion in Infectious Diseases*. 2007;20(2):182-189.

Carlson J. Faulty Gauge? Readmissions are down, but observational-status patients are up and that could

skew Medicare numbers. *Modern healthcare*. 2013.

Coleman EA, Parry C, Chalmers S, et al. The care transitions intervention: results of a randomized

controlled trial. *Arch Intern Med.* 2006;166:1822-1828.

Coleman EA, Smith JD, Frank JC, Min SJ, Parry C, Kramer AM. Preparing patients and caregivers to

participate in care delivered across settings: the Care Transitions Intervention. *J Am Geriatr Soc*

2004;52(11):1817-25.

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*

2009;57(3):395-402.

Domingues FB, Clausell N, Aliti GB, Dominguez DR, Rabelo ER. Education and telephone monitoring by

nurses of patients with heart failure: randomized clinical trial. *Arquivos Brasileiros de Cardiologia*.

2011;96(3):233-239.

Feng Z, Wright B, Mor V. Sharp rise in Medicare enrollees being held in hospitals for observation raises

concerns about causes and consequences. *Health affairs (Project Hope)*. Jun 2012;31(6):1251-1259.

Fisher, ES, JE Wennberg, TA Stukel, et al. Hospital readmission rates for cohorts of Medicare

beneficiaries in Boston and New Haven. *N Engl J Med.* 1994;331(15):989-95.

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.

Gerhardt G, Yemane A, Hickman P, Oelschlaeger A, Rollins E, Brennan N. Medicare Readmission Rates Showed Meaningful Decline in 2012. *Medicare & Medicaid Research Review.* 2013;3(2):E1-E12.

Graham J, Tomcavage J, Salek D, Sciandra J, Davis DE, Stewart WF. Postdischarge monitoring using interactive voice response system reduces 30-day readmission rates in a case-managed Medicare population. *Medical Care.* 2012;50(1):50-57.

Harrison MB, Browne GB, Roberts J, Tugwell P, Gafni A, Graham ID. Quality of life of individuals with

heart failure: a randomized trial of the effectiveness of two models of hospital-to-home transition.

Medical Care. 2002;40(4):271-282.

Harrison PL, Hara PA, Pope JE, Young MC, Rula EY. The impact of post-discharge telephonic follow-up on hospital readmissions. *Population Health Management.* 2011;14(1):27-32.

Hernandez AF, Greiner MA, Fonarow GC, et al. Relationship between early physician follow-up and 30-day readmission among Medicare beneficiaries hospitalized for heart failure. *JAMA.* 2010;303(17):1716-1722.

Jack BW, Chetty VK, Anthony D, Greenwald JL, Sanchez GM, Johnson AE, et al. A reengineered hospital

discharge program to decrease rehospitalization: a randomized trial. Ann Intern Med 2009;150(3):178

87.

Koehler BE, Richter KM, Youngblood L, Cohen BA, Prengler ID, Cheng D, 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. *J Hosp Med*. 2009;4(4):211-218.

Koelling, TM. 2005. Multifaceted outpatient support can improve outcomes for people with heart

failure. Commentary. Evid Based Cardiovasc Med. 2005;9 (2):138-41.

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*. 2002; 39 (1):83-9.

Mistiaen P, Francke AL, Poot E. Interventions aimed at reducing problems in adult patients discharged

from hospital to home: a systematic metareview. BMC Health Serv Res 2007;7:47.

Naylor, MD, Brooten D, Campbell R, et al. Comprehensive discharge planning and home follow-up

of hospitalized elders: a randomized clinical trial. *JAMA*. 1999; 281(7):613-20.

Naylor, MD, Brooten D, Campbell R, et al. 2004. Transitional care of older adults hospitalized with heart

failure: a randomized, controlled trial. *J Am Geriatr Soc*. 2004; 52 (5):675-84.

Naylor M, Brooten D, Jones R, Lavizzo-Mourey R, Mezey M, Pauly M. 1994. Comprehensive discharge

planning for the hospitalized elderly. A randomized clinical trial. *Ann Intern Med*. 1994. 120(12):999-1006.

Nohria A, Lewis E, and Stevenson LW. 2002. Medical management of advanced heart failure. *JAMA*. 2002; 287(5):628-40.

Rising, Kristin L., et al. Emergency department visits after hospital discharge: a missing part of the equation. *Annals of emergency medicine*. 2013;62(2):145-150.

Seow H, Phillips CO, Rich MW, Spertus JA, Krumholz HM, Lynn J. Isolation of health services research

from practice and policy: the example of chronic heart failure management. *Journal of the American*

Geriatrics Society. Mar 2006;54(3):535-540.

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.* 2011;171(14):1238-1243.

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.* 2002;17(3):186-92.

Vashi AA, Fox JP, Carr BG, et al. Use of hospital-based acute care among patients recently discharged

from the hospital. *JAMA*. Jan 23 2013;309(4):364-371.

Weiss M, Yakusheva O, Bobay K. Nurse and patient perceptions of discharge readiness in relation to

Post-discharge utilization. *Med Care*. 2010;48(5):482-6.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

N/A. This is an outcome measure.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

N/A. This is an outcome measure.

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

N/A. This is an outcome measure.

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

N/A. This is an outcome measure.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A. This is an outcome measure.

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

N/A. This is an outcome measure.

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

N/A. This is an outcome measure.

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

N/A. This is an outcome measure.

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

N/A. This is an outcome measure.

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

N/A. This is an outcome measure.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A. This is an outcome measure.

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

N/A. This is an outcome measure.

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

N/A. This is an outcome measure.

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

N/A. This is an outcome measure.

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

N/A. This is an outcome measure.

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

N/A. This is an outcome measure.

1a.7.2. Grade assigned for the quality of the quoted evidence <u>with definition</u> of the grade:

N/A. This is an outcome measure.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

N/A. This is an outcome measure.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: Click here to enter date range

N/A. This is an outcome measure.

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

N/A. This is an outcome measure.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

N/A. This is an outcome measure.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

N/A. This is an outcome measure.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

N/A. This is an outcome measure.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

N/A. This is an outcome measure.

1a.8 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.8.1 What process was used to identify the evidence?

N/A. This is an outcome measure.

1a.8.2. Provide the citation and summary for each piece of evidence.

N/A. This is an outcome measure.

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

Heart_Failure_Excess_Days_in_Acute_Care_NQF_Measure_Evidence_Form_01-29-16_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, and hospitals with information about hospital-level, risk-standardized outcomes following hospitalization for heart failure. Measurement of patient outcomes allows for a broad view of quality of care that cannot be captured entirely by individual process-of-care measures. Safely transitioning patients from hospital to home requires a complex series of tasks which would be cumbersome to capture individually as process measures: timely and effective communication between providers, prevention of and response to complications, patient education about post-discharge care and self-management, timely follow-up, and more. Suboptimal transitions contribute to a variety of adverse events post-discharge, including ED evaluation, need for observation, and readmission. Measures of unplanned readmission already exist, but there are no current measures for ED and observation stay utilization. It is thus difficult for providers and consumers to gain a complete picture of post-discharge outcomes. Moreover, separately reporting each of these outcomes encourages "gaming," such as re-categorizing readmission stays as observation stays to avoid a readmission outcome. By capturing a range of acute care events that are important to patients, we can produce a more complete picture of post-discharge outcomes that better informs consumers about care quality and incentivizes global improvement in transitional care.

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.*

Distribution of EDAC per 100 discharges in the three-year dataset used for measure development. This analysis includes only hospitals that have at least 25 heart failure index admissions in the three-year period.

Time period//2010-2013
Number of hospitals//3,375
Number of discharges//575,672
Mean EDAC (standard deviation)//6.48 (29.63)
Range (minimum - maximum)//263.33 (-67.02 - 196.31
Interquartile range//-14.41– 24.33

Minimum//-67.02
10th percentile//-29.02
20th percentile//-19.06
30th percentile//-10.95
40th percentile//-3.85
50th percentile//3.62
60th percentile//11.24
70th percentile//19.67
80th percentile//29.70
90th percentile//44.41
Maximum//196.31

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

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for 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.*
Distribution of heart failure EDAC (per 100 discharges) by Proportion of Dual-Eligible Patients:
Dates of Data: July 2010 through June 2013 heart failure development dataset
Data Source: Medicare FFS claims
Characteristic//Hospitals with a low proportion (=7.69%) dual-eligible patients)//Hospitals with a high proportion (= 23.08% dual-eligible patients)
Number of Measured Hospitals//1,137//1,139
Number of Patients//131,204 patients in low-proportion hospitals/ 88,954 in high-proportion hospitals
Maximum//223.58 //140.48
90th percentile//37.54 //55.54
75th percentile//15.46 //28.50
Median (50th percentile)//-1.95 //2.03
25th percentile//-18.36//-14.85
10th percentile//-30.28 //-28.23
Minimum //-63.47 //-60.35

Distribution of HF EDAC (per 100 discharges) by Proportion of African-American Patients:
Dates of Data: July 2010 through June 2013 heart failure development dataset
Data Source: Medicare FFS claims
Characteristic// Hospitals with a low proportion (=0.0%) African-American patients//Hospitals with a high proportion (=11.56%) African-American patients
Number of Measured Hospitals//1,851//1,155
Number of Patients//72590 patients in low-proportion hospitals/199,085 in high-proportion hospitals
Maximum//223.58 //144.04
90th percentile//32.52 //51.70
75th percentile//7.98 //31.35
Median (50th percentile)//-7.15// 9.45
25th percentile//-22.81//-7.99
10th percentile//-34.45//-22.63
Minimum//-69.99//-54.39

Low-proportion hospitals are those hospitals whose population of dual-eligible patients or of African-American patients is small enough to place them at or below the 25th percentile among all hospitals; and high proportion are those hospitals whose population of dual eligible patients or African-American patients is large enough to place them at or above the 75th percentile among all hospitals.

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, 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.**
Heart failure was the second most common principal discharge diagnosis among Medicare beneficiaries in 2012 (AHRQ), with nearly half of heart failure patients expected to return to the hospital within six months of discharge (Jencks et al., 2009; Krumholz et al., 1997; Lloyd-Jones et al., 2010). Readmission rates following discharge for heart failure are high and variable across hospitals in the United States (Krumholz et al., 2009; Bernheim et al., 2010). For example, for the time period between July 2012 and June 2013, hospitals' 30-day risk-standardized readmission rates (RSRRs) for heart failure ranged from 17.0% to 28.2% for patients admitted with heart failure (CMS, 2014) Rehospitalization, for any reason, is an undesirable outcome, disruptive to patients and caregivers, costly to the healthcare system, and puts patients at additional risk of hospital-acquired infections and complications. Although some readmissions are unavoidable, others may result from poor quality of care or inadequate transitional care. Transitional care includes effective discharge planning, transfer of information at the time of discharge, patient assessment and education, and coordination of care and monitoring in the post-discharge period. Numerous studies have found an association between quality of inpatient or transitional care and early (typically 30-day) readmission rates for a wide range of conditions including heart failure (Frankl et al., 1991; Corrigan et al., 1992; Oddone et al., 1996; Ashton et al., 1997; Benbassat et al., 2000; Courtney et al., 2003; Halfon et al., 2006; Dean et al., 2006).

Several studies have reported on the relationship between inpatient admissions and other types of hospital care including ED visits and observation stays. ED visits represent a significant proportion of post-discharge acute care utilization. Two recent studies conducted in patients of all ages have shown that 9.5% of patients return to the ED within 30 days of hospital discharge and that about 12% of these patients are discharged from the ED and are not captured by current CMS readmissions measures (Rising et al., 2013; Vashi et al., 2013).

Additionally, over the past decade, the use of observation stays has rapidly increased. Specifically, between 2001 and 2008, the use of observation services increased nearly three-fold (Venkatesh et al., 2011) and significant variation has been demonstrated in the use of observation services for conditions such as chest pain (Schuur et al.,

2011). These rising rates of observation stays among Medicare beneficiaries have gained the attention of patients, providers, and policymakers (Feng et al., 2012; Hockenberry et al., 2014; Rising et al., 2013; Vashi et al., 2013, Wright B. et al., 2014). A report from the Office of the Inspector General (OIG) notes that in 2012, Medicare beneficiaries had 1.5 million observation stays. Many of these observation stays lasted longer than the intended one day. The OIG report also notes the potential relationship between hospital use of observation stays as an alternative to short-stay inpatient hospitalizations as a response to changing hospital payment incentives (Wright, 2013).

Thus, in the context of CMS's publicly reported readmission measures, the increasing use of ED visits and observation stays has raised concerns that current readmission measures do not capture the full range of unplanned acute care in the post-discharge period. By definition, the readmission measures only assess returns to the hospitals for inpatient stays and not for other acute care services, such as observation stays or ED visits. Stakeholders have expressed concerns about whether observation stays should also be evaluated as markers of the quality of care transitions. In particular, there exists concern that high use of observation stays could in some cases replace readmissions, and hospitals with high rates of observation stays in the post-discharge period may therefore have low readmission rates that do not accurately reflect the quality of care (Carlson, 2013).

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

Agency for Healthcare Research and Quality (AHRQ). Healthcare Cost and Utilization Project (HCUP) http://hcupnet.ahrq.gov/.

Ashton CM, Del Junco DJ, Souchek J, Wray NP, Mansyur CL. The association between the quality of inpatient care and early readmission: a meta-analysis of the evidence. Med Care. Oct 1997;35(10):1044-1059.

Carlson J. Faulty Gauge? Readmissions are down, but observational-status patients are up and that could skew Medicare numbers. Modern Healthcare. June 8, 2013 2013.

Benbassat J, Taragin M. Hospital readmissions as a measure of quality of health care: advantages and limitations. Archives of Internal Medicine. Apr 24 2000;160(8):1074-1081.

Carlson J. Faulty Gauge? Readmissions are down, but observational-status patients are up and that could skew Medicare numbers. Modern Healthcare. June 8, 2013 2013.

Centers for Medicare and Medicaid Services (CMS). Medicare Hospital Quality Chartbook Performance Report on Outcome Measures September 2014. September 2014; https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Medicare-Hospital-Quality-Chartbook-2014.pdf.

Corrigan JM, Martin JB. Identification of factors associated with hospital readmission and development of a predictive model. Health Serv Res. Apr 1992;27(1):81-101.

Courtney EDJ, Ankrett S, McCollum PT. 28-Day emergency surgical re-admission rates as a clinical indicator of performance. Ann R Coll Surg Engl. Mar 2003;85(2):75-78.

Feng Z, Wright B, Mor V. Sharp rise in Medicare enrollees being held in hospitals for observation raises concerns about causes and consequences. Health affairs (Project Hope). Jun 2012;31(6):1251-1259.

Frankl SE, Breeling JL, Goldman L. Preventability of emergent hospital readmission. Am J Med. Jun 1991;90(6):667-674.

Halfon P, Eggli Y, Pr, et al. Validation of the potentially avoidable hospital readmission rate as a routine indicator of the quality of hospital care. Medical Care. Nov 2006;44(11):972-981.

Hernandez AF, Greiner MA, Fonarow GC, et al. Relationship between early physician follow-up and 30-day readmission among Medicare beneficiaries hospitalized for heart failure. JAMA : the journal of the American Medical Association. May 5 2010;303(17):1716-1722.

Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. N Engl J Med. 2009;360(14):1418-28.

Hockenberry JM, Mutter R, Barrett M, Parlato J, Ross MA. Factors Associated with Prolonged Observation Services Stays and the Impact of Long Stays on Patient Cost. Health Serv Res. 2014;49(3):893-909.

Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. N Engl J Med. 2009 Apr 2;360(14):1418-28.

Krumholz HM, Parent EM, Tu N, Vaccarino V, Wang Y, Radford MJ, Hennen J. Readmission after hospitalization for congestive heart failure among Medicare beneficiaries. Arch Intern Med. 1997;157:99-104.

Krumholz HM, Merrill AR, Schone EM, Schreiner GC, Chen J, Bradley EH, Wang Y, Wang Y, Lin Z, Straube BM, Rapp MT, Normand SL, Drye EE. 2009. Patterns of hospital performance in acute myocardial infarction and heart failure 30-day mortality and readmission. Circ Cardiovasc Qual Outcomes (2):407-413.
Lloyd-Jones D et al. American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics--2010 update: a report from the American Heart Association. Circulation. 2010 Feb 23;121(7):e46-e215. Epub 2009 Dec 17.
Oddone EZ, Weinberger M, Horner M, et al. Classifying general medicine readmissions. Are they preventable? Veterans Affairs Cooperative Studies in Health Services Group on Primary Care and Hospital Readmissions. Journal of General Internal Medicine. 1996;11(10):597-607.

Rising KL, White LF, Fernandez WG, Boutwell AE. Emergency Department Visits After Hospital Discharge: A Missing Part of the Equation. Annals of Emergency Medicine.

Schuur JD, Baugh CW, Hess EP, Hilton JA, Pines JM, Asplin BR. Critical pathways for post-emergency outpatient diagnosis and treatment: tools to improve the value of emergency care. Academic emergency medicine : official journal of the Society for Academic Emergency Medicine. Jun 2011;18(6):e52-63.

Vashi AA, Fox JP, Carr BG, et al. Use of hospital-based acute care among patients recently discharged from the hospital. JAMA : the journal of the American Medical Association. Jan 23 2013;309(4):364-371.

Venkatesh AK, Geisler BP, Gibson Chambers JJ, Baugh CW, Bohan JS, Schuur JD. Use of observation care in US emergency departments, 2001 to 2008. PloS one. 2011;6(9):e24326.

Wright B., Jung H-Y, Feng Z, Mor V. Hospital, Patient, and Local Health System Characteristics Associated with the Prevalence and Duration of Observation Care. Health Serv Res. 2014;49(4):1088-1107.

Wright S. Hospitals' Use of Observation Stays and Short Inpatient Stays for Medicare Beneficiaries. Washington, DC: OIG;2013.

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

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 : Congestive Heart Failure

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Care Coordination : Readmissions, Safety, Safety : Readmissions

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:** Heart_Failure_Excess_Days_in_Acute_Care_Measure_NQF_Data_Dictionary_01-29-16_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 of the measure is a count of the number of days the patient spends in acute care within 30 days of discharge. We define days in acute care as days spent in an ED, admitted to an observation unit, or admitted as an unplanned readmission for any cause within 30 days from the date of discharge from the index heart failure hospitalization. Each ED treat-and-release visit is counted as one half-day (0.5 days). Observation stays are recorded in terms of hours and are rounded up to the nearest half-day. Each readmission day is counted as one full-day (1 day). We count all eligible outcomes occurring in the 30-day period, even if they are repeat occurrences.

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 the measure as within 30 days of the date of discharge of

the index heart failure hospitalization.

Denominator Time Window: The measure was developed and will be 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.
Outcome Definition
The measure counts ED treat-and-release visits, observation stays, and readmissions to any acute care hospital for any cause within 30 days of the date of discharge of the index heart failure admission, excluding planned readmissions as defined below.

All events which occur within the 30-day window are counted. For example, if a patient returns to the ED three times on three different days, we count each ED visit as a half-day. Similarly, if a patient has two hospitalizations within 30 days, the days spent in each are counted. Therefore, the measure may include multiple ED visits, observation stays, and/or readmissions per patient.

The measure incorporates "exposure time" (the number of days each patient survives after discharge, up to 30). This exposure time is included to account for differential risk for EDAC after discharge among those patients who do not survive the full post-discharge period. If a hospitalization or observation stay extends beyond the 30-day window, only those days within the 30-day window are counted.

Planned Readmission Algorithm
The Planned Readmission Algorithm is a set of criteria for classifying readmissions as planned among the general Medicare population using Medicare administrative claims data. The algorithm identifies admissions that are typically planned and may occur within 30 days of discharge from the hospital.

The Planned Readmission Algorithm has three fundamental principles:
1. A few specific, limited types of care are always considered planned (obstetric delivery, transplant surgery, maintenance chemotherapy/radiotherapy/ immunotherapy, rehabilitation);
2. Otherwise, a planned readmission is defined as a non-acute readmission for a scheduled procedure; and
3. Admissions for acute illness or for complications of care are never planned.

The algorithm was developed in 2011 as part of the Hospital-Wide Readmission measure. In 2013, CMS applied the algorithm to its other readmission measures. In applying the algorithm to condition- and procedure-specific measures, teams of clinical experts reviewed the algorithm in the context of each measure-specific patient cohort and, where clinically indicated, adapted the content of the algorithm to better reflect the likely clinical experience of each measure's patient cohort.

For development of this measure, we used the Planned Readmission Algorithm, Version 3.0. This version and associated code tables are attached in data field S.2b (Data Dictionary or Code Table). For reporting purposes, the measure will use the next version of the Planned Readmission Algorithm, Version 4.0, as will be used in the CMS 30-day heart failure readmission measure.

Definition of Emergency Department Visit and Observation Stay
We defined ED visits and observation stays using specified billing codes or revenue center codes identified in Medicare hospital outpatient claims and physician Carrier claims. The codes that define ED visits and observation stays are in the attached Data Dictionary.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The target population for this measure is Medicare FFS beneficiaries aged 65 years and older hospitalized at non-Federal acute care hospitals for heart failure.

The cohort includes admissions for patients discharged from the hospital with a principal discharge diagnosis of heart failure (see codes below in S.9) and with continuous 12 months Medicare enrollment prior to admission. The measure will be publicly reported by CMS for those patients 65 years and older who are Medicare FFS beneficiaries admitted to non-federal hospitals.

Additional details are provided in S.9 Denominator Details.

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

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)*
To be included in the measure cohort used in public reporting, patients must meet the following inclusion criteria:
1. Having a principal discharge diagnosis of heart failure
2. Enrolled in Medicare fee-for-service (FFS) Part A and Part B for the 12 months prior to the date of the admission, and enrolled in Part A during the index admission;
3. Aged 65 or over;
4. Discharged alive from a non-federal short-term acute care hospital; and,
5. Not transferred to another acute care facility.

International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) codes used to define the cohort for the measure are:
402.01 Malignant hypertensive heart disease with heart failure
402.11 Benign hypertensive heart disease with heart failure
402.91 Unspecified hypertensive heart disease with heart failure
404.01 Hypertensive heart and chronic kidney disease, malignant, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified
404.03 Hypertensive heart and chronic kidney disease, malignant, with heart failure and with chronic kidney disease stage V or end stage renal disease
404.11 Hypertensive heart and chronic kidney disease, benign, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified
404.13 Hypertensive heart and chronic kidney disease, benign, with heart failure and with chronic kidney disease stage V or end stage renal disease
404.91 Hypertensive heart and chronic kidney disease, unspecified, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified
404.93 Hypertensive heart and chronic kidney disease, unspecified, with heart failure and with chronic kidney disease stage V or end stage renal disease
428.0 Congestive heart failure, unspecified
428.1 Left heart failure
428.20 Systolic heart failure, unspecified
428.21 Acute systolic heart failure
428.22 Chronic systolic heart failure
428.23 Acute on chronic systolic heart failure
428.30 Diastolic heart failure, unspecified
428.31 Acute diastolic heart failure

428.32 Chronic diastolic heart failure
428.33 Acute on chronic diastolic heart failure
428.40 Combined systolic and diastolic heart failure, unspecified
428.41 Acute combined systolic and diastolic heart failure
428.42 Chronic combined systolic and diastolic heart failure
428.43 Acute on chronic combined systolic and diastolic heart failure
428.9 Heart failure, unspecified

An ICD-9 to ICD-10 crosswalk is attached in field S.2b. (Data Dictionary or Code Table).

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
The measure excludes index admissions for patients:
1. Without at least 30 days post-discharge enrollment in FFS Medicare.
2. Discharged against medical advice (AMA);
3. Admitted within 30 days of a prior index discharge.

For 2016 public reporting, the measure will also exclude:
4. Admissions with a procedure code for left ventricular assist device (LVAD) implantation or heart transplantation either during the index admission or in the 12 months prior to the index admission. Patients with these procedures are a highly selected group of patients with different risk of the outcome. This exclusion will be added to the heart failure EDAC measure so that it remains fully harmonized with the CMS 30-day heart failure readmission measure. We did not exclude patients with LVAD or heart transplantation from the cohort of admissions used in the analyses for measure development and testing presented here.

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)*
1. Admissions without at least 30 days post-discharge enrollment in FFS Medicare are determined by examining the Medicare Enrollment Database (EDB).
2. Discharges against medical advice (AMA) are identified using the discharge disposition indicator in claims data.
3. Admissions within 30 days of discharge from a qualifying index admission are identified by comparing the discharge date from the index admission with subsequent admission dates.

For 2016 public reporting:
4. Procedure codes for left ventricular assist device (LVAD) implantation or heart transplantation are identified by the corresponding codes included in claims data (see sheet "Cohort Exclusion Codes" in attached Data Dictionary).

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)*
Not applicable. This measure is not stratified.

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).

For risk-adjustment, we used a hierarchical generalized linear model (HGLM). The model consists of two parts, a logit model and a truncated Poisson model. The two-part logit/Poisson model (often called a "hurdle" model) assumes that the outcome results from two related processes: an initial dichotomous event – that a patient has at least one acute care event – which is modeled as the logit of the probability of the event, and for patients with an event (those which clear the "hurdle"), the number of days, which is modeled as a Poisson process. The outcome, number of days, is a half-integer count variable (because ED visits count as 0.5 days). Observation care is counted according to the hours spent in observation care, rounded up to the nearest half-day. For each patient, an exposure variable is defined as the number of survival days post discharge, up to 30. For the hurdle model, exposure time as an offset is included for each part of the model.

There are two random effects for each hospital, one for the logit model and one for the truncated Poisson model, as well as a covariance between the two random effects. The random effects allow us to account for within-hospital correlation of the observed outcome and accommodates the assumption that underlying differences in quality across hospitals lead to systematic differences in outcomes.

We use the existing, NQF-endorsed, CMS 30-day heart failure readmission measure final risk-adjustment variables. We verified the adequacy of this risk-adjustment strategy for our new outcome by comparing the discrimination of models with a full set of all comorbidities to the more parsimonious existing risk models. We found no improvement in model discrimination with the full set, indicating that the existing risk models are adequate.

The measures adjust for variables (i.e., age, comorbid diseases, and indicators of patient frailty) that are clinically relevant and have strong relationships with the outcome. For each patient, risk-adjustment variables are obtained from inpatient, outpatient, and physician Medicare administrative claims data extending 12 months prior to, and including, the index admission.

The model adjusts for case-mix differences based on the clinical status of patients 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 (Pope et al., 2000). A file that contains a list of the ICD-9-CM codes and their groupings into CCs is attached in data field S.2b (Data Dictionary or Code Table). In addition, only comorbidities that convey information about the patient at admission or in the 12 months prior, and not complications that arise during the course of the index hospitalization, are included in the risk adjustment. Hence, we do not risk adjust for CCs that may represent adverse events of care and that are only recorded in the index admission.

The final set of risk-adjustment variables includes the following:
Demographics:
1. Male
2. Age (defined as "Age minus 65" [years above 65, continuous])

Comorbidities:
3. Diabetes mellitus (DM) or DM complications (CC 15-20, 119-120)
4. Iron deficiency or other unspecified anemias and blood disease (CC 47)
5. Congestive heart failure (CC 80)
6. Valvular or rheumatic heart disease (CC 86)
7. Chronic obstructive pulmonary disease (COPD) (CC 108)
8. End-stage renal disease or dialysis (CC 129-130)
9. Other urinary tract disorders (CC 136)
10. Specified arrhythmias and other heart rhythm disorders (CC 92-93)
11. Pneumonia (CC 111-113)
12. Renal failure (CC 131)

13. Vascular or circulatory disease (CC 104-106)
14. Disorders of fluid/electrolyte/acid-base (CC 22-23)
15. Coronary atherosclerosis or angina (CC 83-84)
16. Metastatic cancer or acute leukemia (CC 7)
17. Cancer (CC 8-12)
18. Decubitus ulcer or chronic skin ulcer (CC 148-149)
19. Dementia or other specified brain disorders (CC 49-50)
20. Stroke (CC 95-96)
21. Asthma (CC 110)
22. Acute coronary syndrome (CC 81-82)
23. Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69,100-102,177-178)
24. Protein-calorie malnutrition (CC 21)
25. History of Coronary Artery Bypass Graft (CABG) (ICD-9-CM V45.81, 36.10-36.16)
26. Liver or biliary disease (CC 25-30)
27. Peptic ulcer, hemorrhage, other specified gastrointestinal disorders (CC 34)
28. Other gastrointestinal disorders (CC 36)
29. Severe hematological disorders (CC 44)
30. Drug/alcohol abuse/dependence/psychosis (CC 51-53)
31. Major psychiatric disorders (CC 54-56)
32. Depression (CC 58)
33. Other psychiatric disorders (CC 60)
34. Cardio-respiratory failure or shock (CC 79)
35. Other or unspecified heart disease (CC 94)
36. Fibrosis of lung or other chronic lung disorders (CC 109)
37. Nephritis (CC 132)

References:
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.

Pope GC, et al. 2000. Principal Inpatient Diagnostic Cost Group Models for Medicare Risk Adjustment. Health Care Financing Review 21(3): 93-118.

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*)

S.16. **Type of score:**
Other (specify):
If other: Excess days in acute care (EDAC) per 100 discharges

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.)*
As described above, we used a hierarchical generalized linear model (HGLM). This consists of the two-part logit/truncated Poisson model specifications for days in acute care and includes two random effects for hospitals – one for the logit part and one for the truncated Poisson part – with a non-zero covariance between the two random effects.

This model is used to estimate predicted and expected values for each patient. Predicted values are model predictions that include the hospital random effects, and expected values are model predictions that do not include the hospital random effects. We describe calculation of the predicted and expected values in the attached Appendix (Section 2.7). The measure reports, for each hospital, the difference ("excess") between each hospital's patients' average days in acute care ("predicted days"), and the number of days in acute care that they would have been expected to spend if discharged from an average performing hospital ("expected days"). To be consistent with the reporting of the CMS 30-day heart failure readmission measure, we have multiplied the final score by 100 so that the reported EDAC represents EDAC per 100 discharges.

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.
Administrative claims

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:
1. Medicare Part A inpatient, Part B hospital outpatient claims and physician Carrier claims data: 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.

For development purposes, we obtained the Medicare Part B hospital and physician outpatient claims from the Chronic Condition Data Warehouse (CCW) 100% condition-specific datasets.

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

Reference:
Fleming C, Fisher ES, Chang CH, Bubolz TA, Malenka DJ. 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. Data sources for the all-payer update

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/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

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
Heart_Failure_Excess_Days_in_Acute_Care_NQF_Measure_Testing_Form_01-29-16_v1.1.docx

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

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Excess days in acute care (EDAC) after hospitalization for heart failure
Date of Submission: 1/29/2016
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☒ Outcome (*including PRO-PM*)
☐ 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 decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [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 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:

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, the Medicare Enrollment Database (EDB), and the Chronic Condition Data Warehouse (CCW) 100% condition-specific dataset to capture emergency department (ED) visits and observation stays.

The specific 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 July 1, 2010 through June 30, 2013.

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 hospitals in the United States ([US] including territories) with Medicare Fee-for-Service (FFS) beneficiaries over the age of 65 are included. 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 patients and discharges varies by testing type and samples used. See Section 1.7 for the uses of the development sample and validation sample.

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 hospitals, and number of admissions used in each type of testing are as follows:

For reliability testing (Section 2a2):
The reliability of the model was tested by randomly selecting 50% of the Medicare patients aged 65 years or older in a three-year cohort (July 1, 2010-June 30, 2013) and developing a risk-adjusted model for this group (the "development sample"). We then developed a second model for the remaining 50% of patients (the "validation sample") and compared the two.

The development sample consisted of:
Number of discharges: 590,448
Number of hospitals: 4,626
Patient descriptive characteristics: average (standard deviation [SD]) age = 81.0 (8.2); % male = 44.1%

The validation sample consisted of:
Number of discharges: 590,447
Number of hospitals: 4,634
Patient descriptive characteristics: average (SD) age = 81.0 (8.2); % male = 44.1%

We used the three-year dataset for testing of measure exclusions (Section 2b3).

We used the development sample for calculation of performance score (Section 1b2), model selection (2b4), testing of disparities (Section 1b4), reliability testing (Section 2a2), empirical validity testing (Section 2b2), testing of measure risk adjustment (Section 2b4), and testing to identify meaningful differences in performance (Section 2b5). We also used the development sample to examine 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 (Section 2b4.4b).

We used the validation sample for testing of measure risk adjustment (Section 2b4), and data element and performance measure reliability (Section 2a2).

Data Elements:
• African-American race and dual-eligible status (i.e., enrolled in both Medicare and Medicaid) patient-level data are obtained from Centers for Medicare and Medicaid Services (CMS) enrollment 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. 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, here too though studies suggest a possible increased risk of 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 presumed 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
- African-American race

In selecting variables, our intent was to be responsive to the National Quality Forum (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.

<u>References:</u>

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 aimed to utilize only those data elements from the claims that have both face validity and reliability. We used the final risk-adjustment variables in the existing, NQF-endorsed measure of hospital-level risk-standardized readmission rates following heart failure (NQF #0330).

We avoided the use of fields that are thought to be coded inconsistently across facilities. Specifically, we used fields that are consequential for payment and which are audited. We identified such variables through empiric analyses and our understanding of the CMS auditing and billing policies. We sought to avoid variables which do not meet these standards.

In addition, CMS has in place several hospital auditing programs used to assess overall accuracy of claims-based coding, to ensure appropriate billing, and for overpayment recoupment. CMS routinely conducts data analysis to identify potential problem areas and to 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 assessed the reliability of the data elements by comparing variable frequencies between our development sample and validation sample.

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, is measured again using a second random subset exclusive of the first, then the agreement between the two resulting performance measures across hospitals is calculated (Rousson et al., 2002).

For test-retest reliability, we calculated the EDAC for each hospital using first the development sample, then using the validation sample. Thus, we measured each hospital twice, each time 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) as defined by ICC[2,1] by Shrout and Fleiss (1979) and assessed the values according to conventional standards (Landis and Koch, 1977). We restricted this calculation to hospitals with at least 12 discharges in both samples to approximate the set of hospitals that would have at least 24 discharges over three years and are thus likely to be included in public reporting.

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. In addition, using our split-sample datasets underestimates the test-retest reliability that would be achieved if the measure were reported using three years of data, because the smaller samples for each hospital in one year of data are less reliable. To correct for this underestimate, we used the Spearman-Brown prophecy formula (Spearman 1910, Brown 1910) to adjust the ICC[2,1] to represent three years of data.

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

Risk variable	Development sample (N=590,448)		Validation sample (N=590,447)	
	n	%	n	%
Age, continuous (mean [SD])	81.0 (8.2)	--	81.0 (8.2)	--
Male	260,609	44.1	260,557	44.1
History of Coronary Artery Bypass Graft (CABG) (ICD-9 codes V45.81, 36.10-36.16)	106,935	18.1	106,455	18.0
Diabetes mellitus (DM) or DM complications (CC 15-20, 119-120)	321,646	54.5	320,487	54.3
Disorders of fluid/electrolyte/acid-base (CC 22-23)	291,028	49.3	290,254	49.2
Iron deficiency or other unspecified anemias and blood disease (CC 47)	372,097	63.0	371,958	63.0
Cardio-respiratory failure or shock (CC 79)	160,703	27.2	160,315	27.2
Congestive heart failure (CC 80)	455,321	77.1	454,922	77.1
Vascular or circulatory disease (CC 104-106)	314,822	53.3	314,296	53.2
Chronic obstructive pulmonary disease (COPD) (CC 108)	289,162	49.0	288,601	48.9
Pneumonia (CC 111-113)	270,161	45.8	269,905	45.7
Renal failure (CC 131)	297,618	50.4	296,544	50.2
Other urinary tract disorders (CC 136)	196,599	33.3	196,534	33.3
Decubitus ulcer or chronic skin ulcer (CC 148-149)	86,055	14.6	86,707	14.7
Other gastrointestinal disorders (CC 36)	367,348	62.2	367,211	62.2
Acute coronary syndrome (CC 81-82)	102,372	17.3	102,039	17.3
Valvular or rheumatic heart disease (CC 86)	313,520	53.1	314,086	53.2
Specified arrhythmias and other heart rhythm disorders (CC 92-93)	403,247	68.3	403,252	68.3
Asthma (CC 110)	57,719	9.8	58,061	9.8
Peptic ulcer, hemorrhage, other specified gastrointestinal disorders (CC 34)	92,393	15.7	92,049	15.6

Risk variable	Development sample (N=590,448)		Validation sample (N=590,447)	
	n	%	n	%
Cancer (CC 8-12)	124,978	21.2	125,202	21.2
Drug/alcohol abuse/dependence/psychosis (CC 51-53)	68,980	11.7	69,352	11.8
Major psychiatric disorders (CC 54-56)	62,099	10.5	62,334	10.6
End-stage renal disease or dialysis (CC 129-130)	28,111	4.8	27,637	4.7
Severe hematological disorders (CC 44)	21,696	3.7	21,630	3.7
Nephritis (CC 132)	22,445	3.8	22,597	3.8
Liver or biliary disease (CC 25-30)	62,928	10.7	62,893	10.7
Metastatic cancer or acute leukemia (CC 7)	12,940	2.2	13,280	2.3
Stroke (CC 95-96)	57,476	9.7	57,170	9.7
Dementia or other specified brain disorders (CC 49-50)	143,500	24.3	142,610	24.2
Coronary atherosclerosis or angina (CC 83-84)	440,271	74.6	440,242	74.6
Other or unspecified heart disease (CC 94)	198,418	33.6	197,547	33.5
Other psychiatric disorders (CC 60)	98,793	16.7	98,513	16.7
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)	50,100	8.5	50,231	8.5
Fibrosis of lung or other chronic lung disorders (CC 109)	68,245	11.6	68,297	11.6
Protein-calorie malnutrition (CC 21)	57,291	9.7	57,414	9.7
Depression (CC 58)	117,160	19.8	116,801	19.8

Measure Score Reliability Results

The agreement between the two EDAC values for each hospital was estimated for three years to be ICC[2,1] = 0.73, which according to the conventional interpretation is “substantial” (Landis & Koch, 1977).

Reference

Landis J, Koch G. The measurement of observer agreement for categorical data, Biometric. 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 results are consistent with existing hospital-level measures of patient outcomes. Compared to the development sample, the mean age of patients and the frequencies of the risk-adjustment variables were very similar in the validation sample; this indicates that the data elements are reliable. The ICC [2,1] score of 0.73, estimated for three years of data, demonstrates substantial agreement between samples across the full range of measure values. We interpret this to mean that when used with a full three years of data, the measure will be reliable by the standards of hospital measurement.

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 <u>performance measure score</u> 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)*

We demonstrated measure validity through relevant prior validity testing that we conducted for other claims-based measures, through use of established measure development guidelines, through assessment by external groups, and through systematic assessment of measure face validity by Technical Expert Panel (TEP) of national experts and stakeholder organizations.

<u>Empirical Validity Testing</u>

This measure is closely related in design to the existing, NQF-endorsed readmission measure for patients with heart failure. While the current measure includes additional endpoints and measures them in a different metric (days rather than rates), we would expect that hospitals would have similar – though not identical – performance rankings on the two measures. Thus, as one assessment of validity, we compared the rankings of all hospitals using the two measures to assess the consistency of hospital performance on closely related outcomes. We calculated the Pearson correlation and graphed the readmission measure against the EDAC measure to determine if there were outliers.

<u>Validity of Claims-Based Measures</u>

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 six NQF-endorsed measures currently in public reporting (acute myocardial infarction [AMI], heart failure, 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) (Krumholz et al. 2006; Keenan et al. 2008), acute myocardial infarction (AMI) patients (Cooperative Cardiovascular Project data) (Krumholz, Wang, et al. 2006), and pneumonia patients (National Pneumonia Project dataset) (Bratzler et al. 2011). 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, supporting the use of the claims-based models for public reporting. This measure uses the same risk-adjustment variables that were previously validated in the chart review studies.

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 (National Quality Forum, 2010), CMS Measure Management System guidance, and the guidance articulated in the American Heart Association scientific statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz, Brindis, et al. 2006).

Validity as Assessed by External Groups
Throughout measure development, we obtained expert and stakeholder input via three mechanisms in the initial, early phase of development: a discussion with an advisory Methodology Workgroup, discussions with a national TEP and a 30-day public comment period in order to increase transparency and to gain broader input on the measure.

The Methodology Workgroup meeting addressed key issues related to measure methodology, including weighing the pros and cons of and measure specifications, modeling, and use (e.g., defining the measure cohort and outcome) to ensure the measure is meaningful, useful, and well-designed. The group provided a forum for focused expert review and discussion of technical issues during measure development.

List of Methodology Workgroup Members:
1) Arlene Ash, PhD; University of Massachusetts Medical School (Professor and Division Chief)
2) Jeremiah Brown, MS, PhD; The Dartmouth Institute for Health Policy and Clinical Practice (Assistant Professor of Health Policy and Clinical Practice)
3) Grant Ritter, PhD, MS, MA; Schneider Institute for Health Policy & Heller Graduate School (Senior Scientist)
4) Patrick Romano, MD, MPH; University of California Davis School of Medicine (Professor of Medicine and Pediatrics)

In alignment with the CMS Measures 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 physicians, consumers, purchasers, as well as individuals with experience in quality improvement, performance measurement, and health care disparities. We held two structured TEP conference calls consisting of a presentation of key issues, our proposed approach, and relevant data, followed by open discussion among TEP

members. We solicited additional input and comments from the TEP via e-mail between meetings.

Following completion of the preliminary model, we solicited public comment on the measure through the CMS site link http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/CallforPublicComment.html. The public comments were then posted publicly for 30 days. The resulting input was taken into consideration during the final stages of measure development and led to supplementary analyses reported in the application (1b.4).

Face Validity as Determined by Technical Expert Panel
One means of confirming the validity of this measure was face validity assessed by our Technical Expert Panel (TEP), which included 16 members, including patient representatives, expert clinicians, researchers, providers, and purchasers.

List of TEP members:
1) Kevin E. Driesen, PhD, MPH, MA; Center for Rural Health Mel and Enid Zuckerman College of Public Health, University of Arizona (Assistant Professor & Director of the Arizona Rural Hospital Flexibility Program)
2) David Engler, PhD; America's Essential Hospitals (Senior Vice President for Leadership and Innovation)
3) Timothy Farrell, MD; University of Utah School of Medicine (Assistant Professor of Medicine, Geriatrics; Adjunct Professor of Family Medicine)
4) Karen Farris, PhD; University of Michigan College of Pharmacy (Charles R. Walgreen III Professor of Pharmacy Administration; Director of the Social and Administrative Pharmacy Graduate Program)
5) Maura C. Feldman, MSW; Blue Cross Blue Shield of Massachusetts, Inc (Director for Hospital Performance Measurement and Improvement)
6) Jay A. Gold, MD, JD, MPH; Meta Star, Inc. (Vice President & Chief Medical Officer)
7) Sally Hinkle, DNP, MPA, RN; Temple University Hospital (Director of Performance Improvement & Clinical Value)
8) Amy J.H. Kind, MD, PhD; University of Wisconsin School of Medicine and Public Health (Assistant Professor of Geriatrics)
9) Marjorie King, MD, FACC, MAACVPR; Helen Hayes Hospital (Director of Cardiac Services)
10) Eugene Kroch, PhD; University of Pennsylvania (Adjunct Faculty at the Health Care Systems Department); Premier, Inc. (Vice President & Chief Scientist) University of Pennsylvania; Philadelphia, PA
11) Keith D. Lind, JD, MS, BSN; American Association of Retired Persons (AARP) Public Policy Institute (Senior Policy Advisor)
12) Grace McConnell, PhD; Patient representative
13) Michael A. Ross, MD, FACEP; Emory University School of Medicine (Medical Director of Observation Medicine and Chest Pain Center; Professor of Emergency Medicine)
14) Mark Louis Sanz, MDl; International Heart Institute of Montana (Interventional Cardiologist)

15) Paul Takahashi, MD; Mayo Clinic College of Medicine (Associate Professor of Medicine)
16) Patient representative

We systematically assessed the face validity of the measure score as an indicator of quality by soliciting the TEP members' agreement with the following statement: "*The risk-standardized acute care days obtained from the measures as specified can be used to distinguish between better and worse quality hospitals."*

We measured agreement on a six-point scale: 1=Strongly disagree, 2=Moderately disagree, 3=Somewhat disagree, 4=Somewhat agree, 5=Moderately agree, 6=Strongly agree.

Process Used to Identify International Classification of Diseases, Tenth Revision (ICD-10) Codes Statement of Intent
[X] Goal was to convert this measure to a new code set, fully consistent with the intent of the original measure.
[ ] Goal was to take advantage of the more specific code set to form a new version of the measure, but fully consistent with the original intent.
[ ] The intent of the measure has changed.

Process of Conversion
ICD-10 codes were initially identified using 2013 General Equivalence Mapping (GEM) 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. An ICD-9 to ICD-10 crosswalk is attached in field S.2b. (Data Dictionary or Code Table).

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

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

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

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

National Quality Forum. National voluntary consensus standards for patient outcomes, first report for phases 1 and 2: A consensus report

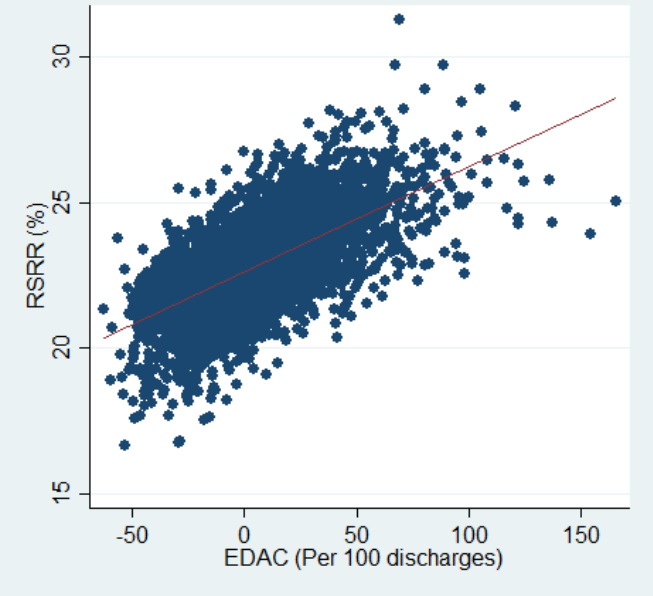
http://www.qualityforum.org/projects/Patient_Outcome_Measures_Phases1-2.aspx. Accessed August 19, 2010.

Krumholz HM, Brindis RG,Brush JE, et al. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. *Circulation.* January 24, 2006 2006;113(3):456-462.

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

Empirical Validity Testing

Comparison of the new measure with the existing CMS 30-day heart failure readmission measure found a Pearson Correlation of 0.714 ($P < 0.0001$). The following figure shows the relationship between risk-standardized readmission rate (RSRR) and EDAC for heart failure:



Systematic Assessment of Face Validity

The results of the TEP rating of agreement with the validity statement were as follows:

N=12

Mean rating=5

Frequency of Ratings of Agreement

Rating	# of Responses	Percent (%)	Cumulative Percent (%)
6 (Strongly agree)	4	33.3%	33.3%
5 (Moderately agree)	6	50.0%	83.3%
4 (Somewhat agree)	1	8.3%	91.7%
3 (Somewhat disagree)	0	0.0%	91.7%

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?*)

Empirical Validity Testing
There was substantial correlation between the two hospital measures, indicating that the proposed measure and the existing readmission measure share underlying properties. This result, and the notable lack of outliers in the figure, provide external empirical validity.

Validity as Assessed by External Groups
The face validity testing results demonstrated TEP agreement with overall face validity of the measure as specified.

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*)

All exclusions were determined by careful clinical review and have been made based on clinically relevant considerations. To ascertain impact of exclusions on the cohort, we examined overall frequencies and proportions of the total cohort excluded for all exclusions, and examined distributions 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. Rationales for the exclusions are detailed in data field S.10 (Denominator Exclusions).

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*)

We examined overall frequencies and proportions of the admissions excluded for each criterion in all heart failure admissions from July 1, 2010 to June 30, 2013.

The exclusion categories are not mutually exclusive.
1. Discharged patients without at least 30 days post-discharge information (0.47%)
2. Discharges against medical advice (AMA) (0.39%)
3. Admissions within 30 days of a prior index admission (8.06%)

For 2016 public reporting, the measure will also exclude:
4. Admissions with a procedure code for left ventricular assist device (LVAD) implantation or heart transplantation either during the index admission or in the 12 months prior to the index admission. This exclusion will be added to the heart failure EDAC measure so that it remains fully harmonized with the CMS 30-day heart failure readmission measure for 2016 reporting. We

did not exclude patients with LVAD or heart transplantation from the cohort of admissions used in the analyses for measure development and testing presented here.

Exclusion	N	%	Distribution across hospitals with ≥ 25 discharges (N=3,996): Minimum, 25th percentile, 50th percentile, 75th percentile, maximum
1. Without at least 30 days post-discharge enrollment in FFS Medicare for index admissions	6,107	0.47	(0.0, 0.0, 0.3, 0.7, 8.0)
2. Discharged against medical advice (AMA)	5,092	0.39	(0.0, 0.0, 0.0, 0.5, 7.4)
3. Heart failure admission within 30 days of a prior heart failure index admission	104,470	8.06	(0.0, 0.7, 7.5, 9.2, 21.4)

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*)

Exclusion 1 (patients without at least 30 days of post-discharge enrollment in FFS Medicare for index admissions) accounts for 0.47% of all index admissions excluded from the initial cohort. This exclusion is needed since the outcome cannot be assessed in this group since claims data are used to determine whether a patient returned to the hospital for an ED visit, was placed under observation care, or was readmitted. Because a very small percent of patients are excluded, this exclusion is unlikely to affect measure score.

Exclusion 2 (patients who are discharged AMA) accounts for 0.39% of all index admissions excluded from the initial index cohort. This exclusion is needed for acceptability of the measure to hospitals, who do not have the opportunity to adequately deliver full care and prepare the patient for discharge. Because a very small percent of patients are excluded, this exclusion is unlikely to affect measure score.

Exclusion 3 (patients with admission within 30 days of a prior index admission) accounts for 8.06% of all index admissions excluded from the initial index cohort. This exclusion is needed to prevent admissions from being counted as both an index admission and a readmission, consistent with the approach taken in the heart failure readmission 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 37 risk factors**

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

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

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (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 is tailored to, and appropriate for, a publicly reported outcome measure as articulated in published scientific guidelines (Krumholz et al. 2006, Normand et al. 2007). We adopted the risk factors from the existing NQF-endorsed CMS 30-day heart failure readmission measure (Dorsey et al. 2015). These risk factors are comprised of age, sex, and condition categories (CCs) for prior 12-month and current claims. These risk factors had been systematically chosen as predictors of any readmission for the same patient cohort as the current measure; the outcome of this measure is dominated by the number of days of a readmission, so we judged it unlikely that repeating the original analysis would produce different results. We confirmed that there were no additional risk factors to consider by comparing the model estimated using the a priori set of risk factors to a model which included all additional CCs.

For risk adjustment, we used a hierarchical generalized linear model (HGLM). The model consists of two parts, a logit model and a truncated Poisson model. The two-part logit/Poisson model (often called a "hurdle" model) assumes that the outcome results from two related processes: an initial dichotomous event – that a patient has at least one acute care event – which is modeled as the logit of the probability of the event, and for patients with an event (those which clear the "hurdle"), the number of days, which is modeled as a Poisson process. The outcome, number of days, is a half-integer count variable (because ED visits count as 0.5 days). Observation care is counted according to the hours spent in observation care, rounded up to the nearest half-day. For each patient, an exposure variable is defined as the number of survival days post-discharge, up to 30. For the hurdle model, exposure time as an offset is included for each part of the model.

There are two random effects for each hospital, one for the logit model and one for the truncated Poisson model, as well as a covariance between the two random effects. The random effects

allow us to account for within-hospital correlation of the observed outcome and accommodates the assumption that underlying differences in quality across hospitals lead to systematic differences in outcomes.

Socioeconomic Status Factors and Race

We selected variables representing 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 days in acute care in the 30 days after discharge.

Our conceptualization of the pathways by which patient SES or race affects days in acute care in the 30 days is informed by the literature on the association of SES and race with heart failure readmissions, since the majority of the EDAC outcome is composed of readmission days, and since there is a much more robust literature about readmission than about observation care and ED visits.

Literature Review of Socioeconomic Status and Race Variables and Heart Failure Excess Days in Acute Care

To examine the relationship between SES and race variables and hospital 30-day, all-cause EDAC following heart failure hospitalization, 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 heart failure readmission. Fifty studies were initially reviewed, and 36 studies were excluded from full-text review based on the above criteria. Studies indicated that SES/race variables were associated with 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; Aseltine et al., 2015; Calvillo-King et al., 2013; McHugh, Carthon, and Kang 2010; Damiani et al., 2015; Berenson and Shih 2012), though there may not be a significant effect on hospital-level profiling (Blum et al., 2014).

Causal Pathways for Socioeconomic Status 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 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 child care), 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 in 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 sufficient 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

We assessed the relationship between the dual-eligible status and race with the outcome and examined the incremental effect of each 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.

Dorsey KB GJ, Desai N, Lindenauer P, Young J, Wang C, DeBuhr, Parisi ML, Bernheim SM, Krumholz HM. 2015 Condition-Specific Measures Updates and Specifications Report Hospital-Level 30-Day Risk-Standardized Readmission Measures: AMI-Version 8.0, HF-Version 8.0, Pneumonia-Version 8.0, COPD-Version 4.0, and Stroke-Version 4.0. 2015; https://www.qualitynet.org/dcs/BlobServer?blobkey=id&blobnocache=true&blobwhere=1228890435217&blobheader=multipart%2Foctet-stream&blobheadername1=Content-Disposition&blobheadervalue1=attachment%3Bfilename%3DRdmn_AMIHFPNCOPDSTK_Msr_UpdtRpt.pdf&blobcol=urldata&blobtable=MungoBlobs. Accessed July 8, 2015.

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 final variables in the model with associated parameter estimates.

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

Risk variable	Part 1: Logit model		Part 2: Poisson model	
	Estimate	CI	Estimate	CI
Age minus 65 (years above 65, continuous)	0.002	(0.001, 0.002)	-0.004	(-0.004, -0.003)
Male	0.033	(0.023, 0.042)	0.006	(0.001, 0.010)
History of coronary artery bypass graft (CABG) (ICD-9 codes V45.81, 36.10-36.16)	-0.049	(-0.066, -0.034)	-0.018	(-0.023, -0.013)
Diabetes mellitus (DM) or DM complications (CC 15-20, 119-120)	0.069	(0.055, 0.082)	0.032	(0.028, 0.037)
Disorders of fluid/electrolyte/acid-base (CC 22-23)	0.130	(0.116, 0.143)	0.034	(0.030, 0.039)
Iron deficiency or other unspecified anemias and blood disease (CC 47)	0.076	(0.063, 0.090)	0.071	(0.066, 0.075)
Cardio-respiratory failure or shock (CC 79)	0.069	(0.056, 0.083)	0.057	(0.051, 0.062)
Congestive heart failure (CC 80)	0.120	(0.104, 0.137)	0.008	(0.001, 0.015)
Vascular or circulatory disease (CC 104-106)	0.064	(0.054, 0.076)	0.011	(0.006, 0.015)
Chronic obstructive pulmonary disease (COPD) (CC 108)	0.124	(0.112, 0.137)	0.062	(0.057, 0.066)
Pneumonia (CC 111-113)	0.080	(0.069, 0.092)	0.063	(0.059, 0.068)
Renal failure (CC 131)	0.136	(0.124, 0.150)	0.095	(0.090, 0.010)
Other urinary tract disorders (CC 136)	0.076	(0.065, 0.087)	0.018	(0.013, 0.022)
Decubitus ulcer or chronic skin ulcer (CC 148-149)	0.078	(0.063, 0.094)	0.080	(0.075, 0.086)
Other gastrointestinal disorders (CC 36)	0.092	(0.078, 0.104)	-0.020	(-0.024, -0.015)
Acute coronary syndrome (CC 81-82)	0.130	(0.115, 0.148)	-0.016	(-0.021, -0.011)
Valvular or rheumatic heart disease (CC 86)	0.032	(0.020, 0.044)	0.022	(0.017, 0.026)

Risk variable	Part 1: Logit model		Part 2: Poisson model	
	Estimate	CI	Estimate	CI
Specified arrhythmias and other heart rhythm disorders (CC 92-93)	0.037	(0.024, 0.048)	0.008	(0.002, 0.014)
Asthma (CC 110)	0.027	(0.008, 0.048)	-0.025	(-0.032, -0.017)
Peptic ulcer, hemorrhage, other specified gastrointestinal disorders (CC 34)	0.049	(0.035, 0.065)	0.032	(0.026, 0.038)
Cancer (CC 8-12)	0.014	(-0.001, 0.027)	0.001	(-0.004, 0.008)
Drug/alcohol abuse/dependence/psychosis (CC 51-53)	0.099	(0.078, 0.117)	-0.042	(-0.048, -0.035)
Major psychiatric disorders (CC 54-56)	0.073	(0.054, 0.092)	0.005	(-0.002, 0.012)
End-stage renal disease or dialysis (CC 129-130)	0.158	(0.134, 0.184)	-0.136	(-0.145, -0.127)
Severe hematological disorders (CC 44)	0.190	(0.157, 0.217)	0.046	(0.036, 0.055)
Nephritis (CC 132)	0.069	(0.037, 0.096)	0.029	(0.021, 0.040)
Liver or biliary disease (CC 25-30)	0.064	(0.046, 0.083)	0.044	(0.038, 0.051)
Metastatic cancer or acute leukemia (CC 7)	0.177	(0.136, 0.211)	0.025	(0.010, 0.039)
Stroke (CC 95-96)	0.047	(0.027, 0.066)	-0.009	(-0.016, -0.003)
Dementia or other specified brain disorders (CC 49-50)	0.084	(0.068, 0.097)	-0.019	(-0.025, -0.014)
Coronary atherosclerosis or angina (CC 83-84)	0.060	(0.044, 0.075)	-0.015	(-0.020, -0.009)
Other or unspecified heart disease (CC 94)	0.053	(0.040, 0.065)	-0.009	(-0.014, -0.004)
Other psychiatric disorders (CC 60)	0.116	(0.102, 0.133)	-0.021	(-0.026, -0.015)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)	0.044	(0.022, 0.065)	0.021	(0.014, 0.028)
Fibrosis of lung or other chronic lung disorders (CC 109)	0.056	(0.038, 0.074)	0.030	(0.024, 0.037)
Protein-calorie malnutrition (CC 21)	0.110	(0.091, 0.126)	0.101	(0.095, 0.109)
Depression (CC 58)	0.030	(0.016, 0.045)	-0.012	(-0.018, -0.007)

References

Dorsey K, Grady J, Wang C, Zhang W et al. 2014 Measures Updates and Specifications Report: Hospital-Level Risk-Standardized Readmission Measures; https://www.qualitynet.org/dcs/BlobServer?blobkey=id&blobnocache=true&blobwhere=1228890294248&blobheader=multipart%2Foctet-stream&blobheadername1=Content-Disposition&blobheadervalue1=attachment%3Bfilename%3DRdmsn_Updts_AMIPNCOPDSTK_032114.pdf&blobcol=urldata&blobtable=MungoBlobs. Accessed September 4, 2014.

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*. 113 (3): 456-462.

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

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 dual-eligible and African-American patients in the heart failure cohort varies across hospitals (number of hospitals = 4,626). The median percentage of dual-eligible patients is 13.78% (interquartile range [IQR] 7.69%-23.08%). The median percentage of black patients is 1.97% (IQR 0%-11.56%).

Empirical Association with the Outcome (Univariate)

The mean patient-level observed days in acute care is higher for dual-eligible patients, 172.70 days in acute care per 100 discharges, compared with 142.84 days in acute care per 100 discharges for all other patients. The mean observed days in acute care for African-American patients was also higher at 174.06 days per 100 discharges compared with 143.52 days per 100 discharges for patients of all other races.

Incremental Effect of Socioeconomic Status Variables and Race in a Multivariable Model

We then examined the strength and significance of the dual-eligible status and race variables in the context of a multivariable model. When we include either of these variables in a multivariate model that includes all of the claims-based clinical variables, the effect size of the variable is small. We also find that the c-statistics for the logit part of the model and the deviance R^2 values for the Poisson part of the model are similar with and without the addition of either of these variables into the model. The c-statistic for the logit model without the dual-eligibility indicator in the model is 0.587 and with the dual-eligibility indicator in the model is 0.588. The c-statistics for the logit model with and without the race indicator are 0.587. The deviance R^2 values for the Poisson model with and without the dual-eligibility indicator are 0.026. The deviance R^2 values for the Poisson model with and without the race indicator are 0.026. Furthermore, we find that the addition of any of these variables into the model has little to no effect on hospital

performance. We examined the change in hospitals' EDAC with the addition of either of these variables. The median absolute change in hospitals' EDAC when adding a dual-eligibility indicator is 0.43 EDAC per 100 discharges (interquartile range [IQR] 0.20-0.75; minimum 0.00-maximum 7.16), with a Spearman correlation coefficient between EDAC for each hospital with and without dual eligibility added of 0.9996. The median absolute change in hospitals' EDAC when adding a race indicator is 0.42 EDAC per 100 discharges (IQR 0.19-0.78; minimum 0.00-maximum 7.91), with a Spearman correlation coefficient between EDAC for each hospital with and without race added of 0.9958.

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

Both the patient-level and hospital-level dual-eligible effects were significant in the logistic part of the heart failure EDAC model, but only the patient-level effect was significant in the Poisson part of the model. This indicates that a) both the patient- and hospital-level dual eligible effects are associated with an increased risk of acute care but b) only the patient-level effect is associated with the expected duration of that care.

Both the patient-level and hospital-level race effects were significantly associated with heart failure EDAC in both the logistic and Poisson models in the decomposition analysis. This indicates that a) both the patient- and hospital-level African-American race effects are associated with a greater risk of having any acute care event and b) both the patient- and hospital-level race effects are associated with the expected duration of acute care following discharge from a heart failure admission.

Because both the hospital- and patient-level effects contribute to the increased risk, if the dual eligible or race variables were used in the model to adjust for patient-level differences, then some of the differences in both risk of acute care and expected duration of care between hospitals would be adjusted for, potentially obscuring a signal of hospital quality.

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

Heart Failure EDAC Decomposition Analysis

Parameter	Logistic model estimate (standard error)	Logistic model p-value	Poisson model estimate (standard error)	Poisson model p-value
Dual Eligible – Patient-Level	0.1101 (0.0084)	<.0001	-0.0196 (0.0031)	<.0001
Dual Eligible – Hospital-Level	0.3441 (0.0376)	<.0001	0.0876 (0.0384)	0.0763
African American – Patient-Level	0.0935 (0.0101)	<.0001	-0.0343 (0.0037)	<.0001

African American – Hospital-Level	0.0545 (0.0236)	<.0001	0.3252 (0.0266)	<.0001

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*)

Dataset

This model selection process was performed using one half (the development sample) of the random three-year split sample.

Approach to Determining Model Specifications

Because the outcome, number of days in acute care, is novel not only for quality measurement but also in the literature as a measure of utilization, we considered a range of model specifications. We performed a number of analyses to determine the best model specification for the number of days in acute care. This is a pseudo-count variable (similar to a count variable, but taking half-integer values for half-days of acute care), and we therefore considered models that were generalized count models. All model development was performed using the development sample.

Inspection of the distribution of the outcome determined that the number of event days was highly skewed, with a large number of zeroes. Thus, we considered models appropriate for skewed data, including approaches that modeled the zero-day outcomes and non-zero day outcomes separately. We only considered approaches that allowed us to incorporate exposure time to account for differential risk.

First, using only patients with non-zero days, we estimated a generalized linear model (GLM) using a Poisson specification, and applied a Park test (Manning and Mullahy, 2001); the Park test indicated that Poisson was the best fit for our outcome. The Poisson model is commonly used for modeling count data and can be generalized to dependent variables that take non-integer values, such as ours.

We then considered three different model specifications for the full set of outcomes (zero and non-zero days): Poisson, zero-inflated Poisson (ZIP), and two-part logit/Poisson ("hurdle" model). For each model, we included an offset for the number of days the patient survived discharge, up to 30 (i.e., the exposure time). For the hurdle model, we included exposure time as an offset for each part because the Poisson part included only observations with non-zero days; it was technically a 'truncated' Poisson model.

For each of the three specifications listed above, we estimated (non-hierarchical) generalized linear models with days in acute care as the outcome. We compared the three different model specifications for the outcome using the following criteria: Akaike information criterion (AIC), Baysian information criterion (BIC), and log-likelihood.

Criterion	Poisson	Zero-inflated Poisson	Two-part logit/Poisson
Akaike information	6,290,000	3,940,000	3,930,000

Criterion	Poisson	Zero-inflated Poisson	Two-part logit/Poisson
criterion (AIC)			
Bayesian information criterion (BIC)	6,290,000	3,940,000	3,930,000
Log-likelihood	-3,095,000	-1,970,000	-1,965,000

We selected the best model based on these statistics and judgment regarding the technical challenges of extending each to a random effects model for the measure. The AIC is a measure of the relative quality of statistical models for a given set of data. The best performing model was the two-part logit/ Poisson model, which had the smallest AIC. This model also made the most sense conceptually, with the likelihood of returning for acute care being modelled separately from the number of days of acute care received.

Assessing Model Discrimination and Calibration

Discrimination: We computed two different statistics – one for the logit part of the model and one for the Poisson part – using the development sample. For the logit model of zero versus non-zero days, which includes all patients in the cohort, we calculated the c-statistic. For the Poisson model of non-zero days, which includes only patients with some acute care, we calculated the deviance R^2. The deviance R^2 is computed from the difference in the log-likelihoods between the final model and an empty model (no covariates) attributed to each observation, averaged over all observations (Cameron, Windmeijer, 1996).

Calibration Statistics

In a generalization of the calibration statistics for logistic models, we calculated the linear prediction Z = XB and W = XC using the coefficients B and C from the development sample and data X from the validation sample. We then estimated a model using the same functional form but only two independent variables, Z for the truncated Poisson part and W for the logit part. The intercepts and coefficients of Z and W in these second models are reported as (γ_0, γ_1), the calibration statistics for each part of the model. The closer they are to (0, 1), the better the model calibration (Harrell, 2013).

Calibration Plot

To further assess model calibration we constructed calibration plots with mean predicted and mean observed days in acute care plotted against decile of predicted utilization rate (predicted days/exposure days).

References

Cameron AC and Windmeijer FAG. R-Squared Measures for Count Data Regression Models with Applications to Health-Care Utilization. Journal of Business & Economic Statistics, Vol. 14, No. 2 (Apr., 1996), pp. 209-220.

Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer New York; 2013.

Manning WG, Mullahy J. Estimating log models: to transform or not to transform? *Journal of health economics.* 2001;20(4):461-494.

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*)**:**

Dataset
The model discrimination statistics were calculated using the development sample:

Discrimination Statistics:
C-statistic for logit part of model: 0.587
Deviance R^2 for truncated Poisson part of model: 0.026 (2.6%)

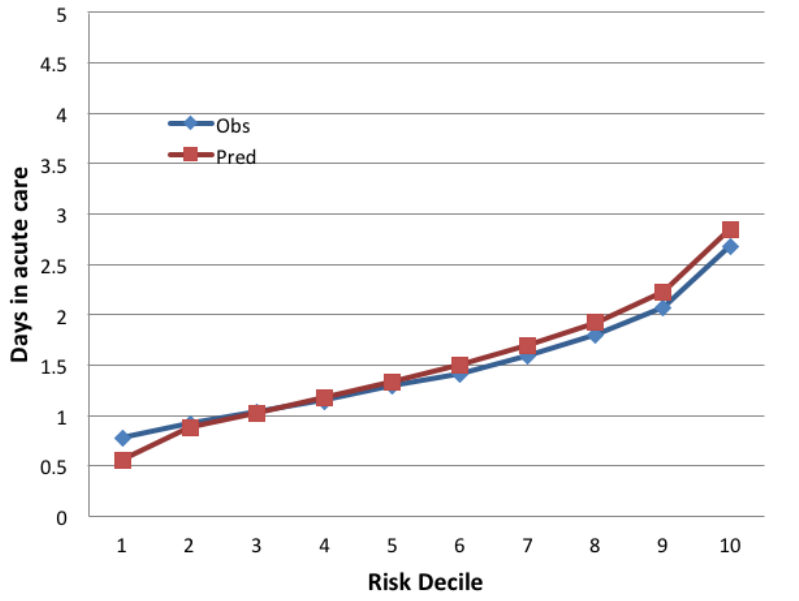
2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Dataset
The model discrimination statistics were calculated using both the development and validation samples; see section 1.7.

Calibration Statistics (y0, y1):
Logit part of model: (0.03, 1.00)
Poisson part of model: (-0.06, 0.97)

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

Calibration Plot:
The plot below shows that the model underestimates risk for the lowest risk decile patients and slightly overestimates risk for the highest risk decile patients.



2b4.9. Results of Risk Stratification Analysis:

N/A. This measure is not risk stratified.

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 for the logit part of the model was 0.59; the deviance R^2 for the Poisson part of 0.026 is consistent with deviance R^2 for other count data models, indicating good model calibration.

Calibration Statistics
Over-fitting (Calibration γ0, γ1)
If the γ_0 in the validation sample is substantially far from zero and the γ_1 is substantially far from one, there is potential evidence of over-fitting. The calibration value of close to 0 at one end and close to 1 to the other end indicates good calibration of the model.

Calibration Plot
The calibration plot shows very good agreement between the mean of predicted days and the mean of observed days within same risk decile.

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*)

N/A.

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)*

To categorize hospital performance, we estimated each hospital's EDAC and the corresponding 95% credible interval (CI) described in the attached Appendix (Section 2.7.2). We assigned hospitals to a performance category by comparing each hospital's EDAC interval estimate to zero. Comparative performance for hospitals with 25 or more eligible cases was classified as follows:

- "Lower than expected" if the entire 95% CI surrounding the hospital's days is below zero.
- "No different than expected" if the 95% CI surrounding the hospital's days includes zero.
- "Higher than expected" if the entire 95% CI surrounding the hospital's days is above zero.

If a hospital has fewer than 25 eligible cases for a measure, we assigned the hospital to a separate category: “The number of cases is too small (fewer than 25) to reliably assess the hospital’s EDAC.”

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*)

Of 4,654 hospitals in the study cohort (data from July 1, 2010 through June 30, 2013), 532 had EDACs "lower than expected," 2,501 were "no different than expected," and 915 had EDACs "higher than expected." 706 were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing. The mean EDAC per 100 discharges for hospitals in the top decile of performance is -29.0, compared to 196.3 for hospitals in the bottom decile.

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 hospital-level EDAC suggests there are meaningful differences in the quality of care received across hospitals for the heart failure EDACmeasure.

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 criterion is directed 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).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of 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 demonstrating comparability of 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.**
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.
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.

Planned	Current Use (for current use provide URL)
Public Reporting Not in use	

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

N/A. The measure is not yet in use.

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?*)

This measure is not currently publicly reported or used in an accountability application because it only recently completed development.

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 has been finalized for use in CMS's Hospital Inpatient Quality Reporting (IQR) program starting in Fiscal Year (FY) 2018 (80 FR 49690).

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**

N/A

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.

Since this measure is not yet in use, there are no performance results to assess improvement.

We expect there will be improvement in measure scores over time since publicly reported measure scores can reduce adverse patient outcomes associated with days spent in acute care for heart failure by capturing and making acute care utilization following the index hospitalization more visible to providers and patients.

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 or model testing. 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)**

0229 : Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following heart failure (HF) hospitalization for patients 18 and older

0506 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization

1551 : Hospital-level 30-day all-cause risk-standardized readmission rate (RSRR) following elective primary total hip arthroplasty (THA) and total knee arthroplasty (TKA)

1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

1891 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization

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?

No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**

We developed the measure in the Medicare Fee-for-Service (FFS) population and completely harmonized the cohort definition and risk-adjustment strategy with those of the existing CMS 30-day heart failure readmission measure. However, while the existing measure counts readmissions as a dichotomous outcome, the proposed measure counts the number of days for all readmissions during the follow-up period, as well as the number of days of observation stays and ED visits. This difference in the outcome measure imposes differences on the statistical modeling and reporting format. There are no differences in data collection burden.

5b. Competing Measures

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

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
N/A

Appendix

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

Attachment **Attachment:** Heart_Failure_Excess_Days_in_Acute_Care_NQF_Appendix_01-29-16_v1.0.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services (CMS)
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 (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.
Yale New Haven Health Services Corporation/Center for Outcomes Research (YNHHSC/CORE) Measure Development Team Members
1. Faseeha K. Altaf, MPH- Lead Project Coordinator. Provided experience relevant to performance measurement.
2. Susannah Bernheim, MD, MHS- Project Director. Provided experience relevant to clinical content and performance measurement.
3. Nihar Desai, MD, MPH- Clinical Consultant. Provided experience relevant to clinical content and performance measurement.
4. Jacqueline Grady, MS- Supporting Analyst. Provided experience relevant to performance measurement.
5. Jeph Herrin, PhD- Statistician. Provided experience relevant to performance measurement.
6. Leora Horwitz, MD, MHS- Project Lead. Provided experience relevant to clinical content and performance measurement.
7. Zhenqiu Lin, PhD- Director of Analytics. Provided experience relevant to performance measurement.
8. Shuling Liu, PhD- Statistical Consultant. Provided experience relevant to performance measurement.
9. Chi Ngo, MPH- Research Associate. Provided experience relevant to performance measurement.
10. Arjun Venkatesh, MD, MBA- Clinical Consultant. Provided experience relevant to clinical content and performance measurement.
11. Changqin Wang, MD, MS - Lead Analyst. Provided experience relevant to performance measurement.
12. Yongfei Wang- Supporting Analyst. Provided experience relevant to performance measurement.
13. Sharon-Lise Normand, Ph.D.* - Statistical Consultant. Provided statistical expertise for the project.

*Harvard Medical School

Technical Expert Panel (TEP) Members
1. Anonymous Patient- Patient Representative. Provided patient perspective.
2. Kevin E. Driesen, PhD, MPH, MA- Assistant Professor, Mel and Enid Zuckerman College of Public Health; Director, Arizona Rural

Hospital Flexibility Program. Provided experience relevant to performance measurement.
3. David Engler, PhD- Senior Vice President for Leadership and Innovation, America's Essential Hospitals. Provided experience relevant to clinical content, performance measurement, and coding and informatics.
4. Timothy Farrell, MD- Assistant Professor of Medicine, Adjunct Professor of Family Medicine, Physician Investigator; University of Utah School of Medicine. Provided experience relevant to clinical content and performance measurement.
5. Karen Farris, PhD- Charles R. Walgreen III Professor of Pharmacy Administration, Director of the Social and Administrative Pharmacy Graduate Program; University of Michigan College of Pharmacy. Provided experience relevant to performance measurement.
6. Maura C. Feldman, MSW- Director for Hospital Performance Measurement and Improvement, Blue Cross Blue Shield of Massachusetts. Provided consumer perspective.
7. Jay A. Gold, MD, JD, MPH- Senior Vice President and Chief Medical Officer, MetaStar. Provided experience relevant to clinical content and performance measurement.
8. Sally Hinkle, DNP, MPA, RN- Director of Performance Improvement and Clinical Value, Temple University Hospital. Provided experience relevant to performance measurement.
9. Amy Jo Haavisto Kind, MD, PhD - Assistant Professor of Geriatrics, University of Wisconsin School of Medicine and Public Health; Attending Physician, William S. Middleton VA. Provided experience relevant to clinical content and performance measurement.
10. Marjorie King, MD, FACC, MAACVPR- Director of Cardiac Services, Helen Hayes Hospital. Provided experience relevant to clinical content and performance measurement.
11. Eugene Kroch, PhD- Vice President and Chief Scientist, Premier. Provided experience relevant to performance measurement.
12. Keith D. Lind, JD, MS, BSN- Senior Policy Advisor, American Association of Retired Persons (AARP) Public Policy Institute. Provided consumer perspective.
13. Grace McConnell, PhD- Patient Representative. Provided patient perspective.
14. Michael A. Ross, MD, FACEP- Medical Director, Professor of Emergency Medicine; Emory University School of Medicine. Provided experience relevant to clinical content and performance measurement.
15. Mark Louis Sanz, MD- Interventional Cardiologist, International Heart Institute of Montana. Provided experience relevant to clinical content and performance measurement.
16. Paul Takahashi, MD- Associate Professor of Medicine, Mayo Clinic College of Medicine. Provided experience relevant to performance measurement.

Methodology Work Group Members
1. Arlene Ash, PhD- Professor and Division Chief, University of Massachusetts Medical School. Provided experience relevant to performance measurement.
2. Jeremiah Brown, PhD, MS- Assistant Professor of Health Policy and Clinical Practice, The Dartmouth Institute for Health Policy and Clinical Practice. Provided experience relevant to performance measurement.
4. Grant Ritter, PhD, MS, MA- Senior Scientist, Schneider Institute for Health Policy & Heller Graduate School. Provided experience relevant to performance measurement.
5. Patrick Romano, MD, MPH- Professor of Medicine and Pediatrics, University of California Davis School of Medicine. Provided experience relevant to performance measurement.

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:**
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

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 #: 2881
De.2. **Measure Title:** Excess days in acute care (EDAC) after hospitalization for acute myocardial infarction (AMI)
Co.1.1. **Measure Steward:** Centers for Medicare & Medicaid Services (CMS)
De.3. **Brief Description of Measure:** This measure assesses days spent in acute care within 30 days of discharge from an inpatient hospitalization for acute myocardial infarction (AMI) to provide a patient-centered assessment of the post-discharge period. This measure is intended to capture the quality of care transitions provided to discharged patients hospitalized with AMI by collectively measuring a set of adverse acute care outcomes that can occur post-discharge: emergency department (ED) visits, observation stays, and unplanned readmissions at any time during the 30 days post-discharge. In order to aggregate all three events, we measure each in terms of days. In 2016, CMS will begin annual reporting of the measure for patients who are 65 years or older, are enrolled in fee-for-service (FFS) Medicare, and are hospitalized in non-federal hospitals.
1b.1. **Developer Rationale:** The goal of this measure is to improve patient outcomes by providing patients, physicians, and hospitals with information about hospital-level, risk-standardized outcomes following hospitalization for AMI. Measurement of patient outcomes allows for a broad view of quality of care that cannot be captured entirely by individual process-of-care measures. Safely transitioning patients from hospital to home requires a complex series of tasks which would be cumbersome to capture individually as process measures: timely and effective communication between providers, prevention of and response to complications, patient education about post-discharge care and self-management, timely follow-up, and more. Suboptimal transitions contribute to a variety of adverse events post-discharge, including ED evaluation, need for observation, and readmission. Measures of unplanned readmission already exist, but there are no current measures for ED and observation stay utilization. It is thus difficult for providers and consumers to gain a complete picture of post-discharge outcomes. Moreover, separately reporting each of these outcomes encourages "gaming," such as re-categorizing readmission stays as observation stays to avoid a readmission outcome. By capturing a range of acute care events that are important to patients, we can produce a more complete picture of post-discharge outcomes that better informs consumers about care quality and incentivizes global improvement in transitional care.

S.4. **Numerator Statement:** The outcome of the measure is a count of the number of days the patient spends in acute care within 30 days of discharge. We define days in acute care as days spent in an ED, admitted to an observation unit, or admitted as an unplanned readmission for any cause within 30 days from the date of discharge from the index AMI hospitalization. Each ED treat-and-release visit is counted as one half-day (0.5 days). Observation stays are recorded in terms of hours and are rounded up to the nearest half-day. Each readmission day is counted as one full day (1 day). We count all eligible outcomes occurring in the 30-day period, even if they are repeat occurrences.
S.7. **Denominator Statement:** The target population for this measure is Medicare FFS beneficiaries aged 65 years and older hospitalized at non-federal acute care hospitals for AMI.

The cohort includes admissions for patients discharged from the hospital with a principal discharge diagnosis of AMI (see codes below in S.9) and with continuous 12 months Medicare enrollment prior to admission. The measure will be publicly reported by CMS for those patients 65 years and older who are Medicare FFS beneficiaries admitted to non-federal hospitals.

Additional details are provided n S.9 Denominator Details.
S.10. **Denominator Exclusions:** The measure excludes index admissions for patients:
1. Without at least 30 days post-discharge enrollment in FFS Medicare;
2. Discharged against medical advice (AMA);
3. Admitted within 30 days of a prior index discharge;
4. Admitted and then discharged on the same day (because it is unlikely these are clinically significant AMIs).

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Administrative claims
S.26. **Level of Analysis:** Facility

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

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 any measure; however, it is harmonized with a measure of hospital-level, all-cause, 30-day, risk-standardized readmission following AMI hospitalization.

New Measure -- Preliminary Analysis

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.

- This measure calculates excess days in acute care (EDAC) for patients with AMI. This measure is intended to capture the quality of care transitions provided to discharged patients hospitalized with AMI by collectively measuring a set of adverse acute care outcomes that can occur post-discharge: emergency department (ED) visits, observation stays, and unplanned readmissions at any time during the 30 days post-discharge. In order to aggregate all three events, this measure assesses each in terms of days.
- As a rationale for measuring this health outcome, the developer suggests that hospitals are able to influence readmission rates through a broad range of clinical activities including communication between providers, prevention of, and response to, complications, patient safety and coordinated transitions to the outpatient environment.
- The developer cites that "in the context of the Centers for Medicare and Medicaid Services' (CMS's) publicly reported readmission measures, the increasing use of ED visits and observation stays has raised concerns that current readmission measures do not capture the full range of unplanned acute care in the post-discharge period (Vashi et al., 2013; Rising et al., 2012; Feng et al., 2012). Observation stays can occur in many different parts of the hospital, including dedicated treatment rooms, the ED, or inpatient units. In particular, there is concern that high use of observation stays could in some cases replace readmissions, and that hospitals with high rates of observation stays in the post-discharge period may therefore have low readmission rates that do not accurately reflect the quality of care (Vashi et al., 2013)."
- The developer also explains for AMI specifically "studies suggest that appropriate care for AMI during and after the index hospitalization may reduce the risk of subsequent readmission (Carroll et al., 2007; Young et al., 2003; Bondestam et al., 1995; Ades et al, 1992; Carlhed et al., 2009)."

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

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 developer provides performance data from one measurement period from 2010-2013, covering a total of 232,954 discharges.
- The data show that during the measurement period of 2010-2013, AMI readmission rates ranged from a minimum of -54.82% to a maximum of 170.44%, with the 10th percentile at -23%, the 50th percentile at 5.46%, and the 90th percentile at 46.05%.

Disparities

- To help in assessment of potential disparities, the developers also provide performance scores for hospitals serving a low proportion of dual eligible patients vs. those serving a high proportion of dual eligible patients and performance scores for hospitals serving a low proportion of African-American patients vs. those serving a high proportion of African-American patients.

- By proportion of **Dual Eligible Patients:**

  // Low proportion (=0%) dual-eligible patients//Hospitals with a high proportion (=21.05%) dual eligible patients
  Number of Measured Hospitals//1,082 //1,046
  Number of Patients//5,142 patients in low-proportion hospitals/ 24,494 in high-proportion hospitals
  Maximum//214.66 //386.10
  90th percentile//26.44 //61.03
  75th percentile//0.74 //20.35
  Median (50th percentile)//-0.68 //-0.30
  25th percentile//-12.08//-15.59
  10th percentile//-28.08//-31.99
  Minimum //-93.19//-97.78

- By proportion of **African-American Patients:**

  // Low proportion (=0%) African-American patients//Hospitals with a high proportion (=7.32%) African-American patients
  Number of Measured Hospitals//2,229 //1,038
  Number of Patients//42,537 patients in low-proportion hospitals/ 82,236 in high-proportion hospitals
  Maximum//386.1 //322.32
  90th percentile//34.63//59.20
  75th percentile//3.43//29.37
  Median (50th percentile)//-1.02//6.81
  25th percentile//-17.35//-9.48
  10th percentile//-31.36//-25.12
  Minimum//-97.78//-88.77

• The developer explains that: “low-proportion hospitals are those hospitals whose population of dual-eligible patients or of African-American patients is small enough to place them at or below the 25th percentile among all hospitals; and high proportion are those hospitals whose population of dual eligible patients or African-American patients is large enough to place them at or above the 75th percentile among all hospitals.”

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)

1. Importance to Measure and Report

1a. Evidence to Support Measure Focus

Comments: **The evidence as outlined may support the measure, however the inclusion of observations, and ED visits in the unplanned admissions may have unintended consequences, patient post AMI 30 days may need observation as a clinically sound intervention to prevent hospitalization. Most patient instructions educate the patient when to call the MD and when to go to ED.

**The published evidence is weak that there are interventions that will reduce days of acute care following a myocardial infarction. There are reasons to believe that such interventions may exist.

1b. Performance Gap

Comments: **Yes. Yes. Unclear, it appears that in the 90th percentile, there is the biggest differences in findings, and warrant improvement. However, if the measure includes observation and ED, the low income, dual eligible, poor may have a disproportionate readmission, however the standard of care may be the same as the general population.

**Unknown performance gap in that we don't know what number of days post discharge represents good care. However, there is a large variation across hospitals.

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

Comments: **NA

**NA

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

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): Administrative claims

- This measure calculates the number of days spent in acute care within 30 days of discharge from an inpatient hospitalization for acute myocardial infarction (AMI) to provide a patient-centered assessment of the post-discharge period.
- The outcome of the measure is a count of the number of days the patient spends in acute care within 30 days of discharge. The measure defines days in acute care as days spent in an ED, admitted to an observation unit, or admitted as an unplanned readmission for any cause within 30 days from the date of discharge from the index heart failure hospitalization. Each ED treat-and-release visit is counted as one half-day (0.5 days). Observation stays are recorded in terms of hours and are rounded up to the nearest half-day. Each readmission day is counted as one full-day (1 day). The measure counts all eligible outcomes occurring in the 30-day period, even if they are repeat occurrences.
- The Numerator if the number of days the patient spends in acute care within 30 days of discharge. Days in acute care is defined as days spent in an ED, admitted to an observation unit, or admitted as an unplanned readmission for any cause within 30 days from the date of discharge from the index AMI hospitalization
- The Denominator is the Medicare FFS beneficiaries aged 65 years and older hospitalized at non-federal acute care hospitals for AMI.
- The data sources for this measure are Medicare Part A inpatient, Part B hospital outpatient claims and physician Carrier claims, and the Medicare Enrollment Database (EDB).
- The measure's time window is three years.
- The measure is risk-adjusted using a statistical risk model (see details below).

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**

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**

- The developer has assessed reliability at both the data element and the performance score levels.
- The datasets used for testing included Medicare Parts A and B claims, the Medicare Enrollment Database (EDB), and the Chronic Condition Data Warehouse (CCW) 100% condition-specific dataset to capture emergency department (ED) visits and observation stays. Additionally, census data were used to assess socio-demographic factors.
  - **Data element reliability**:
    - With regard to data element reliability, the developer notes that the measure has been developed to avoid the use of claims data elements that are thought to be coded inconsistently across hospitals or providers, instead using fields that are consequential for payment and which are audited by CMS. Additionally, the developer used the final risk-adjustment variables in the existing, NQF-endorsed measure of hospital-level risk-standardized readmission rates following AMI (NQF #0505).
    - Additionally, the developer compared variable frequencies between the development and validation samples.
  - **Performance score reliability**:
    - The developer defines performance score reliability as the degree to which repeated measurements of the same entity agree with each other.
    - In line with this thinking, the developer's approach to assessing score-level reliability was to consider the extent to which assessments of a hospital using different but randomly-selected subsets of patients produce similar measures of hospital performance. The developers refer to this as a "test-retest" approach; it may also be called a "split-half" method.
    - For test-retest reliability, the developer calculated the EDAC for each hospital using first the development sample, then the validation sample. Thus, each hospital twice was measured twice, each time using an entirely distinct set of patients. The developer states that the extent to which the calculated measures of these two subsets agree is evidence that the measure is assessing an attribute of the hospital, not of the patients. As a metric of agreement, the developer calculated the intra-class correlation coefficient (ICC) as defined by ICC[2,1] by Shrout and Fleiss (1979) and assessed the values according to conventional standards (Landis and Koch, 1977).
    - A total of 496,716 admissions were examined, with 248,358 in each sample.
    - The agreement between the two EDAC values for each hospital (as measured by an intra-class correlation coefficient (ICC)) was **0.54**; the developer states that according to the conventional interpretation, this is considered a "moderate" level of agreement.
    - The developer notes that this analysis was limited to hospitals with to hospitals with at least 12 discharges in both samples to approximate the set of hospitals that would have at least 24 discharges over three years and are thus likely to be included in public reporting. [*Note: It is unclear whether the measure itself is limited to hospitals with 12 or more cases and if three years of data are needed to calculate the measure; if it is not, then testing was not conducted with the measure as specified*.]

- The developer expects that the correlation coefficient would be higher using a full three-year sample since it would include more patients. To correct this problem, the developer used the Spearman-Brown prophecy formula (Spearman 1910, Brown 1910) to adjust the ICC[2,1] to represent three years of data.

- The developer's overall interpretation of reliability testing results is that the compared to the development sample, the mean age of patients and the frequencies of the risk-adjustment variables were very similar in the validation sample; this indicates that the data elements are reliable and that the ICC score from performance score analysis demonstrates moderate agreement across samples. The developer notes that the ICC [2,1] score of 0.54, estimated for three years of data, demonstrates moderate agreement between samples across the full range of measure values. We interpret this to mean that when used with a full three years of data, the measure will be reliable by the standards of hospital measurement.

Guidance from the Reliability Algorithm

- Question 1. Submitted specifications are precise, unambiguous, and complete. Measure can be consistently implemented.
- Question 2. Empirical reliability testing was conducted using statistical tests with the measure as specified.
- Question 3. Empirical validity testing of patient-level data was conducted.
- Question 4. Reliability testing was conducted with computed performance measure scores for each measured entity.
- Question 5. Random split-half correlation was used to assess the proportion of variability due to real differences among the measured entities.
- Question 6. The ICC was 0.54 which is considered a moderate level of agreement

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?*
- *Does the measure need three years of data to achieve this level of reliability?*

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

2b. Validity

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**

- This measure calculates the number of days spent in acute care within 30 days of discharge from an inpatient hospitalization for AMI to provide a patient-centered assessment of the post-discharge period.
- As a rationale for measuring this health outcome, the developer suggests that hospitals are able to influence readmission rates through a broad range of clinical activities including communication between providers, prevention of, and response to, complications, patient safety and coordinated transitions to the outpatient environment.

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.

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**

- The developer demonstrated measure validity through prior validity testing done on their claims-based measures, through use of established measure development guidelines, and by systematic assessment of measure face validity by a Technical Expert Panel (TEP).
    - Empirical Validity Testing
        - The developer notes this measure is closely related in design to the existing, NQF-endorsed readmission measure for patients with AMI. While this measure includes additional endpoints and measures them in a different metric (days rather than rates), the developer expects that hospitals would have similar – though not identical – performance rankings on the two measures. Therefore as one assessment of validity, they compared the rankings of all hospitals using the two measures to assess the consistency of hospital performance on closely related outcomes. The developer calculated the Pearson correlation, and graphed the readmission measure against the EDAC measure to determine if there were outliers.
        - Comparison of the new measure with the existing CMS 30-day AMI readmission measure found a Pearson correlation of 0.610 ($P < 0.0001$)
    - Validity of Claims-Based Measures:
        - The developer states that they have demonstrated for a number of other readmission measures the validity of claims-based measures by comparing either the measure result or the individual data elements against medical records.
        - Claims model validation was conducted by building comparable models using abstracted medical chart data for risk adjustment. 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.
    - Validity Indicated by Established Measure Development Guidelines
        - The developer states that this measure was developed in consultation with national guidelines for publicly reported outcomes measures, with outside experts, and with the public.
    - Validity as Assessed by External Groups:
        - Input was obtained through regular discussions with an advisory working group, a TEP, and a 30-day public comment period.
    - Face Validity as Determined by TEP:
        - The developer asked members of the TEP to note their agreement with the statement "The risk-standardized acute care days obtained from the measures as specified can be used to distinguish between better and worse quality hospitals."
        - Of the TEP members who responded, 91% agreed (83% moderately or strongly agreed) that the measure will provide an accurate reflection of quality.
        - The developer interpreted this as a moderate level of agreement.

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?*
- *Is the claims model validation sufficiently similar to the measure?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- Patients in the following categories are excluded from the measure:
  - Discharged patients without at least 30 days post-discharge information
  - Discharges against medical advice (AMA)
  - Admissions within 30 days of a prior index admission
  - Same-day discharges
- The developer notes that all exclusions were determined by careful clinical review and have been made based on clinically relevant decisions and to ensure accurate calculation of the measure
- To determine the impact of exclusions, the developer examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion.
- The number and percentage of patients excluded for each criterion are as follows:
  - Without at least 30 days post-discharge enrollment in FFS Medicare for index admissions: **3,169 (0.62%)**
  - Discharged against medical advice (AMA): **2,393 (.47%)**
  - Admissions within 30 days of a prior index admission: **8,907 (1.74%)**
  - Same-day discharges: **2,429 (0.47%)**
- The developer also provides the distribution across hospitals for each exclusion criterion.
- The developer notes that the first exclusion criterion, is needed since the outcome cannot be assess in this group since claims data are used to determine whether a patient returned to the hospital for an ED visit, was placed under observation care, or was readmitted.
- The developer states that the second exclusion criterion is needed for acceptability of the measure to hospitals, who do not have the opportunity to adequately deliver full care and prepare the patient for discharge.
- The developer notes that exclusion criterion 2 is needed to prevent admissions from being counted as both an index admission and a readmission, consistent with the approach taken in the AMI readmission measure.
- The developer states that exclusion criterion 4 is meant to ensure a clinically coherent cohort. This exclusion prevents the inclusion of patients who likely did not suffer a clinically significant AMI. For most hospitals this results in very few patients being excluded.

Questions for the Committee:

- *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: **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

- For this measure the develop adopted the risk factors from the existing NQF-endorsed CMS 30-day AMI readmission measure. These risk factors are comprised of age, sex, and condition categories (CCs) for prior 12-

month and current claims.

- The developer notes these risk factors had been systematically chosen as predictors of any readmission for the same patient cohort as the current measure; the outcome of this measure is dominated by the number of days of a readmission, so they judged it unlikely that repeating the original analysis would produce different results.
- The developer confirmed that there were no additional risk factors to consider by comparing the model estimated using the a priori set of risk factors to a model which included all additional CCs.
- The measure employs a hierarchical generalized linear model [HGLM]) that consists of two parts, a logit model and a truncated Poisson model. The two-part logit/Poisson model (often called a "hurdle" model) assumes that the outcome results from two related processes: an initial dichotomous event – that a patient has at least one acute care event – which is modeled as the logit of the probability of the event, and for patients with an event (those which clear the "hurdle"), the number of days, which is modeled as a Poisson process. The outcome, number of days, is a half-integer count variable (because ED visits count as 0.5 days).
- There are two random effects for each hospital, one for the logit model and one for the truncated Poisson model, as well as a covariance between the two random effects. The developer suggests that the random effects allows for within-hospital correlation of the observed outcome and accommodates the assumption that underlying differences in quality across hospitals lead to systematic differences in outcomes.
- The final set of 31 risk-adjustment variables is included in the testing attachment; the odds ratio associated with each variable is also provided.
- The developers also considered a number of variables related to sociodemographic status (SDS) for potential inclusion in the risk-adjustment model. Candidate SDS variables were selected for examination based on a review of literature and national data sources.
- **Conceptual analysis of the need for SDS adjustment:**
    - The developers note there is a large body of literature linking various SES factors and African-American race to worse health status and higher readmission risk with income, education, and occupational level being the most commonly examined variables. The developers state that the literature directly examining how SES factors or race might influence the likelihood of older, insured, Medicare patient of being readmitted within 30 days of an admission for heart failure is more limited.
    - The developers state that few studies directly address causal pathways for SDS factors to affect 30-day readmission rates or examine the role of the hospital in these pathways.
    - There are at least four potential pathways for SDS factors to affect 30-day readmission rates:
        - One potential pathway is the relationship to health status at the time of admission. SDS factors 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. The developers note that this pathway should be largely accounted for by their clinical risk-adjustment model.
        - The next potential path way is that patients with low income and African-American patient are more likely to be seen in lower quality hospitals, which can contribute to increased risk of readmission.
        - The third major pathway is that a patient's race or SDS status cause them to experience differential, lower quality care or may not receive the differentiated care they require.
        - Finally, some SES risk factors may affect the likelihood of readmission without directly affecting health status at admission or the quality of care received during the hospitalization. Patients may have worse outcomes due to competing economic priorities or a lack of access to care outside the hospital.

- **Empirical analysis of SDS factors:**
    - The developers considered African-American race, and dual-eligible status-i.e. enrolled in both Medicare and Medicaid. The developers assessed the relationship between the SES variables and race with the outcome and examined the incremental effect in a multivariable mode.
    - The developer assessed the relationship between the SDS variables and the days in acute care and examined the incremental effect of SDS in a multivariable model, evaluating the extent to which the addition of any one of these variables improved model performance or changed hospital results.
    - The developer notes that 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. Therefore, the developers performed a decomposition analysis to assess the independent effects of the SES and race variables at the patient level and the hospital level.
    - The developers' analysis found that the prevalence of SDS factors in the AMI cohort does vary across measured entities.
    - With regard to the empirical association of each SDS variable with the outcome (univariate), the analysis found that patient-level observed days in acute care for dual-eligible patients was higher, at 141.75 per 100 discharges compared with 106.65 days in acute care per 100 discharges for all other patients. The readmission rate for African-American patients was also higher at 148.38 days per 100 discharges compared with 107.62 days per 100 discharges for patients of all other races.
    - With regard to the strength and significance of the SDS variables in the context of a multivariable model, the developers' analysis found that the effect size of each of these variables is small. The developers also found that the c-statistics (i.e., predictive value) for the logit part of the model and the deviance R2 values for the Poisson part of the model are similar with and without the addition of either of these variables into the model. Additionally the developers found the addition of these variables has little to no effect on hospital performance.
        - The median absolute change in hospitals' EDAC when adding a dual-eligibility indicator is 0.50 EDAC per 100 discharges (interquartile range [IQR] 0.23-0.98; minimum 0.00-maximum 24.59), with a Spearman correlation coefficient between EDAC for each hospital with and without dual eligibility added of 0.9933.
        - The median absolute change in hospitals' EDAC when adding a race indicator is 0.5002 EDAC per 100 discharges (IQR 0.23-0.97; minimum 0.00-maximum 12.71), with a Spearman correlation coefficient between EDAC for each hospital with and without race added of 0.9936.
    - The developers state that both the patient-level and hospital-level dual eligible and race effects were significant in the logistic part of the AMI EDAC model, but only the hospital-level effect was significant in the Poisson part of the model. This indicates that a) both the patient- and hospital-level dual eligible and race effects are associated with an increased risk of acute care but b) only the hospital-level effect is associated with the expected duration of that care. The developers note that if the dual eligible or race are 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 developers state that given these findings and complex pathways that could explain any relationship between SDS and readmission, they did not incorporate SDS variables into the measure.
- **Risk Model Diagnostics:**
    - To assess model discrimination the developers computed two different statistics: one for the logit part of the model and one for the Poisson part.
        - For the logit model of zero versus non-zero days, which includes all patients in the cohort, the

developers calculated the c-statistic.
        - C-statistic for logit part of model: 0.60
    - For the Poisson model of non-zero days, which includes only patients with some acute care, the developers calculated the deviance R2. The deviance R2 is computed from the difference in the log-likelihoods between the final model and an empty model (no covariates) attributed to each observation, averaged over all observations.
        - Deviance R2 for truncated Poisson part of model: 0.040 (4.0%)
    - The developers interpret these results as good model calibration.
- In a generalization of the calibration statistics for logistic models, the developers calculated the linear prediction Z = XB and W = XC using the coefficients B and C from the development sample and data X from the validation sample. The developers then estimated a model using the same functional form but only two independent variables, Z for the truncated Poisson part and W for the logit part. The intercepts and coefficients of Z and W in these second models are reported as (γ_0, γ_1), , the calibration statistics for each part of the model. The closer they are to (0, 1), the better the model calibration
    - Calibration Statistics (y0, y1):
        - Logit part of model: (-0.10, 0.98)
        - Poisson part of model: (-0.04, 0.97)

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?*
- *Does the Committee agree with the developer's use of current claims data for risk adjustment variables?*
- *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 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)*:

- To categorize hospital performance, the developers estimated each hospital's EDAC and the corresponding 95% credible interval (CI).
- The developers then assigned hospitals to a performance category by comparing each hospital's EDAC interval estimate to zero. Comparative performance for hospitals with 25 or more eligible cases was classified as follows:
    - "Lower than expected" if the entire 95% CI surrounding the hospital's days is below zero.
    - "No different than expected" if the 95% CI surrounding the hospital's days includes zero.
    - "Higher than expected" if the entire 95% CI surrounding the hospital's days is above zero.
- Hospitals with fewer than 25 eligible cases were assigned to a separate category: "The number of cases is too small (fewer than 25) to reliably assess the hospital's EDAC."
- Of 4,286 hospitals in the three-year study cohort, 254 had EDACs "lower than expected," 1,440 were "no different than expected," and 579 had EDACs "higher than expected." 2,013 were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing. The mean EDAC per 100 discharges for hospitals in the top decile of performance is -23.3, compared to 170.4 for hospitals in the bottom decile.
- The developer states that the variation in hospital-level EDAC suggests there are meaningful differences in the quality of care received across hospitals for the AMI EDAC measure.

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
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)

2. Scientific Acceptability of Measure Properties

2a1. & 2b1. Specifications

Comments: **Patient instructions indicate when to present a hospital for observation, ED observation is largely the only source for 24 observation that is clinically supervised. Same day discharges are considered in exclusions. What is the tipping point in days where observations are generally deemed unnecessary? Can this be excluded? The target population will go to the ED assuming appropriate care when following well coordinated care instructions.

**High face validity though seems to be moderately unstable for a hospital from year to year suggesting it is influenced by other forces beyond the quality of hospital care.

2a2. Reliability Testing

Comments: **yes

**The data are highly reliable in measuring acute care, though the measure for different patient sets varies suggesting that it is not a reliable measure of quality.

2b2. Validity Testing

Comments: **The measure is a valid metric for resource use. I am not convinced it is a valid measure of quality.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **The developer does mention that the it uses days rather than rates,and data differences may occur, but does not feel this poses a threat. Although SDS was not incorporated, but rationale was given that the differences, (even significant) that the findings and complex pathways could explain differences. I would like to hear from the developer if observation visits were materially different in the SDS findings.

**Do not believe differences in resource use represent differences in quality.

Criterion 3. Feasibility

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 based on administrative claims data (e.g., DRG, ICD-9/10), which the developers note are routinely generated and collected as part of hospitals' billing processes.
- The developer indicates that all data elements are in defined fields in electronic claims.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*
- *If an eMeasure, does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?*

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3. Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **All are collected in claims

**Highly feasible.

Criterion 4: Usability and Use

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**

OR

Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

- This measure has been finalized for use in CMS's Hospital Inpatient Quality Reporting (IQR) program starting in Fiscal Year (FY) 2018 (80 FR 49690).

Improvement results

- Since this measure is not in use, there are no performance results to assess improvement at this time.
- The developer states that they expect that "there will be improvement in measure scores over time since publicly reported measure scores can reduce adverse patient outcomes associated with days spent in acute care for AMI by capturing and making acute care utilization following the index hospitalization more visible to providers and patients."

Potential harms

- The developer noted that there were no unintended consequences during development, testing or re-specification. They are committed to ongoing monitoring of potential unintended consequences, such as the inappropriate shifting of care, increased patient morbidity and mortality, and other negative intended consequences over time.

Feedback :

- MAP reviewed this measure during its 2014-2015 pre-rulemaking deliberations for consideration in the IQR program. MAP was conditionally supportive of this measure on the condition that this measure is reviewed by NQF and endorsed. In particular, members noted that the measure should be considered for SDS adjustment in the upcoming NQF trial period, reviewed for the empirical and conceptual relationship between SDS factors and risk-standardized days following acute care, and endorsed with appropriate consideration of SDS factors as determined by NQF standing committees. Some MAP members noted this measure could help address concerns about the growing use of observation stays.

***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

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **Public reporting planned. SThe developer states "they are committed to ongoing monitoring of potential unintended consequences, such as the inappropriate shifting of care, increased patient morbidity and mortality, and other negative intended consequences over time.

The inclusion of observation and ED in a unplanned admit is worrisome, for the reasons stated above. If the standard of care for prevention of hospitalization is the effective use of observation visits, then how does the inclusion of observation visits improve care? I do not believe the benefits of the measure outweighs this concern without further understanding.

**Not yet publically reported. Unintended consequences include refusing to admit acutely ill patients which would worsen health status. Not sure if the benefits outweigh the risks

Criterion 5: Related and Competing Measures

Related or competing measures

- 0505: Hospital 30-day all-cause risk-standardized readmission rate (RSRR) following acute myocardial infarction (AMI) hospitalization

Harmonization

• The developers note that both measures are harmonized.

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: Excess days in acute care (EDAC) after hospitalization for acute myocardial infarction (AMI)

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: 1/29/2016

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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: Single measure: quality outcome measure

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

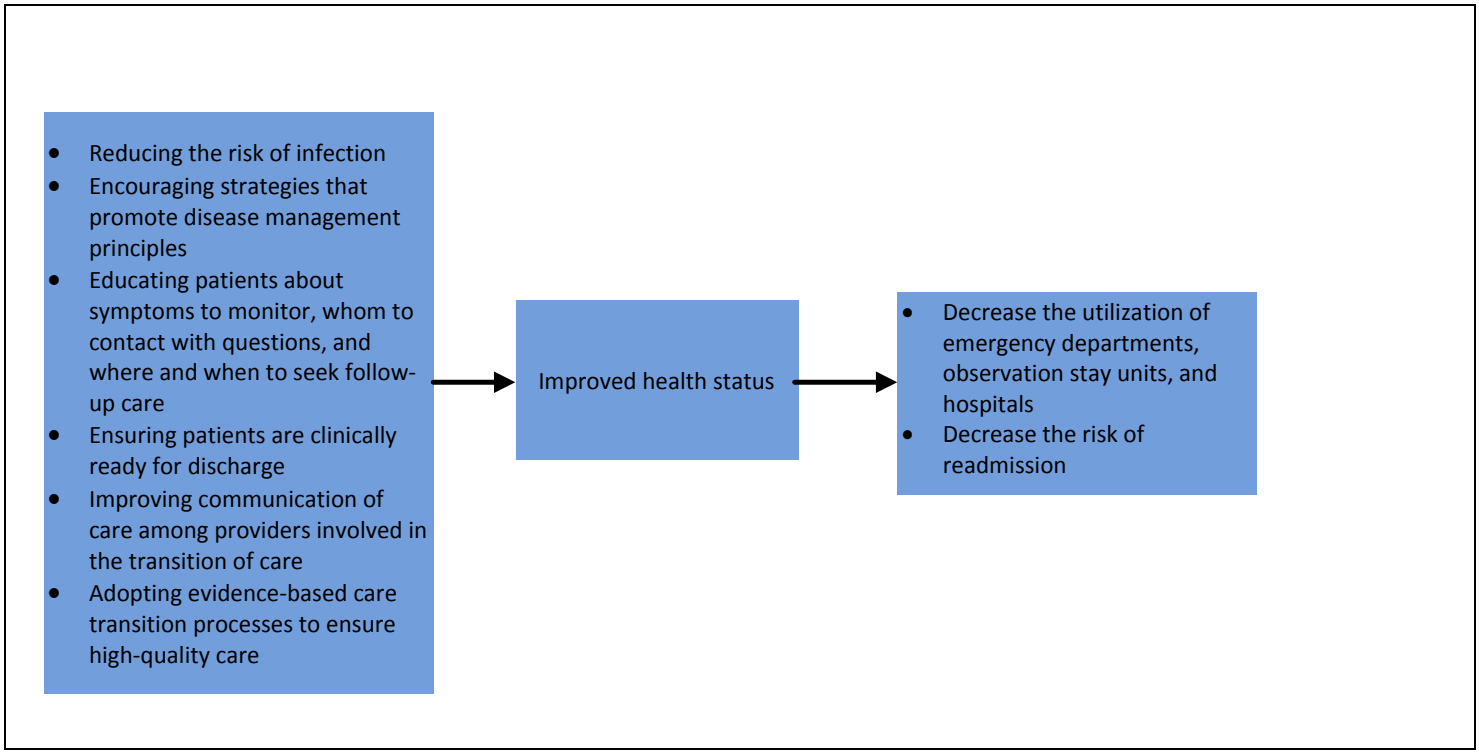
☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.



1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

The diagram above indicates some of the many care processes that can influence post-discharge acute care utilization. These complex and critical aspects of care – such as communication between providers, patient education, patient safety, and coordinated transitions to the outpatient environment – all contribute to patient outcomes but are difficult to measure by individual process measures. Numerous studies have shown that improvement in the following areas can favorably impact utilization rates: communication with patients, caregivers, and other providers; patient education; and quality of care during the initial inpatient admission.

Interventions during and after a hospitalization can be effective in reducing readmission rates in geriatric populations (Benbassat et al., 2000; Naylor et al., 1999; Coleman et al., 2006; Courtney et al., 2009; Koehler et al., 2009) and, particularly, for older patients with AMI (Carroll et al., 2007; Young et al., 2003; Bondestam et al., 1995; Ades et al, 1992; Carlhed et al., 2009). Several randomized trials have reduced 30-day readmission rates by 20-40% (Jack et al., 2009; Coleman et al., 2004; Courtney et al., 2009; Garasen et al., 2007; Koehler et al., 2009; Mistiaen et al., 2007; Naylor et al., 1994; Naylor et al., 1999; van Walraven et al., 2002; Weiss et al., 2010; Krumholz et al., 2012; Balaban et al., 2008). These types of interventions have also been demonstrated to be cost-saving (Naylor et al., 1999; Naylor et al., 2004; Koelling et al., 2005; Krumholz et al., 2002; Stauffer et al., 2011). Outside the randomized controlled trial setting, there is also increasing evidence that hospitals and health plans have been able to reduce readmission rates through more generalizable quality improvement initiatives (Gerhardt et al., 2012; Stauffer et al., 2011; Graham et al., 2012; Harrison et al., 2011; Hernandez et al., 2010).

In the case of AMI, specifically, studies suggest that appropriate care for AMI during and after the index hospitalization may reduce the risk of subsequent readmission (Carroll et al., 2007; Young et al., 2003; Bondestam et al., 1995; Ades et al, 1992; Carlhed et al., 2009). Studies have also reported reductions in emergency department (ED) visit rates for patients with other conditions after implementation of interventions that focused on the inpatient and outpatient settings (Bondestam et al., 1995).

The current process-based performance measures cannot capture all the ways that care within the hospital might influence outcomes. As a result, many stakeholders, including patient organizations, are interested in outcomes measures that allow patients and providers to assess relative outcomes performance among hospitals (Bratzler et al., 2007).

In the context of the Centers for Medicare and Medicaid Services' (CMS's) publicly reported readmission measures, the increasing use of ED visits and observation stays has raised concerns that current readmission measures do not capture the full range of unplanned acute care in the post-discharge period (Vashi et al., 2013; Rising et al., 2012; Feng et al., 2012). Observation stays can occur in many different parts of the hospital, including dedicated treatment rooms, the ED, or inpatient units. In particular, there is concern that high use of observation stays could in some cases replace readmissions, and that hospitals with high rates of observation

stays in the post-discharge period may therefore have low readmission rates that do not accurately reflect the quality of care (Vashi et al., 2013).

Citations:

Ades PA, Huang D, Weaver SO. 1992. Cardiac rehabilitation participation predicts lower rehospitalization costs. Am Heart J 123(4 Pt 1):916-921.

Balaban RB, Weissman JS, Samuel PA, Woolhandler S. Redefining and redesigning hospital discharge to enhance patient care: a randomized controlled study. *J Gen Intern Med.* 2008;23(8):1228-1233.

Benbassat, J., and M. Taragin. 2000. Hospital readmissions as a measure of quality of health care: advantages and limitations. Arch Intern Med 160 (8):1074-81.

Bondestam E, Breikss A, Hartford M. 1995. Effects of early rehabilitation on consumption of medical care during the first year after acute myocardial infarction in patients > or = 65 years of age. Am J Cardiol 75(12):767-771.

Bratzler, DW, Nsa W, Houck PM. Performance measures for pneumonia: are they valuable, and are process measures adequate. Current Opinion in Infectious Diseases. 20(2):182-189, April 2007.

Carlhed R, Bojestig M, Peterson A, et al. Improved clinical outcome after acute myocardial infarction in hospitals participating in a Swedish quality improvement initiative. Circulation. Cardiovascular Quality & Outcomes. 2009;2(5):458-464.

Carroll DL, Rankin SH, Cooper BA. 2007. The effects of a collaborative peer advisor/advanced practice nurse intervention: cardiac rehabilitation participation and rehospitalization in older adults after a cardiac event. J Cardiovasc Nurs 22(4):313-319.

Coleman EA, Parry C, Chalmers S, et al. 2006. The care transitions intervention: results of a randomized controlled trial. Arch Intern Med 166:1822-1828.

Coleman EA, Smith JD, Frank JC, Min SJ, Parry C, Kramer AM. Preparing patients and caregivers to participate in care delivered across settings: the Care Transitions Intervention. J Am Geriatr Soc 2004;52(11):1817-25.

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 2009;57(3):395-402.

Feng Z, Wright B, Mor V. Sharp rise in Medicare enrollees being held in hospitals for observation raises concerns about causes and consequences. Health affairs (Project Hope). Jun 2012;31(6):1251-1259.

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.

Gerhardt G, Yemane A, Hickman P, Oelschlaeger A, Rollins E, Brennan N. Medicare Readmission Rates Showed Meaningful Decline in 2012. *Medicare & Medicaid Research Review.* 2013;3(2):E1-E12.

Graham J, Tomcavage J, Salek D, Sciandra J, Davis DE, Stewart WF. Postdischarge monitoring using interactive voice response system reduces 30-day readmission rates in a case-managed Medicare population. *Medical Care.* 2012;50(1):50-57.

Harrison PL, Hara PA, Pope JE, Young MC, Rula EY. The impact of postdischarge telephonic follow-up on hospital readmissions. *Population Health Management.* 2011;14(1):27-32.

Hernandez AF, Greiner MA, Fonarow GC, et al. Relationship between early physician follow-up and 30-day readmission among Medicare beneficiaries hospitalized for heart failure. *Jama.* 2010;303(17):1716-1722.

Jack BW, Chetty VK, Anthony D, Greenwald JL, Sanchez GM, Johnson AE, et al. A reengineered hospital discharge program to decrease rehospitalization: a randomized trial. Ann Intern Med 2009;150(3):178-87.

Koehler BE, Richter KM, Youngblood L, Cohen BA, Prengler ID, Cheng D, 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. J Hosp Med 2009;4(4):211-218.

Koelling, TM. 2005. Multifaceted outpatient support can improve outcomes for people with heart failure. Commentary. Evid Based Cardiovasc Med 9 Koelling, TM. 2005. Multifaceted outpatient support can improve outcomes for people with heart failure. Commentary. Evid Based Cardiovasc Med 9 (2):138-41.

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.

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

Naylor, MD, Brooten D, Campbell R, et al. 1999. Comprehensive discharge planning and home follow-up of hospitalized elders: a randomized clinical trial. JAMA 281 (7):613-20.

Naylor MD, 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 1994;120(12):999-1006.

Naylor MD, Brooten D, Campbell R, et al. 2004. Transitional care of older adults hospitalized with heart failure: a randomized, controlled trial. J Am Geriatr Soc 52 (5):675-84.

Rising KL, White LF, Fernandez WG, Boutwell AE. Emergency Department Visits After Hospital Discharge: A Missing Part of the Equation. Annals of Emergency Medicine. (0).

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.* 2011;171(14):1238-1243.

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 2002;17(3):186-92.

Vashi AA, Fox JP, Carr BG, et al. Use of hospital-based acute care among patients recently discharged from the hospital. JAMA. Jan 23 2013;309(4):364-371.

Weiss M, Yakusheva O, Bobay K. Nurse and patient perceptions of discharge readiness in relation to postdischarge utilization. Med Care 2010;48(5):482-6.

Young W, Rewa G, Goodman SG, et al. 2003. Evaluation of a community-based inner-city disease management program for postmyocardial infarction patients: a randomized controlled trial. CMAJ Canadian Med Assn J 169(9):905-910.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

N/A. This is an outcome measure.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

N/A. This is an outcome measure.

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

N/A. This is an outcome measure.

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

N/A. This is an outcome measure.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A. This is an outcome measure.

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

N/A. This is an outcome measure.

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

N/A. This is an outcome measure.

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

N/A. This is an outcome measure.

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

N/A. This is an outcome measure.

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

N/A. This is an outcome measure.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A. This is an outcome measure.

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

N/A. This is an outcome measure.

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

N/A. This is an outcome measure.

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

N/A. This is an outcome measure.

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

N/A. This is an outcome measure.

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

N/A. This is an outcome measure.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

N/A. This is an outcome measure.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

N/A. This is an outcome measure.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

N/A. This is an outcome measure.

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

N/A. This is an outcome measure.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

N/A. This is an outcome measure.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

N/A. This is an outcome measure.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

N/A. This is an outcome measure.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

N/A. This is an outcome measure.

1a.8 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.8.1 What process was used to identify the evidence?

N/A. This is an outcome measure.

1a.8.2. Provide the citation and summary for each piece of evidence.

N/A. This is an outcome measure.

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
AMI_Excess_Days_in_Acute_Care_NQF_Measure_Evidence_Form_01_29_16_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, and hospitals with information about hospital-level, risk-standardized outcomes following hospitalization for AMI. Measurement of patient outcomes allows for a broad view of quality of care that cannot be captured entirely by individual process-of-care measures. Safely transitioning patients from hospital to home requires a complex series of tasks which would be cumbersome to capture individually as process measures: timely and effective communication between providers, prevention of and response to complications, patient education about post-discharge care and self-management, timely follow-up, and more. Suboptimal transitions contribute to a variety of adverse events post-discharge, including ED evaluation, need for observation, and readmission. Measures of unplanned readmission already exist, but there are no current measures for ED and observation stay utilization. It is thus difficult for providers and consumers to gain a complete picture of post-discharge outcomes. Moreover, separately reporting each of these outcomes encourages "gaming," such as re-categorizing readmission stays as observation stays to avoid a readmission outcome. By capturing a range of acute care events that are important to patients, we can produce a more complete picture of post-discharge outcomes that better informs consumers about care quality and incentivizes global improvement in transitional care.

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.*
Distribution of EDAC per 100 discharges in the three-year dataset used for measure development:
This analysis includes hospitals that have at least 25 AMI index admissions in the three-year period.

Time period//2010-2013
Number of hospitals//1,855
Number of discharges//232,954
Mean EDAC (standard deviation)//9.27 (28.49)
Range (minimum - maximum)//225.26 (-54.82 - 170.44)
Interquartile range//-10.50– 24.48
Minimum//-54.82
10th percentile//-23.33
20th percentile//-13.94
30th percentile//-7.22
40th percentile//-0.82
50th percentile//5.46
60th percentile//12.24
70th percentile//19.97
80th percentile//29.22
90th percentile//46.05
Maximum//170.44

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

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for 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.*
Distribution of AMI EDAC (per 100 discharges) by Proportion of Dual Eligible Patients:
Dates of Data: July 2010 through June 2013 AMI development dataset
Data Source: Medicare FFS claims
Characteristic//Hospitals with a low proportion (=0%) dual-eligible patients//Hospitals with a high proportion (=21.05%) dual eligible patients
Number of Measured Hospitals//1,082 //1,046
Number of Patients//5,142 patients in low-proportion hospitals/ 24,494 in high-proportion hospitals
Maximum//214.66 //386.10
90th percentile//26.44 //61.03
75th percentile//0.74 //20.35
Median (50th percentile)//-0.68 //-0.30
25th percentile//-12.08//-15.59
10th percentile//-28.08//-31.99
Minimum //-93.19//-97.78

Distribution of AMI EDAC (per 100 discharges) by Proportion of African-American Patients:
Dates of Data: July 2010 through June 2013 AMI development dataset
Data Source: Medicare FFS claims
Characteristic// Hospitals with a low proportion (=0%) African-American patients//Hospitals with a high proportion (=7.32%) African-American patients
Number of Measured Hospitals//2,229 //1,038
Number of Patients//42,537 patients in low-proportion hospitals/ 82,236 in high-proportion hospitals
Maximum//386.1 //322.32
90th percentile//34.63//59.20
75th percentile//3.43//29.37
Median (50th percentile)//-1.02//6.81
25th percentile//-17.35//-9.48
10th percentile//-31.36//-25.12
Minimum//-97.78//-88.77

Low-proportion hospitals are those hospitals whose population of dual-eligible patients or of African-American patients is small enough to place them at or below the 25th percentile among all hospitals; and high proportion are those hospitals whose population of dual eligible patients or African-American patients is large enough to place them at or above the 75th percentile among all hospitals.

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, 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.**
AMI is among the most common principal hospital discharge diagnoses among Medicare beneficiaries, and in 2008, it was the sixth most expensive condition billed to Medicare, accounting for 4.8% of Medicare's hospital bill (Wier and Andrews, 2011). Readmission rates following discharge for AMI are high. For example, between July 2005 and June 2008, the median 30-day readmission rate for AMI was 19.9%, with a range of 15.3% to 29.4% (Krumholz et al., 2009). For the time period of July 2012-June 2013, publicly reported 30-day risk-standardized readmission rates (RSRRs) ranged from 14.1% to 20.6% for patients admitted with AMI (CMS, 2014).

Hospital readmission, for any reason, is disruptive to patients and caregivers, is costly to the healthcare system, and puts patients at additional risk of hospital-acquired infections and complications. Although some readmissions are unavoidable, others may result from poor quality of care or inadequate transitional care. Transitional care includes effective discharge planning, transfer of information at the time of discharge, patient assessment and education, and coordination of care and monitoring in the post-discharge period. Numerous studies have found an association between quality of inpatient or transitional care and early (typically 30-day) readmission rates for a wide range of conditions including AMI (Frankl et al., 1991; Corrigan et al., 1992; Oddone et al., 1996; Ashton et al., 1997; Benbassat et al., 2000; Courtney et al., 2003; Halfon et al., 2006; Bondestam et al., 1995; Carlhed et al., 2009).

Several studies have reported on the relationship between inpatient admissions and other types of hospital care including ED visits and observation stays. ED visits represent a significant proportion of post-discharge acute care utilization. Two recent studies conducted in patients of all ages have shown that 9.5% of patients return to the ED within 30 days of hospital discharge and that about 12% of these patients are discharged from the ED and are not captured by current CMS readmissions measures (Rising et al., 2013; Vashi et al., 2013).

Additionally, over the past decade, the use of observation stays has rapidly increased. Specifically, between 2001 and 2008, the use of observation services increased nearly three-fold (Venkatesh et al., 2011) and significant variation has been demonstrated in the use of observation services for conditions such as chest pain (Schuur et al., 2011). These rising rates of observation stays among Medicare beneficiaries have gained the attention of patients, providers, and policymakers (Feng et al., 2012; Hockenberry et al., 2014; Rising et al., 2013; Vashi et al., 2013, Wright B. et al., 2014). A report from the Office of the Inspector General (OIG) notes that in 2012, Medicare beneficiaries had 1.5 million observation stays. Many of these observation stays lasted longer than the intended one day. The OIG report also notes the potential relationship between hospital use of observation stays as an alternative to short-stay inpatient hospitalizations as a response to changing hospital payment incentives (Wright, 2013).

Thus, in the context of the CMS's publicly reported readmission measures, the increasing use of ED visits and observation stays has raised concerns that current readmission measures do not capture the full range of unplanned acute care in the post-discharge period. By definition, the readmission measures only assess returns to the hospitals for inpatient stays and not for other acute care services, such as observation stays or ED visits. Stakeholders have expressed concerns about whether observation stays should also be evaluated as markers of the quality of care transitions. In particular, there exists concern that high use of observation stays could in some cases replace readmissions, and hospitals with high rates of observation stays in the post-discharge period may therefore have low readmission rates that do not accurately reflect the quality of care (Carlson et al., 2013).

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
Ashton CM, Del Junco DJ, Souchek J, Wray NP, Mansyur CL. The association between the quality of inpatient care and early readmission: a meta-analysis of the evidence. Med Care. Oct 1997;35(10):1044-1059.

Benbassat J, Taragin M. Hospital readmissions as a measure of quality of health care: advantages and limitations. Archives of Internal Medicine. Apr 24 2000;160(8):1074-1081.

Bondestam E, Breikss A, Hartford M. Effects of early rehabilitation on consumption of medical care during the first year after acute myocardial infarction in patients > or = 65 years of age. American Journal of Cardiology. 1995;75(12):767-771.

Carlhed R, Bojestig M, Peterson A, et al. Improved clinical outcome after acute myocardial infarction in hospitals participating in a Swedish quality improvement initiative. Circulation. Cardiovascular Quality & Outcomes. 2009;2(5):458-464.

Carlson J. Faulty Gauge? Readmissions are down, but observational-status patients are up and that could skew Medicare numbers. Modern Healthcare. June 8, 2013.

Centers for Medicare and Medicaid Services (CMS). Medicare Hospital Quality Chartbook Performance Report on Outcome Measures September 2014. September 2014; https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Medicare-Hospital-Quality-Chartbook-2014.pdf.

Corrigan JM, Martin JB. Identification of factors associated with hospital readmission and development of a predictive model. Health Serv Res. Apr 1992;27(1):81-101.

Courtney EDJ, Ankrett S, McCollum PT. 28-Day emergency surgical re-admission rates as a clinical indicator of performance. Ann R Coll Surg Engl. Mar 2003;85(2):75-78.

Feng Z, Wright B, Mor V. Sharp rise in Medicare enrollees being held in hospitals for observation raises concerns about causes and consequences. Health affairs (Project Hope). Jun 2012;31(6):1251-1259.

Frankl SE, Breeling JL, Goldman L. Preventability of emergent hospital readmission. Am J Med. Jun 1991;90(6):667-674.

Halfon P, Eggli Y, Pr, et al. Validation of the potentially avoidable hospital readmission rate as a routine indicator of the quality of hospital care. Medical Care. Nov 2006;44(11):972-981.

Hockenberry JM, Mutter R, Barrett M, Parlato J, Ross MA. Factors Associated with Prolonged Observation Services Stays and the Impact of Long Stays on Patient Cost. Health Serv Res. 2014;49(3):893-909.

Krumholz HM, Merrill AR, Schone EM, Schreiner GC, Chen J, Bradley EH, Wang Y, Wang Y, Lin Z, Straube BM, Rapp MT, Normand SL, Drye EE. 2009. Patterns of hospital performance in acute myocardial infarction and heart failure 30-day mortality and readmission. Circ Cardiovasc Qual Outcomes (2):407-413.
Oddone EZ, Weinberger M, Horner M, et al. Classifying general medicine readmissions. Are they preventable? Veterans Affairs Cooperative Studies in Health Services Group on Primary Care and Hospital Readmissions. Journal of General Internal Medicine. 1996;11(10):597-607.

Rising KL, White LF, Fernandez WG, Boutwell AE. Emergency Department Visits After Hospital Discharge: A Missing Part of the Equation. Annals of Emergency Medicine.

Schuur JD, Baugh CW, Hess EP, Hilton JA, Pines JM, Asplin BR. Critical pathways for post-emergency outpatient diagnosis and treatment: tools to improve the value of emergency care. Academic emergency medicine : official journal of the Society for Academic Emergency Medicine. Jun 2011;18(6):e52-63.

Vashi AA, Fox JP, Carr BG, et al. Use of hospital-based acute care among patients recently discharged from the hospital. JAMA : the journal of the American Medical Association. Jan 23 2013;309(4):364-371.

Venkatesh AK, Geisler BP, Gibson Chambers JJ, Baugh CW, Bohan JS, Schuur JD. Use of observation care in US emergency departments, 2001 to 2008. PloS one. 2011;6(9):e24326.

Wier, L.M. (Thomson Reuters), and Andrews, R.M. (AHRQ).The National Hospital Bill: The Most Expensive Conditions by Payer, 2008. HCUP Statistical Brief #107. March 2011. Agency for Healthcare Research and Quality, Rockville, MD. http://www.hcupus.ahrq.gov/reports/statbriefs/sb107.pdf.

Wright B., Jung H-Y, Feng Z, Mor V. Hospital, Patient, and Local Health System Characteristics Associated with the Prevalence and Duration of Observation Care. Health Serv Res. 2014;49(4):1088-1107.

Wright S. Hospitals' Use of Observation Stays and Short Inpatient Stays for Medicare Beneficiaries. Washington, DC: OIG; 2013.

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

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 : Acute Myocardial Infarction

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Care Coordination : Readmissions, Safety, Safety : Readmissions

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:** AMI_Excess_Days_in_Acute_Care_Measure_NQF_Data_Dictionary_01-29-16_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 of the measure is a count of the number of days the patient spends in acute care within 30 days of discharge. We define days in acute care as days spent in an ED, admitted to an observation unit, or admitted as an unplanned readmission for any cause within 30 days from the date of discharge from the index AMI hospitalization. Each ED treat-and-release visit is counted as one half-day (0.5 days). Observation stays are recorded in terms of hours and are rounded up to the nearest half-day. Each readmission day is counted as one full day (1 day). We count all eligible outcomes occurring in the 30-day period, even if they are repeat occurrences.

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 the measure as within 30 days of the date of discharge of the index AMI hospitalization.

Denominator Time Window: The measure was developed and will be reported with using 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.

Outcome Definition

The measure counts ED treat-and-release visits, observation stays, and readmissions to any acute care hospital for any cause within 30 days of the date of discharge of the index AMI admission, excluding planned readmissions as defined below.

All events which occur within the 30-day window are counted. For example, if a patient returns to the ED three times on three different days, we count each ED visit as a half-day. Similarly, if a patient has two hospitalizations within 30 days, the days spent in each are counted. Therefore, the measure may include multiple ED visits, observation stays, and/or readmissions per patient.

The measure incorporates "exposure time" (the number of days each patient survives after discharge, up to 30). This exposure time is included to account for differential risk for EDAC after discharge among those patients who do not survive the full post-discharge period. If a hospitalization or observation stay extends beyond the 30-day window, only those days within the 30-day window are counted.

Planned Readmission Algorithm

The Planned Readmission Algorithm is a set of criteria for classifying readmissions as planned among the general Medicare population using Medicare administrative claims data. The algorithm identifies admissions that are typically planned and may occur within 30 days of discharge from the hospital.

The Planned Readmission Algorithm has three fundamental principles:
1. A few specific, limited types of care are always considered planned (obstetric delivery, transplant surgery, maintenance chemotherapy/radiotherapy/ immunotherapy, rehabilitation);
2. Otherwise, a planned readmission is defined as a non-acute readmission for a scheduled procedure; and
3. Admissions for acute illness or for complications of care are never planned.

The algorithm was developed in 2011 as part of the Hospital-Wide Readmission measure. In 2013, CMS applied the algorithm to its other readmission measures. In 2013, CMS applied the algorithm to its other readmission measures. In applying the algorithm to condition- and procedure-specific measures, teams of clinical experts reviewed the algorithm in the context of each measure-specific patient cohort and, where clinically indicated, adapted the content of the algorithm to better reflect the likely clinical experience of each measure's patient cohort. For the CMS 30-day AMI EDAC measure, CMS used the Planned Readmission Algorithm without making any changes.

For development, we used the Planned Readmission Algorithm, Version 3.0. This version and associated code tables are attached in data field S.2b (Data Dictionary or Code Table). For reporting purposes, the measure will use the next version of the Planned Readmission Algorithm, Version 4.0, as will be used in the CMS 30-day AMI readmission measure.

Definition of Emergency Department Visit and Observation Stay

We defined ED visits and observation stays using specified billing codes or revenue center codes identified in Medicare hospital outpatient claims and physician carrier claims. The codes that define ED visits and observation stays are in the attached Data Dictionary.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*

The target population for this measure is Medicare FFS beneficiaries aged 65 years and older hospitalized at non-federal acute care hospitals for AMI.

The cohort includes admissions for patients discharged from the hospital with a principal discharge diagnosis of AMI (see codes below in S.9) and with continuous 12 months Medicare enrollment prior to admission. The measure will be publicly reported by CMS

for those patients 65 years and older who are Medicare FFS beneficiaries admitted to non-federal hospitals.

Additional details are provided n S.9 Denominator Details.

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

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)*
To be included in the measure cohort used in public reporting, patients must meet the following inclusion criteria:
1. Having a principal discharge diagnosis of AMI
2. Enrolled in Medicare fee-for-service (FFS) Part A and Part B for the 12 months prior to the date of the admission, and enrolled in Part A during the index admission;
3. Aged 65 or over;
4. Discharged alive from a non-federal short-term acute care hospital; and,
5. Not transferred to another acute care facility.

International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) codes used to define the cohort for the measure are:
410.00 Acute myocardial infarction of anterolateral wall, episode of care unspecified
410.01 Acute myocardial infarction of anterolateral wall, initial episode of care
410.10 Acute myocardial infarction of other anterior wall, episode of care unspecified
410.11 Acute myocardial infarction of other anterior wall, initial episode of care
410.20 Acute myocardial infarction of inferolateral wall, episode of care unspecified
410.21 Acute myocardial infarction of inferolateral wall, initial episode of care
410.30 Acute myocardial infarction of inferoposterior wall, episode of care unspecified
410.31 Acute myocardial infarction of inferoposterior wall, initial episode of care
410.40 Acute myocardial infarction of other inferior wall, episode of care unspecified
410.41 Acute myocardial infarction of other inferior wall, initial episode of care
410.50 Acute myocardial infarction of other lateral wall, episode of care unspecified
410.51 Acute myocardial infarction of other lateral wall, initial episode of care
410.60 True posterior wall infarction, episode of care unspecified
410.61 True posterior wall infarction, initial episode of care
410.70 Subendocardial infarction, episode of care unspecified
410.71 Subendocardial infarction, initial episode of care
410.80 Acute myocardial infarction of other specified sites, episode of care unspecified
410.81 Acute myocardial infarction of other specified sites, initial episode of care
410.90 Acute myocardial infarction of unspecified site, episode of care unspecified
410.91 Acute myocardial infarction of unspecified site, initial episode of care

An ICD-9 to ICD-10 crosswalk is attached in field S.2b. (Data Dictionary or Code Table).

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
The measure excludes index admissions for patients:
1. Without at least 30 days post-discharge enrollment in FFS Medicare;
2. Discharged against medical advice (AMA);
3. Admitted within 30 days of a prior index discharge;
4. Admitted and then discharged on the same day (because it is unlikely these are clinically significant AMIs).

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)*
1. Admissions without at least 30 days post-discharge enrollment in FFS Medicare are determined by examining the Medicare Enrollment Database (EDB).

2. Discharges against medical advice (AMA) are identified using the discharge disposition indicator in claims data.
3. Admissions within 30 days of discharge from a qualifying index admission are identified by comparing the discharge date from the index admission with subsequent admission dates.
4. Index admissions for patients admitted and then discharged on the same day are identified when the admission and discharge dates are equal.

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. This measure is not stratified.

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).

For risk-adjustment, we used a hierarchical generalized linear model (HGLM). This model consists of two parts, a logit model and a truncated Poisson model. The two-part logit/Poisson model (often called a "hurdle" model) assumes that the outcome results from two related processes: an initial dichotomous event – that a patient has at least one acute care event – which is modeled as the logit of the probability of the event, and for patients with an event (those which clear the "hurdle"), the number of days, which is modeled as a Poisson process. The outcome, number of days, is a half-integer count variable (because ED visits count as 0.5 days). Observation care is counted according to the hours spent in observation care, rounded up to the nearest half-day. For each patient, an exposure variable is defined as the number of survival days post discharge, up to 30. For the hurdle model, exposure time as an offset is included for each part of the model.

There are two random effects for each hospital, one for the logit model and one for the truncated Poisson model, as well as a covariance between the two random effects. The random effects allow us to account for within-hospital correlation of the observed outcome and accommodates the assumption that underlying differences in quality across hospitals lead to systematic differences in outcomes.

We use the existing, NQF-endorsed, CMS 30-day AMI readmission measure final risk-adjustment variables. We verified the adequacy of this risk-adjustment strategy for our new outcome by comparing the discrimination of models with a full set of all comorbidities to the more parsimonious existing risk models. We found no improvement in model discrimination with the full set, indicating that the existing risk models are adequate.

The measures adjust for variables (i.e., age, comorbid diseases, and indicators of patient frailty) that are clinically relevant and have strong relationships with the outcome. For each patient, risk-adjustment variables are obtained from inpatient, outpatient, and physician Medicare administrative claims data extending 12 months prior to, and including, the index admission.

The model adjusts for case-mix differences based on the clinical status of patients 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 (Pope et al., 2000). A file that contains a list of the ICD-9-CM codes and their groupings into CCs is attached in data field S.2b (Data Dictionary or Code Table). In addition, only comorbidities that convey information about the patient at admission or in the 12 months prior, and not complications that arise during the course of the index hospitalization, are included in the risk adjustment. Hence, we do not risk adjust for CCs that may represent adverse events of care and that are only recorded in the index admission.

The final set of risk-adjustment variables includes the following:
Demographics:
1. Male

2. Age (defined as "Age-65" [years above 65, continuous])

Comorbidities:
3. Diabetes mellitus (DM) and DM complications (CC 15-20, 119-120)
4. Iron deficiency and other anemias and blood disease (CC 47)
5. Congestive heart failure (CC 80)
6. Valvular and rheumatic heart disease (CC 86)
7. COPD (CC108)
8. End-stage renal disease or dialysis (CC130)
9. Other urinary tract disorders (CC136)
10. Arrhythmias (CC 92-93)
11. Pneumonia (CC 111-113)
12. Renal failure (CC 131)
13. Vascular or circulatory disease (CC 104-106)
14. Disorders of fluid/electrolyte/acid-base (CC 22-23)
15. Coronary atherosclerosis/other chronic ischemic heart disease (CC 84)
16. History of infection (CC 1,3-6)
17. Cerebrovascular disease (CC 97-99,103)
18. Metastatic cancer and acute leukemia (CC 7)
19. Cancer (CC 8-12)
20. Decubitus ulcer or chronic skin ulcer (CC 148-149)
21. Dementia and other specified brain disorders (senility)(CC 49-50)
22. Angina pectoris, old myocardial infarction (CC 83)
23. Stroke (CC 95-96)
24. Asthma (CC 110)
25. Acute coronary syndrome (CC 81-82)
26. Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69,100-102,177-178)
27. Protein-calorie malnutrition (CC 21)
28. Anterior myocardial infarction (ICD-9-CM 410.00-410.19)
29. Other location of myocardial infarction (ICD-9-CM 410.20-410.69)
30. History of CABG (ICD-9-CM V45.81, 36.10-36.16)
31. History of PTCA (ICD-9-CM V45.82, 00.66, 36.01, 36.02, 36.05, 36.06, 36.07)

References:
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.

Pope GC, et al. 2000. Principal Inpatient Diagnostic Cost Group Models for Medicare Risk Adjustment. Health Care Financing Review 21(3): 93-118.

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*)

S.16. Type of score:
Other (specify):
If other: Excess days in acute care (EDAC) per 100 discharges

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.)*
As described above, we used a hierarchical generalized linear model (HGLM). This consists of the two-part logit/truncated Poisson model specifications for days in acute care and includes two random effects for hospitals – one for the logit part and one for the truncated Poisson part – with a non-zero covariance between the two random effects.

This model is used to estimate predicted and expected values for each patient. Predicted values are model predictions that include the hospital random effects, and expected values are model predictions that do not include the hospital random effects. We describe calculation of the predicted and expected values in the attached Appendix (Section 2.7). The measure reports, for each hospital, the difference ("excess") between each hospital's patients' average days in acute care ("predicted days"), and the number of days in acute care that they would have been expected to spend if discharged from an average performing hospital ("expected days"). To be consistent with the reporting of the CMS 30-day AMI readmission measure, we have multiplied the final score by 100 so that the reported EDAC represents EDAC per 100 discharges.

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.
Administrative claims

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:
1. Medicare Part A inpatient claims, Part B hospital outpatient claims, and physician carrier claims data: 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.

For development purposes, we obtained the Medicare Part B hospital and physician outpatient claims from the Chronic Condition Data Warehouse (CCW) 100% condition-specific datasets.

2. Medicare Enrollment Database (EDB): This database contains Medicare beneficiary demographic, benefit/coverage, and vital status information. This data source was used to obtain information on several inclusion/exclusion indicators such as Medicare

status on admission as well as vital status. These data have previously been shown to accurately reflect patient vital status (Fleming et al., 1992).

Reference:
Fleming C, Fisher ES, Chang CH, Bubolz TA, Malenka DJ. 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/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

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
AMI_Excess_Days_in_Acute_Care_NQF_Measure_Testing_Form_01-29-16_v1.1.docx

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

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Excess days in acute care (EDAC) after hospitalization for acute myocardial infarction (AMI)
Date of Submission: 1/29/2016
Type of Measure:

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

The datasets used for testing included Medicare Parts A and B claims, the Medicare Enrollment Database (EDB), and the Chronic Condition Data Warehouse (CCW) 100% condition-specific dataset to capture emergency department (ED) visits and observation stays.

The specific 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 July 1, 2010 through June 30, 2013.

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

Measure Specified to Measure Performance of: **(*must be consistent with levels entered in item S.26*)**	**Measure Tested at Level of:**

☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

For this measure, hospitals are the measured entities. All non-Federal, acute inpatient hospitals in the United States ([US] including territories) with Medicare Fee-for-Service (FFS) beneficiaries over the age of 65 are included. See section 1.7 for details

1.6. How many and which patients 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 patients and discharges varies by testing type and samples used. See Section 1.7 for the uses of the development sample and validation sample.

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 hospitals, and number of admissions used in each type of testing are as follows:

For reliability testing (Section 2a2):
The reliability of the model was tested by randomly selecting 50% of the Medicare patients aged 65 years or older in a three-year cohort (July 1, 2010-June 30, 2013) and developing a risk-adjusted model for this group (the "development sample"). We then developed a second model for the remaining 50% of patients (the "validation sample") and compared the two.

The development sample consisted of:
Number of discharges: 248,358
Number of hospitals: 4,163
Patient descriptive characteristics: average (standard deviation [SD]) age = 78.9 (8.3); % male = 50.6%

The validation sample consisted of:

Number of discharges: 248,358
Number of hospitals: 4,176
Patient descriptive characteristics: average (SD) age = 78.9 (8.3); % male = 50.4%

We used the three-year dataset for testing of measure exclusions (Section 2b3).

We used the development sample for calculation of performance score (Section 1b2), model selection (2b4), testing of disparities (Section 1b4), reliability testing (Section 2a2), empirical validity testing (Section 2b2), testing of measure risk adjustment (Section 2b4), and testing to identify meaningful differences in performance (Section 2b5). We also used the development sample to examine 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 (Section 2b4.4b).

We used the validation sample for testing of measure risk adjustment (Section 2b4), and data element and performance measure reliability (Section 2a2).

Data Elements:
• African-American race and dual-eligible status (i.e., enrolled in both Medicare and Medicaid) patient-level data are obtained from Centers for Medicare and Medicaid Services (CMS) enrollment 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 (SDS) 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 (Gilman et al., 2014; Hu et al., 2014; Joynt and Jha, 2013). Income, education, and occupational level are the most commonly examined variables. 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 AMI is more limited, here too studies indicate an association between SES/race and increased risk of AMI readmission (Bernheim et al., 2007; Damiani et al., 2015; Herrin et al., 2015; Joynt, Orav, and Jha 2011; Lindenauer et al., 2013). The presumed 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
• African-American race

In selecting variables, our intent was to be responsive to the National Quality Forum (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.

References

Bernheim SM, Spertus JA, Reid KJ, et al. Socioeconomic disparities in outcomes after acute myocardial infarction. American heart journal. 2007;153(2):313-319.

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.

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.

Herrin J, St Andre J, Kenward K, Joshi MS, Audet AM, Hines SC. Community factors and hospital readmission rates. Health services research. 2015;50(1):20-39.

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

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.

Lindenauer PK, Lagu T, Rothberg MB, et al. Income inequality and 30 day outcomes after acute myocardial infarction, heart failure, and pneumonia: retrospective cohort study. BMJ (Clinical research ed.). 2013;346:f521.

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 aimed to utilize only those data elements from the claims that have both face validity and reliability. We used the final risk-adjustment variables in the existing, NQF-endorsed measure of hospital-level risk-standardized readmission rates following AMI (NQF #0505).

We avoided the use of fields that are thought to be coded inconsistently across facilities. Specifically, we used fields that are consequential for payment and which are audited. We identified such variables through empiric analyses and our understanding of the CMS auditing and billing policies. We sought to avoid variables which do not meet these standards.

In addition, CMS has in place several hospital auditing programs used to assess overall accuracy of claims-based coding, to ensure appropriate billing, and for overpayment recoupment. CMS routinely conducts data analysis to identify potential problem areas and to 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 assessed the reliability of the data elements by comparing variable frequencies between our development sample and validation sample.

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, is measured again using a second

random subset exclusive of the first, and then the agreement between the two resulting performance measures across hospitals is calculated (Rousson et al., 2002).

For test-retest reliability, we calculated the EDAC for each hospital using first the development sample, then the validation sample. Thus, we measured each hospital twice, each time 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) as defined by ICC[2,1] by Shrout and Fleiss (1979) and assessed the values according to conventional standards (Landis and Koch, 1977). We restricted this calculation to hospitals with at least 12 discharges in both samples to approximate the set of hospitals that would have at least 24 discharges over three years and are thus likely to be included in public reporting.

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. In addition, using our split sample datasets underestimates the test-retest reliability that would be achieved if the measure were reported using three years of data, because the smaller samples for each hospital in one year of data are less reliable. To correct this problem, we used the Spearman-Brown prophecy formula (Spearman 1910, Brown 1910) to adjust the ICC[2,1] to represent three years of data.

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

Risk variable	Development sample (N=248,358)		Validation sample (N=248,358)	
	n	%	n	%
Age, continuous (mean [SD])	78.9 (8.3)		78.9 (8.3)	
Male	125,742	50.6	125,274	50.4

Risk variable	Development sample (N=248,358)		Validation sample (N=248,358)	
	n	%	n	%
History of Coronary Artery Bypass Graft (CABG) surgery (ICD-9 codes V45.81, 36.10-36.16)	27,375	11.0	27,420	11.0
History of Percutaneous Transluminal Coronary Angioplasty (PTCA) (ICD-9 codes V45.82, 00.66, 36.06, 36.07)	40,889	16.5	40,461	16.3
Angina pectoris/old myocardial infarction (CC 83)	68,310	27.5	68,521	27.6
Congestive heart failure (CC 80)	81,338	32.8	81,389	32.8
Coronary atherosclerosis (CC 84)	212,635	85.6	212,363	85.6
Acute coronary syndrome (CC 81-82)	56,101	22.6	56,191	22.6
Specified arrhythmias and other heart rhythm disorders (CC 92- 93)	88,925	35.8	88,574	35.7
Valvular or rheumatic heart disease (CC 86)	78,943	31.8	78,402	31.6
Cerebrovascular disease (CC 97-99, 103)	52,767	21.3	52,904	21.3
Stroke (CC 95-96)	18,727	7.5	18,521	7.5
Vascular or circulatory disease (CC 104-106)	90,995	36.6	90,988	36.6
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)	15,976	6.4	15,977	6.4
Diabetes mellitus (DM) or DM complications (CC 15-20, 119-120)	116,027	46.7	116,364	46.9
Renal failure (CC 131)	66,781	26.9	67,028	27.0
End-stage renal disease or dialysis (CC 129-130)	8,021	3.2	7,920	3.2
Other urinary tract disorders (CC 136)	54,899	22.1	54,905	22.1
Chronic obstructive pulmonary disease (COPD) (CC 108)	76,222	30.7	76,037	30.6
Pneumonia (CC 111-113)	57,962	23.3	58,263	23.5
Asthma (CC 110)	16,797	6.8	16,894	6.8
Disorders of fluid/electrolyte/acid-base (CC 22-23)	70,748	28.5	70,975	28.6
History of infection (CC 1, 3-6)	67,351	27.1	67,463	27.2
Metastatic cancer or acute leukemia (CC 7)	5,135	2.1	5,034	2.0

Risk variable	Development sample (N=248,358)		Validation sample (N=248,358)	
	n	%	n	%
Cancer (CC 8-12)	47,525	19.1	46,948	18.9
Iron deficiency or other unspecified anemias and blood disease (CC 47)	117,321	47.2	117,241	47.2
Decubitus ulcer or chronic skin ulcer (CC 148-149)	19,811	8.0	19,738	8.0
Dementia or other specified brain disorders (CC 49-50)	48,891	19.7	49,244	19.8
Protein-calorie malnutrition (CC 21)	15,435	6.2	15,301	6.2
Anterior myocardial infarction (ICD-9 410.00-410.19)	18,146	7.3	18,288	7.4
Other location of myocardial infarction (ICD-9 410.20-410.69)	27,825	11.2	27,646	11.1

Measure Score Reliability Results

The agreement between the two EDAC values for each hospital was estimated for three years to be ICC[2,1] = 0.54, which according to the conventional interpretation is "moderate" (Landis & Koch, 1977).

Reference

Landis J, Koch G. The measurement of observer agreement for categorical data, Biometric. 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 results are consistent with existing hospital-level measures of patient outcomes. Compared to the development sample, the mean age of patients and the frequencies of the risk-adjustment variables were very similar in the validation sample; this indicates that the data elements are reliable. The ICC [2,1] score of 0.54, estimated for three years of data, demonstrates moderate agreement between samples across the full range of measure values. We interpret this to mean that when used with a full three years of data, the measure will be reliable by the standards of hospital measurement.

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)*

We demonstrated measure validity through relevant prior validity testing that we conducted for other claims-based measures, through use of established measure development guidelines, through assessment by external groups, and through systematic assessment of measure face validity by Technical Expert Panel (TEP) of national experts and stakeholder organizations.

Empirical Validity Testing

This measure is closely related in design to the existing, NQF-endorsed readmission measure for patients with AMI. While the current measure includes additional endpoints and measures them in a different metric (days rather than rates), we would expect that hospitals would have similar – though not identical – performance rankings on the two measures. Thus, as one assessment of validity, we compared the rankings of all hospitals using the two measures to assess the consistency of hospital performance on closely related outcomes. We calculated the Pearson correlation, and graphed the readmission measure against the EDAC measure to determine if there were outliers.

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 six NQF-endorsed measures currently in public reporting (AMI, heart failure, 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) (Krumholz et al. 2006; Keenan et al. 2008), AMI patients (Cooperative Cardiovascular Project data) (Krumholz, Wang, et al. 2006), and pneumonia patients (National Pneumonia Project dataset) (Bratzler et al. 2011). 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, supporting the use of the claims-based models for public reporting. This measure uses the same risk-adjustment variables that were previously validated in the chart review studies.

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 (National Quality Forum, 2010), CMS Measures Management System guidance, and the guidance articulated in the American Heart Association scientific statement, “Standards for

Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz, Brindis, et al. 2006).

Validity as Assessed by External Groups

Throughout measure development, we obtained expert and stakeholder input via three mechanisms in the initial, early phase of development: a discussion with an advisory Methodology Workgroup, discussions with a national TEP, and a 30-day public comment period in order to increase transparency and to gain broader input on the measure.

The Methodology Workgroup meeting addressed key issues related to measure methodology, including weighing the pros and cons of and measure specifications, modeling, and use (e.g., defining the measure cohort and outcome) to ensure the measure is meaningful, useful, and well-designed. The group provided a forum for focused expert review and discussion of technical issues during measure development.

List of Methodology Workgroup Members:
1) Arlene Ash, PhD; University of Massachusetts Medical School (Professor and Division Chief)
2) Jeremiah Brown, MS, PhD; The Dartmouth Institute for Health Policy and Clinical Practice (Assistant Professor of Health Policy and Clinical Practice)
3) Grant Ritter, PhD, MS, MA; Schneider Institute for Health Policy & Heller Graduate School (Senior Scientist)
4) Patrick Romano, MD, MPH; University of California Davis School of Medicine (Professor of Medicine and Pediatrics)

In alignment with the CMS MMS, 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 physicians, consumers, purchasers, as well as individuals with experience in quality improvement, performance measurement, and health care disparities. We held two structured TEP conference calls consisting of a presentation of key issues, our proposed approach, and relevant data, followed by open discussion among TEP members. We solicited additional input and comments from the TEP via e-mail between meetings.

Following completion of the preliminary model, we solicited public comment on the measure through the CMS site link http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/CallforPublicComment.html. The public comments were then posted publicly for 30 days. The resulting input was taken into consideration during the final stages of measure development, and led to supplementary analyses reported in the application (1b.4).

Face Validity as Determined by Technical Expert Panel

One means of confirming the validity of this measure was face validity assessed by our TEP, which included 16 members, including patient representatives, expert clinicians, researchers, providers, and purchasers.

List of TEP members:

1) Kevin E. Driesen, PhD, MPH, MA; Center for Rural Health Mel and Enid Zuckerman College of Public Health, University of Arizona (Assistant Professor & Director of the Arizona Rural Hospital Flexibility Program)
2) David Engler, PhD; America's Essential Hospitals (Senior Vice President for Leadership and Innovation)
3) Timothy Farrell, MD; University of Utah School of Medicine (Assistant Professor of Medicine, Geriatrics; Adjunct Professor of Family Medicine)
4) Karen Farris, PhD; University of Michigan College of Pharmacy (Charles R. Walgreen III Professor of Pharmacy Administration; Director of the Social and Administrative Pharmacy Graduate Program)
5) Maura C. Feldman, MSW; Blue Cross Blue Shield of Massachusetts, Inc. (Director for Hospital Performance Measurement and Improvement)
6) Jay A. Gold, MD, JD, MPH; Meta Star, Inc. (Vice President & Chief Medical Officer)
7) Sally Hinkle, DNP, MPA, RN; Temple University Hospital (Director of Performance Improvement & Clinical Value)
8) Amy J.H. Kind, MD, PhD; University of Wisconsin School of Medicine and Public Health (Assistant Professor of Geriatrics)
9) Marjorie King, MD, FACC, MAACVPR; Helen Hayes Hospital (Director of Cardiac Services)
10) Eugene Kroch, PhD; University of Pennsylvania (Adjunct Faculty at the Health Care Systems Department); Premier, Inc. (Vice President & Chief Scientist) University of Pennsylvania; Philadelphia, PA
11) Keith D. Lind, JD, MS, BSN; American Association of Retired Persons (AARP) Public Policy Institute (Senior Policy Advisor)
12) Grace McConnell, PhD; Patient representative
13) Michael A. Ross, MD, FACEP; Emory University School of Medicine (Medical Director of Observation Medicine and Chest Pain Center; Professor of Emergency Medicine)
14) Mark Louis Sanz, MDl; International Heart Institute of Montana (Interventional Cardiologist)
15) Paul Takahashi, MD; Mayo Clinic College of Medicine (Associate Professor of Medicine)
16) Patient representative

We systematically assessed the face validity of the measure score as an indicator of quality by soliciting the TEP members' agreement with the following statement: "*The risk-standardized acute care days obtained from the measures as specified can be used to distinguish between better and worse quality hospitals."*

We measured agreement on a six-point scale: 1=Strongly disagree, 2=Moderately disagree, 3=Somewhat disagree, 4=Somewhat agree, 5=Moderately agree, 6=Strongly agree.

<u>Process Used to Identify International Classification of Diseases, Tenth Revision (ICD-10) Codes Statement of Intent</u>
[X] Goal was to convert this measure to a new code set, fully consistent with the intent of the original measure.

[ ] Goal was to take advantage of the more specific code set to form a new version of the measure, but fully consistent with the original intent.
[ ] The intent of the measure has changed.

Process of Conversion
ICD-10 codes were initially identified using 2013 General Equivalence Mapping (GEM) 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. An ICD-9 to ICD-10 crosswalk is attached in field S.2b. (Data Dictionary or Code Table).

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

Krumholz HM, Wang Y, Mattera JA, et al. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with heart failure. *Circulation* 2006;113(13):1693-1701.

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

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

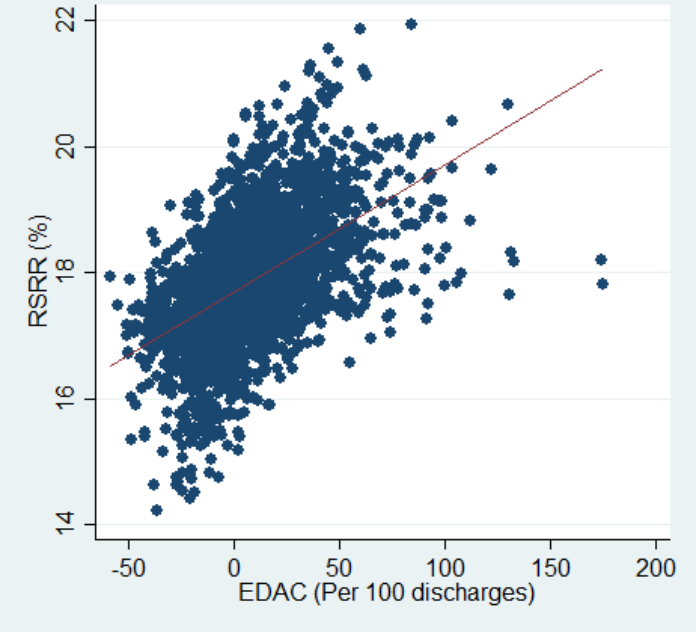
National Quality Forum. National voluntary consensus standards for patient outcomes, first report for phases 1 and 2: A consensus report http://www.qualityforum.org/projects/Patient_Outcome_Measures_Phases1-2.aspx. Accessed August 19, 2010.

Krumholz HM, Brindis RG,Brush JE, et al. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. *Circulation.* January 24, 2006 2006;113(3):456-462.

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

Empirical Validity Testing

Comparison of the new measure with the existing CMS 30-day AMI readmission measure found a Pearson correlation of 0.610 ($P < 0.0001$). The following figure shows the relationship between risk-standardized readmission rate (RSRR) and EDAC for AMI:



Systematic Assessment of Face Validity

The results of the TEP rating of agreement with the validity statement were as follows:

N=12

Mean rating=5

Rating	# of Responses	Percent (%)	Cumulative Percent (%)
6 (Strongly agree)	4	33.3%	33.3%
5 (Moderately agree)	6	50.0%	83.3%
4 (Somewhat agree)	1	8.3%	91.7%
3 (Somewhat disagree)	0	0.0%	91.7%

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?*)

Empirical Validity Testing

There was substantial correlation between the two hospital measures, indicating that the proposed measure and the existing readmission measure share underlying properties. This result, and the notable lack of outliers in the figure, provide external empirical validity.

Validity as Assessed by External Groups
The face validity testing results demonstrated TEP agreement with overall face validity of the measure as specified.

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*)

All exclusions were determined by careful clinical review and have been made based on clinically relevant considerations. To ascertain impact of exclusions on the cohort, we examined overall frequencies and proportions of the total cohort excluded for all exclusions, and examined distributions 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. Rationales for the exclusions are detailed in data field S.10 (Denominator Exclusions).

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*)

We examined overall frequencies and proportions of the admissions excluded for each criterion in all AMI admissions from July 1, 2010 to June 30, 2013.

The exclusion categories are not mutually exclusive.
1. Discharged patients without at least 30 days post-discharge information (0.62%)
2. Discharges against medical advice (AMA) (0.47%)
3. Admissions within 30 days of a prior index admission (1.74%)
4. Same-day discharges (0.47%)

Measure Exclusions

Exclusion	**N**	**%**	**Distribution across hospitals with ≥ 25 discharges (N=2,297): Minimum, 25th percentile, 50th percentile, 75th percentile, maximum**
1. Without at least 30 days post-discharge enrollment in FFS Medicare for index admissions	3,169	0.62	(0.0, 0.0, 0.4, 1.0, 8.1)
2. Discharged against medical	2,393	0.47	(0.0, 0.0, 0.0, 0.8, 8.3)

advice (AMA)			
3. Admissions within 30 days of a prior index admission	8,907	1.74	(0.0, 0.8, 1.6, 2.6, 11.9)
4. Same-day discharges	2,429	0.47	(0.0, 0.0, 0.0, 0.8, 10.7)

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*)

Exclusion 1 (patients without at least 30 days of post-discharge enrollment in FFS Medicare for index admissions) accounts for 0.47% of all index admissions excluded from the initial cohort. This exclusion is needed since the outcome cannot be assessed in this group since claims data are used to determine whether a patient returned to the hospital for an ED visit, was placed under observation care, or was readmitted. Because a very small percent of patients are excluded, this exclusion is unlikely to affect measure score.

Exclusion 2 (patients who are discharged AMA) accounts for 0.62% of all index admissions excluded from the initial index cohort. This exclusion is needed for acceptability of the measure to hospitals, who do not have the opportunity to adequately deliver full care and prepare the patient for discharge. Because a very small percent of patients are excluded, this exclusion is unlikely to affect measure score.

Exclusion 3 (patients with admission within 30 days of a prior index admission) accounts for 1.74% of all index admissions excluded from the initial index cohort. This exclusion is needed to prevent admissions from being counted as both an index admission and a readmission, consistent with the approach taken in the AMI readmission measure.

Exclusion 4 (same-day discharges) accounts for 0.47% of the cohort. The exclusion is meant to ensure a clinically coherent cohort. This exclusion prevents the inclusion of patients who likely did not suffer a clinically significant AMI. For most hospitals this results in very few patients being excluded. For those hospitals with greater proportions of excluded patients, the measure is likely excluding less severe patients that may not be considered as AMI at other hospitals. This exclusion was guided by the input of clinical experts at time of measure development.

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 31 risk factors**

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

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

2b4.2. If an outcome or resource use measure is <u>not risk adjusted or stratified</u>, provide <u>rationale and analyses</u> 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 is tailored to, and appropriate for, a publicly reported outcome measure as articulated in published scientific guidelines (Krumholz et al. 2006, Normand et al. 2007). We adopted the risk factors from the existing NQF-endorsed CMS 30-day AMI readmission measure (Dorsey et al. 2015). These risk factors are comprised of age, sex, and condition categories (CCs) for prior 12-month and current claims. These risk factors had been systematically chosen as predictors of any readmission for the same patient cohort as the current measure; the outcome of this measure is dominated by the number of days of a readmission, so we judged it unlikely that repeating the original analysis would produce different results. We confirmed that there were no additional risk factors to consider by comparing the model estimated using the a priori set of risk factors to a model which included all additional CCs.

For risk adjustment, we used a hierarchical generalized linear model (HGLM). The model consists of two parts, a logit model and a truncated Poisson model. The two-part logit/Poisson model (often called a "hurdle" model) assumes that the outcome results from two related processes: an initial dichotomous event – that a patient has at least one acute care event – which is modeled as the logit of the probability of the event, and for patients with an event (those which clear the "hurdle"), the number of days, which is modeled as a Poisson process. The outcome, number of days, is a half-integer count variable (because ED visits count as 0.5 days). Observation care is counted according to the hours spent in observation care, rounded up to the nearest half-day. For each patient, an exposure variable is defined as the number of survival days post discharge, up to 30. For the hurdle model, exposure time as an offset is included for each part of the model.

There are two random effects for each hospital, one for the logit model and one for the truncated Poisson model, as well as a covariance between the two random effects. The random effects allow us to account for within-hospital correlation of the observed outcome and accommodates the assumption that underlying differences in quality across hospitals lead to systematic differences in outcomes.

Socioeconomic Status Factors and Race

We selected variables representing 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 days in acute care in the 30-days after discharge.

Our conceptualization of the pathways by which patient SES or race affects days in acute care in the 30-days is informed by the literature on the association of SES and race with AMI readmissions, since the majority of the EDAC outcome is composed of readmission days, and since there is a much more robust literature about readmission than about observation care and ED visits.

Literature Review of Socioeconomic Status and Race Variables and AMI Excess Days in Acute Care

To examine the relationship between SES and race variables and hospital 30-day, all-cause EDAC following AMI hospitalization, 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 AMI readmission. Twenty-one studies were initially reviewed, and 16 studies were excluded from full-text review based on the above criteria. Studies indicated that SES/race variables were associated with increased risk of AMI readmission (Bernheim et al., 2007; Damiani et al., 2015; Herrin et al., 2015; Joynt, Orav, and Jha 2011; Lindenauer et al., 2013).

Causal Pathways for Socioeconomic Status 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 in 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 sufficient 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

We assessed the relationship between the dual-eligible status and race variables with the outcome and examined the incremental effect of each 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 Xij be a binary indicator of the SES or race status of the ith patient at the jth hospital, and Xj the percent of patients at hospital j with Xij = 1. Then we rewrote Xij = (Xij- Xj) + Xj □ Xpatient+ Xhospital. The first variable, Xpatient, represents the effect of the risk factor at the patient level (sometimes called the "within" hospital effect), and the second, Xhospital, 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.

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:

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.

Bernheim SM, Spertus JA, Reid KJ, et al. Socioeconomic disparities in outcomes after acute myocardial infarction. American heart journal. 2007;153(2):313-319.

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.

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.

Dorsey KB GJ, Desai N, Lindenauer P, Young J, Wang C, DeBuhr, Parisi ML, Bernheim SM, Krumholz HM. 2015 Condition-Specific Measures Updates and Specifications Report Hospital-Level 30-Day Risk-Standardized Readmission Measures: AMI-Version 8.0, HF-Version 8.0, Pneumonia-Version 8.0, COPD-Version 4.0, and Stroke-Version 4.0. 2015; https://www.qualitynet.org/dcs/BlobServer?blobkey=id&blobnocache=true&blobwhere=1228890435217&blobheader=multipart%2Foctet-stream&blobheadername1=Content-

Disposition&blobheadervalue1=attachment%3Bfilename%3DRdmn_AMIHFPNCOPDSTK_Msr_UpdtRpt.pdf&blobcol=urldata&blobtable=MungoBlobs. Accessed July 8, 2015.

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.

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.

Herrin J, St Andre J, Kenward K, Joshi MS, Audet AM, Hines SC. Community factors and hospital readmission rates. Health services research. 2015;50(1):20-39.

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. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. *Circulation*. January 24, 2006 2006;113(3):456-462.

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.

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

Below is a table showing the final variables in the model with associated parameter estimates.

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

Risk variable	Part 1: Logit model		Part 2: Poisson model	
	Estimate	CI	Estimate	CI
Age minus 65 (years above 65, continuous)	0.009	(0.008, 0.010)	0.004	(0.004, 0.005)
Male	-0.088	(-0.105, -0.071)	-0.003	(-0.012, 0.004)
History of Coronary Artery Bypass Graft (CABG) surgery (ICD-9 codes V45.81, 36.10-36.16)	-0.014	(-0.042, 0.016)	-0.039	(-0.051, -0.026)
History of Percutaneous Transluminal Coronary Angioplasty (PTCA) (ICD-9 codes V45.82, 00.66, 36.06, 36.07)	-0.107	(-0.131, -0.082)	-0.041	(-0.053, -0.030)
Angina pectoris/old myocardial infarction (CC 83)	0.038	(0.014, 0.058)	-0.040	(-0.051, -0.031)
Congestive heart failure (CC 80)	0.140	(0.116, 0.162)	0.095	(0.086, 0.104)
Coronary atherosclerosis (CC 84)	0.010	(-0.016, 0.033)	-0.046	(-0.058, -0.032)
Acute coronary syndrome (CC 81-82)	0.024	(0.005, 0.047)	-0.035	(-0.045, -0.026)
Specified arrhythmias and other heart rhythm disorders (CC 92- 93)	0.099	(0.079, 0.125)	-0.014	(-0.023, -0.004)

Risk variable	Part 1: Logit model		Part 2: Poisson model	
	Estimate	CI	Estimate	CI
Valvular or rheumatic heart disease (CC 86)	0.083	(0.066, 0.105)	0.064	(0.055, 0.073)
Cerebrovascular disease (CC 97-99, 103)	0.030	(0.007, 0.053)	-0.008	(-0.017, 0.002)
Stroke (CC 95-96)	0.045	(0.010, 0.080)	0.021	(0.006, 0.034)
Vascular or circulatory disease (CC 104-106)	0.069	(0.048, 0.089)	0.033	(0.022, 0.042)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)	0.079	(0.041, 0.121)	0.043	(0.028, 0.058)
Diabetes mellitus (DM) or DM complications (CC 15-20, 119-120)	0.111	(0.095, 0.129)	0.083	(0.075, 0.092)
Renal failure (CC 131)	0.107	(0.086, 0.132)	0.102	(0.092, 0.113)
End-stage renal disease or dialysis (CC 129-130)	0.344	(0.294, 0.398)	-0.036	(-0.053, -0.017)
Other urinary tract disorders (CC 136)	0.084	(0.062, 0.104)	0.019	(0.010, 0.028)
Chronic obstructive pulmonary disease (COPD) (CC 108)	0.203	(0.184, 0.223)	0.112	(0.104, 0.121)
Pneumonia (CC 111-113)	0.147	(0.124, 0.170)	0.114	(0.105, 0.123)
Asthma (CC 110)	0.047	(0.011, 0.084)	-0.049	(-0.064, -0.035)
Disorders of fluid/electrolyte/acid-base (CC 22-23)	0.135	(0.109, 0.157)	0.018	(0.009, 0.028)
History of infection (CC 1, 3-6)	0.033	(0.015, 0.055)	0.010	(0.000, 0.020)
Metastatic cancer or acute leukemia (CC 7)	0.204	(0.130, 0.272)	0.097	(0.069, 0.122)
Cancer (CC 8-12)	0.016	(-0.009, 0.040)	-0.023	(-0.034, -0.013)
Iron deficiency or other unspecified anemias and blood disease (CC 47)	0.163	(0.143, 0.181)	0.172	(0.162, 0.182)
Decubitus ulcer or chronic skin ulcer (CC 148-149)	0.128	(0.097, 0.155)	0.059	(0.047, 0.071)
Dementia or other specified brain disorders (CC 49-50)	0.076	(0.054, 0.099)	-0.013	(-0.024, -0.003)
Protein-calorie malnutrition (CC 21)	0.139	(0.100, 0.179)	0.144	(0.130, 0.157)
Anterior myocardial infarction (ICD-9 410.00-	0.181	(0.148, 0.214)	0.103	(0.089, 0.120)

Risk variable	Part 1: Logit model		Part 2: Poisson model	
	Estimate	CI	Estimate	CI
410.19)				
Other location of Myocardial Infarction (ICD-9 410.20-410.69)	-0.032	(-0.064, 0.000)	-0.051	(-0.065, -0.035)

References

Dorsey KB GJ, Desai N, Lindenauer P, Young J, Wang C, DeBuhr, Parisi ML, Bernheim SM, Krumholz HM. 2015 Condition-Specific Measures Updates and Specifications Report Hospital-Level 30-Day Risk-Standardized Readmission Measures: AMI-Version 8.0, HF-Version 8.0, Pneumonia-Version 8.0, COPD-Version 4.0, and Stroke-Version 4.0. 2015; https://www.qualitynet.org/dcs/BlobServer?blobkey=id&blobnocache=true&blobwhere=1228890435217&blobheader=multipart%2Foctet-stream&blobheadername1=Content-Disposition&blobheadervalue1=attachment%3Bfilename%3DRdmn_AMIHFPNCOPDSTK_Msr_UpdtRpt.pdf&blobcol=urldata&blobtable=MungoBlobs. Accessed July 8, 2015.

Krumholz HM, Brindis RG,Brush JE, et al. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. *Circulation*. January 24, 2006 2006;113(3):456-462.

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

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 dual-eligible and African-American patients in the AMI cohort varies across hospitals (number of hospitals=4,163). The median percentage of dual-eligible patients is 10.0% (interquartile range [IQR] 0%-21.1%). The median percentage of black patients is 0% (IQR 0%-7.3%).

Empirical Association with the Outcome (univariate)

The mean patient-level observed days in acute care is higher for dual-eligible patients, 141.75 days in acute care per 100 discharges, compared with 106.65 days in acute care per 100 discharges for all other patients. The mean observed days in acute care for African-American patients was also higher at 148.38 days per 100 discharges compared with 107.62 days per 100 discharges 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 dual-eligible status and race variables in the context of a multivariable model. When we include either of these variables in a multivariate model that includes all of the claims-based clinical variables, the effect size of the variable is small. We also find that the c-statistics for the logit part of the model and the deviance R^2 values for the Poisson part of the model are similar with and without the addition of either of these variables into the model. The c-statistic for the logit model without the dual-eligibility indicator in the model is 0.596 and with the dual-eligibility indicator in the model is 0.597. The c-statistic for the logit model without the race indicator is 0.596 and with the race indicator is 0.597. The deviance R^2 values for the Poisson model with and without dual-eligibility indicator are 0.04. The deviance R^2 values for the Poisson model with and without the race indicator are 0.04. Furthermore, we find that the addition of any of these variables into the model has little to no effect on hospital performance. We examined the change in hospitals' EDAC with the addition of either of these variables. The mean We examined the change in hospitals' EDAC with the addition of either of these variables. The median absolute change in hospitals' EDAC when adding a dual-eligibility indicator is 0.50 EDAC per 100 discharges (interquartile range [IQR] 0.23-0.98; minimum 0.00-maximum 24.59), with a Spearman correlation coefficient between EDAC for each hospital with and without dual eligibility added of 0.9933. The median absolute change in hospitals' EDAC when adding a race indicator is 0.5002 EDAC per 100 discharges (IQR 0.23-0.97; minimum 0.00-maximum 12.71), with a Spearman correlation coefficient between EDAC for each hospital with and without race added of 0.9936.

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

Both the patient-level and hospital-level dual eligible effects were significant in the logistic part of the AMI EDAC model, but only the hospital-level effect was significant in the Poisson part of the model. This indicates that a) both the patient- and hospital-level dual eligible effects are associated with an increased risk of acute care but b) only the hospital-level effect is associated with the expected duration of that care.

Both the patient-level and hospital-level race effects were significant in the logistic part of the AMI EDAC model, but only the hospital-level effect was significant in the Poisson part of the model. This indicates that a) both the patient- and hospital-level race effects are associated with an increased risk of acute care but b) only the hospital-level effect is associated with the expected duration of that care.
Because both the hospital- and patient-level effects contribute to the increased risk, if the dual eligible or race variables were used in the model to adjust for patient-level differences, then some of the differences in both risk of acute care and expected duration of care between hospitals would also be adjusted for, potentially obscuring a signal of hospital quality.

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

AMI EDAC Decomposition Analysis

Parameter	Logistic model estimate (standard error)	Logistic model p-value	Poisson model estimate (standard error)	Poisson model p-value
Dual Eligible – Patient-Level	0.1403 (0.0144)	<.0001	0.0035 (0.0060)	0.5622
Dual Eligible – Hospital-Level	0.1777 (0.0567)	<.0001	0.1029 (0.0515)	0.0389
African American – Patient-Level	0.0777 (0.0186)	<.0001	0.0072 (0.0075)	0.3411
African American – Hospital-Level	0.0509 (0.0439)	0.0034	0.3677 (0.0482)	<.0001

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*)

Dataset
This model selection process was performed using one half (the development sample) of the random three-year split sample.

Approach to Determining Model Specifications
Because the outcome, number of days in acute care, is novel not only for quality measurement but also in the literature as a measure of utilization, we considered a range of model specifications. We performed a number of analyses to determine the best model specification for the number of days in acute care. This is a pseudo-count variable (similar to a count variable, but taking half-integer values for half-days of acute care), and we therefore considered models that were generalized count models. All model development was performed using the development sample.

Inspection of the distribution of the outcome determined that the number of event days was highly skewed, with a large number of zeroes. Thus, we considered models appropriate for skewed data, including approaches that modeled the zero-day outcomes and non-zero day outcomes separately. We only considered approaches that allowed us to incorporate exposure time to account for differential risk.

First, using only patients with non-zero days, we estimated a generalized linear model (GLM) using a Poisson specification, and applied a Park test (Manning and Mullahy, 2001); the Park test indicated that Poisson was the best fit for our outcome. The Poisson model is commonly used for modeling count data and can be generalized to dependent variables that take non-integer values, such as ours.
We then considered three different model specifications for the full set of outcomes (zero and non-zero days): Poisson, zero-inflated Poisson (ZIP), and two-part logit/Poisson ("hurdle" model). For each model, we included an offset for the number of days the patient survived discharge, up to 30 (i.e., the exposure time). For the hurdle

model, we included exposure time as an offset for each part because the Poisson part included only observations with non-zero days, it was technically a 'truncated' Poisson model.
For each of the three specifications listed above, we estimated (non-hierarchical) generalized linear models with days in acute care as the outcome. We compared the three different model specifications for the outcome using the following criteria: Akaike information criterion (AIC), Baysian information criterion (BIC), and log-likelihood.

Criterion	Poisson	Zero-inflated Poisson	Two-part logit/Poisson
Akaike information criterion (AIC)	2,190,000	1,450,000	1,440,000
Bayesian information criterion (BIC)	2,190,000	1,450,000	1,440,000
Log-likelihood	-1,095,000	-725,000	-720,000

We selected the best model based on these statistics and judgment regarding the technical challenges of extending each to a random effects model for the measure. The AIC is a measure of the relative quality of statistical models for a given set of data. The best performing model was the two-part logit/Poisson model. This model also made the most sense conceptually, with the likelihood of returning for acute care being modelled separately from the number of days of acute care received.

Assessing Model Discrimination and Calibration

Discrimination: We computed two different statistics – one for the logit part of the model and one for the Poisson part – using the development sample. For the logit model of zero versus non-zero days, which includes all patients in the cohort, we calculated the c-statistic. For the Poisson model of non-zero days, which includes only patients with some acute care, we calculated the deviance R^2. The deviance R^2 is computed from the difference in the log-likelihoods between the final model and an empty model (no covariates) attributed to each observation, averaged over all observations (Cameron, Windmeijer, 1996).

Calibration Statistics

In a generalization of the calibration statistics for logistic models, we calculated the linear prediction Z = XB and W = XC using the coefficients B and C from the development sample and data X from the validation sample. We then estimated a model using the same functional form but only two independent variables, Z for the truncated Poisson part and W for the logit part. The intercepts and coefficients of Z and W in these second models are reported as (γ_0, γ_1), the calibration statistics for each part of the model. The closer they are to (0, 1), the better the model calibration (Harrell, 2013).

Calibration Plot

To further assess model calibration we constructed calibration plots with mean predicted and mean observed days in acute care plotted against decile of predicted utilization rate (predicted days/exposure days).

References

Cameron AC and Windmeijer FAG. R-Squared Measures for Count Data Regression Models with Applications to Health-Care Utilization. Journal of Business & Economic Statistics, Vol. 14, No. 2 (Apr., 1996), pp. 209-220.

Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer New York; 2013.

Manning WG, Mullahy J. Estimating log models: to transform or not to transform? *Journal of health economics*. 2001;20(4):461-494.

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*)**:**

Dataset
The model discrimination statistics were calculated using the development sample:

Discrimination Statistics:
C-statistic for logit part of model: 0.60
Deviance R^2 for truncated Poisson part of model: 0.040 (4.0%)

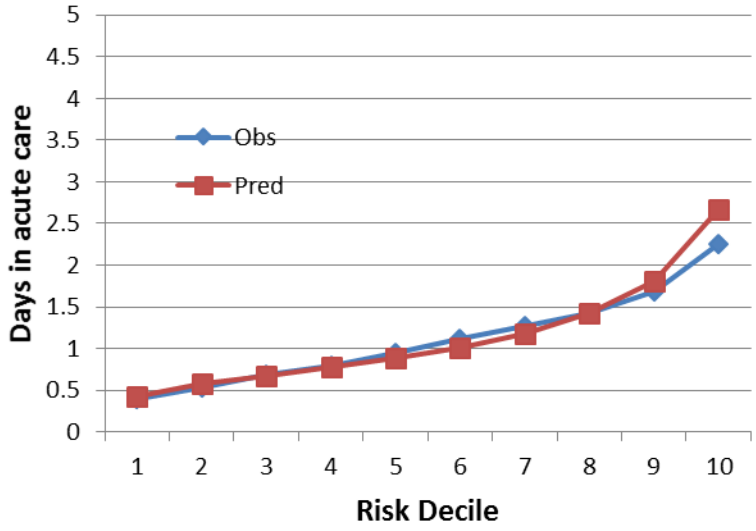
2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Dataset
The model discrimination statistics were calculated using both the development and validation samples; see section 1.7.

Calibration Statistics (y0, y1):
Logit part of model: (-0.10, 0.98)
Poisson part of model: (-0.04, 0.97)

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

Calibration Plot:
The plot below shows that the model underestimates risk for the lowest risk decile patients and slightly overestimates risk for the highest risk decile patients.



2b4.9. Results of Risk Stratification Analysis:

N/A. This measure is not risk-stratified.

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 for the logit part of the model was 0.59; the deviance R^2 for the Poisson part of 0.040 is consistent with deviance R^2 for other count data models, indicating good model calibration.

Calibration Statistics
Over-fitting (Calibration γ0, γ1)
If the γ_0 in the validation sample is substantially far from zero and the γ_1 is substantially far from one, there is potential evidence of over-fitting. The calibration value of close to 0 at one end and close to 1 to the other end indicates good calibration of the model.

Calibration Plot
The calibration plot shows very good agreement between the mean of predicted days and the mean of observed days within same risk decile.

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*)

N/A.

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)*

To categorize hospital performance, we estimated each hospital's EDAC and the corresponding 95% credible interval (CI) described in the attached Appendix (Section 2.7.2). We assigned hospitals to a performance category by comparing each hospital's EDAC interval estimate to zero. Comparative performance for hospitals with 25 or more eligible cases was classified as follows:

- "Lower than expected" if the entire 95% CI surrounding the hospital's days is below zero.
- "No different than expected" if the 95% CI surrounding the hospital's days includes zero.
- "Higher than expected" if the entire 95% CI surrounding the hospital's days is above zero.

If a hospital has fewer than 25 eligible cases for a measure, we assigned the hospital to a separate category: "The number of cases is too small (fewer than 25) to reliably assess the hospital's EDAC."

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*)

Of 4,286 hospitals in the three-year study cohort, 254 had EDACs "lower than expected," 1,440 were "no different than expected," and 579 had EDACs "higher than expected." 2,013 were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing. The mean EDAC per 100 discharges for hospitals in the top decile of performance is -23.3, compared to 170.4 for hospitals in the bottom decile.

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 hospital-level EDAC suggests there are meaningful differences in the quality of care received across hospitals for the AMI EDAC measure.

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 criterion is directed 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).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of 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 demonstrating comparability of 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.**

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)
Public Reporting Not in use	

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

N/A. The measure is not yet in use.

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?*)

This measure is not currently publicly reported or used in an accountability application because it only recently completed development.

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 has been finalized for use in CMS's Hospital Inpatient Quality Reporting (IQR) program starting in Fiscal Year 2018 (80 FR 49690).

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**

N/A

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.

Since this measure is not yet in use, there are no performance results to assess improvement.

We expect there will be improvement in measure scores over time since publicly reported measure scores can reduce adverse patient outcomes associated with days spent in acute care for AMI by capturing and making acute care utilization following the index hospitalization more visible to providers and patients.

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 or model testing. 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)**
0229 : Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following heart failure (HF) hospitalization for patients 18 and older
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 all-cause risk-standardized readmission rate (RSRR) following elective primary total hip arthroplasty (THA) and total knee arthroplasty (TKA)
1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)
1891 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization

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?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
We developed the measure in the Medicare Fee-for-Service (FFS) population and completely harmonized the cohort definition and risk-adjustment strategy with those of the existing CMS 30-day AMI readmission measure. However, while the existing measure counts readmissions as a dichotomous outcome, the proposed measure counts the number of days for all readmissions during the follow-up period, as well as the number of days of observation stays and ED visits. This difference in the outcome measure imposes differences on the statistical modeling and reporting format. There are no differences in data collection burden.

5b. Competing Measures
The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR

Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
N/A

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** AMI_Excess_Days_in_Acute_Care_NQF_Appendix_01-29-16_v1.0.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services (CMS)
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 (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.
Yale New Haven Health Services Corporation/Center for Outcomes Research (YNHHSC/CORE) Measure Reevaluation Team Members
1. Faseeha K. Altaf, MPH- Lead Project Coordinator. Provided experience relevant to performance measurement.
2. Susannah Bernheim, MD, MHS- Project Director. Provided experience relevant to clinical content and performance measurement.
3. Nihar Desai, MD, MPH- Clinical Consultant. Provided experience relevant to clinical content and performance measurement.
4. Jacqueline Grady, MS- Supporting Analyst. Provided experience relevant to performance measurement.
5. Jeph Herrin, PhD- Statistician. Provided experience relevant to performance measurement.
6. Leora Horwitz, MD, MHS- Project Lead. Provided experience relevant to clinical content and performance measurement.
7. Zhenqiu Lin, PhD- Director of Analytics. Provided experience relevant to performance measurement.
8. Shuling Liu, PhD- Statistical Consultant. Provided experience relevant to performance measurement.
9. Chi Ngo, MPH- Research Associate. Provided experience relevant to performance measurement.
10. Arjun Venkatesh, MD, MBA- Clinical Consultant. Provided experience relevant to clinical content and performance measurement.
11. Changqin Wang, MD, MS- Co-Lead Analyst. Provided experience relevant to performance measurement.
12. Yongfei Wang- Supporting Analyst. Provided experience relevant to performance measurement.
13. Sharon-Lise Normand, Ph.D.* - Statistical Consultant. Provided statistical expertise for the project.

*Harvard Medical School

Technical Expert Panel (TEP) Members
1. Anonymous Patient- Patient Representative. Provided patient perspective.
2. Kevin E. Driesen, PhD, MPH, MA- Assistant Professor, Mel and Enid Zuckerman College of Public Health; Director, Arizona Rural Hospital Flexibility Program. Provided experience relevant to performance measurement.
3. David Engler, PhD- Senior Vice President for Leadership and Innovation, America's Essential Hospitals. Provided experience relevant to clinical content, performance measurement, and coding and informatics.
4. Timothy Farrell, MD- Assistant Professor of Medicine, Adjunct Professor of Family Medicine, Physician Investigator; University of Utah School of Medicine. Provided experience relevant to clinical content and performance measurement.
5. Karen Farris, PhD- Charles R. Walgreen III Professor of Pharmacy Administration, Director of the Social and Administrative

Pharmacy Graduate Program; University of Michigan College of Pharmacy. Provided experience relevant to performance measurement.
6. Maura C. Feldman, MSW- Director for Hospital Performance Measurement and Improvement, Blue Cross Blue Shield of Massachusetts. Provided consumer perspective.
7. Jay A. Gold, MD, JD, MPH- Senior Vice President and Chief Medical Officer, MetaStar. Provided experience relevant to clinical content and performance measurement.
8. Sally Hinkle, DNP, MPA, RN- Director of Performance Improvement and Clinical Value, Temple University Hospital. Provided experience relevant to performance measurement.
9. Amy Jo Haavisto Kind, MD, PhD - Assistant Professor of Geriatrics, University of Wisconsin School of Medicine and Public Health; Attending Physician, William S. Middleton VA. Provided experience relevant to clinical content and performance measurement.
10. Marjorie King, MD, FACC, MAACVPR- Director of Cardiac Services, Helen Hayes Hospital. Provided experience relevant to clinical content and performance measurement.
11. Eugene Kroch, PhD- Vice President and Chief Scientist, Premier. Provided experience relevant to performance measurement.
12. Keith D. Lind, JD, MS, BSN- Senior Policy Advisor, American Association of Retired Persons (AARP) Public Policy Institute. Provided consumer perspective.
13. Grace McConnell, PhD- Patient Representative. Provided patient perspective.
14. Michael A. Ross, MD, FACEP- Medical Director, Professor of Emergency Medicine; Emory University School of Medicine. Provided experience relevant to clinical content and performance measurement.
15. Mark Louis Sanz, MD- Interventional Cardiologist, International Heart Institute of Montana. Provided experience relevant to clinical content and performance measurement.
16. Paul Takahashi, MD- Associate Professor of Medicine, Mayo Clinic College of Medicine. Provided experience relevant to performance measurement.

Methodology Work Group Members
1. Arlene Ash, PhD- Professor and Division Chief, University of Massachusetts Medical School. Provided experience relevant to performance measurement.
2. Jeremiah Brown, PhD, MS- Assistant Professor of Health Policy and Clinical Practice, The Dartmouth Institute for Health Policy and Clinical Practice. Provided experience relevant to performance measurement.
4. Grant Ritter, PhD, MS, MA- Senior Scientist, Schneider Institute for Health Policy & Heller Graduate School. Provided experience relevant to performance measurement.
5. Patrick Romano, MD, MPH- Professor of Medicine and Pediatrics, University of California Davis School of Medicine. Provided experience relevant to performance measurement.

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:**
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

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 #: 2882
De.2. **Measure Title:** Excess days in acute care (EDAC) after hospitalization for pneumonia
Co.1.1. **Measure Steward:** Centers for Medicare & Medicaid Services (CMS)
De.3. **Brief Description of Measure:** This measure assesses days spent in acute care within 30 days of discharge from an inpatient hospitalization for pneumonia to provide a patient-centered assessment of the post-discharge period. This measure is intended to capture the quality of care transitions provided to discharged patients hospitalized with pneumonia by collectively measuring a set of adverse acute care outcomes that can occur post-discharge: emergency department (ED) visits, observation stays, and unplanned readmissions at any time during the 30 days post-discharge. In order to aggregate all three events, we measure each in terms of days. In 2016, the Center for Medicare and Medicaid Services (CMS) will begin annual reporting of the measure for patients who are 65 years or older, are enrolled in fee-for-service (FFS) Medicare, and are hospitalized in non-federal hospitals.
1b.1. **Developer Rationale:** The goal of this measure is to improve patient outcomes by providing patients, physicians, and hospitals with information about hospital-level, risk-standardized outcomes following hospitalization for pneumonia. Measurement of patient outcomes allows for a broad view of quality of care that cannot be captured entirely by individual process-of-care measures. Safely transitioning patients from hospital to home requires a complex series of tasks which would be cumbersome to capture individually as process measures: timely and effective communication between providers, prevention of and response to complications, patient education about post-discharge care and self-management, timely follow-up, and more. Suboptimal transitions contribute to a variety of adverse events post-discharge, including ED evaluation, need for observation, and readmission. Measures of unplanned readmission already exist, but there are no current measures for ED and observation stay utilization. It is thus difficult for providers and consumers to gain a complete picture of post-discharge outcomes. Moreover, separately reporting each of these outcomes encourages "gaming," such as re-categorizing readmission stays as observation stays to avoid a readmission outcome. By capturing a range of acute care events that are important to patients, we can produce a more complete picture of post-discharge outcomes that better informs consumers about care quality and incentivizes global improvement in transitional care.

S.4. **Numerator Statement:** The outcome of the measure is a count of the number of days the patient spends in acute care within 30 days of discharge. We define days in acute care as days spent in an ED, admitted to an observation unit, or admitted as an unplanned readmission for any cause within 30 days from the date of discharge from the index pneumonia hospitalization. Each ED treat-and-release visit is counted as one half-day (0.5 days). Observation stays are recorded in terms of hours and are rounded up to the nearest half-day. Each readmission day is counted as one full day (1 day). We count all eligible outcomes occurring in the 30-day period, even if they are repeat occurrences.
S.7. **Denominator Statement:** The target population for this measure is Medicare FFS beneficiaries aged 65 years and older hospitalized at non-Federal acute care hospitals for pneumonia.

The cohort includes admissions for patients discharged from the hospital with a principal discharge diagnosis of pneumonia (see codes below in S.9) and with continuous 12 months Medicare enrollment prior to admission. The measure will be publicly reported by CMS for those patients 65 years and older who are Medicare FFS beneficiaries admitted to non-federal hospitals.

Additional details are provided n S.9 Denominator Details.
S.10. **Denominator Exclusions:** The measure excludes index admissions for patients:
1. Without at least 30 days post-discharge enrollment in FFS Medicare.
2. Discharged against medical advice (AMA);
3. Admitted within 30 days of a prior index discharge;

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Administrative claims
S.26. **Level of Analysis:** Facility

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

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 any measure; however, it is harmonized with a measure of hospital-level, all-cause, 30-day, risk-standardized readmission following pneumonia hospitalization.

New Measure -- Preliminary Analysis

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.

- This measure calculates excess days in acute care (EDAC) for patients with pneumonia. This measure is intended to capture the quality of care transitions provided to discharged patients hospitalized with pneumonia by collectively measuring a set of adverse acute care outcomes that can occur post-discharge: emergency department (ED) visits, observation stays, and unplanned readmissions at any time during the 30 days post-discharge. In order to aggregate all three events, this measure assesses each in terms of days.
- As a rationale for measuring this health outcome, the developer suggests that hospitals are able to influence readmission rates through a broad range of clinical activities including communication between providers, prevention of, and response to, complications, patient safety and coordinated transitions to the outpatient environment.
- The developer cites that "in the context of the Centers for Medicare and Medicaid Services' (CMS's) publicly reported readmission measures, the increasing use of ED visits and observation stays has raised concerns that current readmission measures do not capture the full range of unplanned acute care in the post-discharge period (Vashi et al., 2013; Rising et al., 2012; Feng et al., 2012). Observation stays can occur in many different parts of the hospital, including dedicated treatment rooms, the ED, or inpatient units. In particular, there is concern that high use of observation stays could in some cases replace readmissions, and that hospitals with high rates of observation stays in the post-discharge period may therefore have low readmission rates that do not accurately reflect the quality of care (Vashi et al., 2013)."

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

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 developer provides performance data from one measurement period from 2010-2012, covering a total of 495,130 discharges. The analysis includes hospitals that have at least 25 pneumonia index admissions in the 2-year

period.

- The data show that during the measurement period of 2010-2012, pneumonia readmission rates ranged from a minimum of -67.59% to a maximum of 229.99%, with the 10th percentile at -29%, the 50th percentile at 4.28%, and the 90th percentile at 50.56%.

Disparities

- To help in assessment of potential disparities, the developers also provide performance scores for hospitals serving a low proportion of dual eligible patients vs. those serving a high proportion of dual eligible patients and performance scores for hospitals serving a low proportion of African-American patients vs. those serving a high proportion of African-American patients.

- By proportion of **Dual Eligible Patients:**

  // Low proportion (=10%) dual-eligible patients//Hospitals with a high proportion (=25.81%) dual-eligible patients
  Number of Measured Hospitals//1,187//1,158
  Number of Patients//118,183 patients in low-proportion hospitals/87,732 in high-proportion hospitals
  Maximum//194.15 //356.31
  90th percentile//39.60//59.76
  75th percentile//16.02//29.12
  Median (50th percentile)//-2.51//4.53
  25th percentile//-19.62//-15.02
  10th percentile//-30.93//-29.75
  Minimum //-67.59//-60.52

- By proportion of **African-American Patients:**

  // Low proportion (=0%) African-American patients//Hospitals with a high proportion (=8.05%) African-American patients
  Number of Measured Hospitals//1,978 //1,164
  Number of Patients//96,720 patients in low-proportion hospitals//144,724 in high-proportion hospitals
  Maximum//194.15//356.31
  90th percentile//26.94//68.17
  75th percentile//6.71//42.18
  Median (50th percentile)//-8.27//16.26
  25th percentile//-24.52//-5.90
  10th percentile//-34.63//-24.02
  Minimum//-73.40//-67.59

- The developer explains that: "low-proportion hospitals are those hospitals whose population of dual-eligible patients or of African-American patients is small enough to place them at or below the 25th percentile among all hospitals; and high proportion are those hospitals whose population of dual eligible patients or African-American patients is large enough to place them at or above the 75th percentile among all hospitals."

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)

1. Importance to Measure and Report

1a. Evidence to Support Measure Focus

Comments: **The evidence is that the readmission measures may not be capturing emergency department visits and observation visits, because the patient is not actually admitted into the hospital. These visits are included as "acute care," and their use may mask the quality of care when using a standard readmission measure. Once a high rate has been identified, it appears that the hospital would implement operational changes very similar to the changes implemented following an increase in standard readmission rates.

There was no evidence presented that determined the number of days in acute care reduced by implementing various clinical or operational processes.

**Yes, there is information about potential gaming of readmission metric. Thus this metric accounts for ED, observation stays, and readmissions after discharge from a hospitalization for pneumonia

1b. Performance Gap

Comments: **The authors refer to their measurement of readmission rates from 2010-2012 but do not indicate whether these rates include ED and observation visits. The results include negative readmission rates, with no explanation; based on the description of the numerator and denominator, it is not clear how negative rates were calculated.

**There is variability demonstrated with hospitals ranging from -67 days to 230 risk standardized EDAC days per 100 discharges. I do not see disparity data presented

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

Comments: **NA

**not a composite rating but weighting ED visits as 0.5 days seems reasonable.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

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): Administrative claims

Specifications:

- This measure calculates the number of days spent in acute care within 30 days of discharge from an inpatient hospitalization for pneumonia to provide a patient-centered assessment of the post-discharge period
- The outcome of the measure is a count of the number of days the patient spends in acute care within 30 days of discharge. The measure defines days in acute care as days spent in an ED, admitted to an observation unit, or admitted as an unplanned readmission for any cause within 30 days from the date of discharge from the index heart failure hospitalization. Each ED treat-and-release visit is counted as one half-day (0.5 days). Observation stays are recorded in terms of hours and are rounded up to the nearest half-day. Each readmission day is counted as one full-day (1 day). The measure counts all eligible outcomes occurring in the 30-day period, even if they are repeat occurrences.
- The Numerator is the number of days the patient spends in acute care within 30 days of discharge. Days in acute care is defined as days spent in an ED, admitted to an observation unit, or admitted as an unplanned readmission for any cause within 30 days from the date of discharge from the index pneumonia hospitalization.
- The Denominator is the Medicare FFS beneficiaries aged 65 years and older hospitalized at non-Federal acute care hospitals for pneumonia.
- The denominator population is defined using ICD-9 and ICD-10 codes; a list of applicable codes is included in the submission.

- The data sources for this measure are Medicare Part A inpatient, Part B hospital outpatient claims and physician Carrier claims, and the Medicare Enrollment Database (EDB).
- The measure's time window is three years.
- The measure is risk-adjusted using a statistical risk model (see details below).

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]

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**

- The developer has assessed reliability at both the data element and the performance score levels.
- The datasets used for testing included Medicare Parts A and B claims, the Medicare Enrollment Database (EDB), and the Chronic Condition Data Warehouse (CCW) 100% condition-specific dataset to capture emergency department (ED) visits and observation stays. Additionally, census data were used to assess socio-demographic factors.
  - **Data element reliability**:
    - With regard to data element reliability, the developer notes that the measure has been developed to avoid the use of claims data elements that are thought to be coded inconsistently across hospitals or providers, instead using fields that are consequential for payment and which are audited by CMS. Additionally, the developer used the final risk-adjustment variables in the current CMS 30-day pneumonia readmission measure.
    - Additionally, the developer compared variable frequencies between the development and validation samples.
  - **Performance score reliability**:
    - The developer defines performance score reliability as the degree to which repeated measurements of the same entity agree with each other.
    - In line with this thinking, the developer's approach to assessing score-level reliability was to consider the extent to which assessments of a hospital using different but randomly-selected subsets of patients produce similar measures of hospital performance. The developers refer to this as a "test-retest" approach; it may also be called a "split-half" method.
    - For test-retest reliability, the developer calculated the EDAC for each hospital using first the development sample, then the validation sample. Thus, each hospital twice was measured twice, each time using an entirely distinct set of patients. The developer states that the extent to which the calculated measures of these two subsets agree is evidence that the measure is assessing an attribute of the hospital, not of the patients. As a metric of agreement, the developer calculated the intra-class correlation coefficient (ICC) as defined by ICC[2,1] by Shrout and Fleiss (1979) and assessed the values according to conventional standards (Landis and Koch, 1977).
    - A total of 990,260 admissions were examined, with 495,130 in each sample.
    - The agreement between the two EDAC values for each hospital (as measured by an intra-class correlation coefficient (ICC)) was **0.80**; the developer states that according to the conventional interpretation, this is considered a "substantial" level of agreement.

- The developer notes that this analysis was limited to hospitals with to hospitals with at least 8 discharges in both samples to approximate the set of hospitals that would have at least 24 discharges over three years and are thus likely to be included in public reporting. [*Note: It is unclear whether the measure itself is limited to hospitals with 8 or more cases and if three years of data are needed to calculate the measure; if it is not, then testing was not conducted with the measure as specified*.]
- The developer expects that the correlation coefficient would be higher using a full three-year sample since it would include more patients. To correct this problem, the developer used the Spearman-Brown prophecy formula (Spearman 1910, Brown 1910) to adjust the ICC[2,1] to represent three years of data.

- The developer's overall interpretation of reliability testing results is that the compared to the development sample, the mean age of patients and the frequencies of the risk-adjustment variables were very similar in the validation sample; this indicates that the data elements are reliable and that the ICC score from performance score analysis demonstrates moderate agreement across samples. The developer notes that the ICC [2,1] score of 0.80, estimated for three years of data, demonstrates moderate agreement between samples across the full range of measure values. We interpret this to mean that when used with a full three years of data, the measure will be reliable by the standards of hospital measurement.

Guidance from the Reliability Algorithm

- Question 1. Submitted specifications are precise, unambiguous, and complete. Measure can be consistently implemented.
- Question 2. Empirical reliability testing was conducted using statistical tests with the measure as specified.
- Question 3. Empirical validity testing of patient-level data was conducted.
- Question 4. Reliability testing was conducted with computed performance measure scores for each measured entity.
- Question 5. Random split-half correlation was used to assess the proportion of variability due to real differences among the measured entities.
- Question 6. The ICC was 0.80 which is considered a moderate level of agreement.

Questions for the Committee:

- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*
- *Does the measure need three years of data to achieve this level of reliability?*

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

2b. Validity

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**

- This measure calculates the number of days spent in acute care within 30 days of discharge from an inpatient hospitalization for pneumonia to provide a patient-centered assessment of the post-discharge period.
- As a rationale for measuring this health outcome, the developer suggests that hospitals are able to influence readmission rates through a broad range of clinical activities including communication between providers, prevention of, and response to, complications, patient safety and coordinated transitions to the outpatient

environment.

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.

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**

- The developer demonstrated measure validity through prior validity testing done on their claims-based measures, through use of established measure development guidelines, and by systematic assessment of measure face validity by a Technical Expert Panel (TEP).
  - Empirical Validity Testing
    - The developer notes this measure is closely related in design to the existing readmission measure for patients with pneumonia. While this measure includes additional endpoints and measures them in a different metric (days rather than rates), the developer expects that hospitals would have similar – though not identical – performance rankings on the two measures. Therefore as one assessment of validity, they compared the rankings of all hospitals using the two measures to assess the consistency of hospital performance on closely related outcomes. The developer calculated the Pearson correlation, and graphed the readmission measure against the EDAC measure to determine if there were outliers.
    - Comparison of the new measure with the existing CMS 30-day pneumonia readmission measure found a Pearson correlation of 0.732 ($P < 0.0001$).
  - Validity of Claims-Based Measures:
    - The developer states that they have demonstrated for a number of other readmission measures the validity of claims-based measures by comparing either the measure result or the individual data elements against medical records.
    - Claims model validation was conducted by building comparable models using abstracted medical chart data for risk adjustment. 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.
  - Validity Indicated by Established Measure Development Guidelines
    - The developer states that this measure was developed in consultation with national guidelines for publicly reported outcomes measures, with outside experts, and with the public.
  - Validity as Assessed by External Groups:
    - Input was obtained through regular discussions with an advisory working group, a TEP, and a 30-day public comment period.
  - Face Validity as Determined by TEP:
    - The developer asked members of the TEP to note their agreement with the statement "The risk-standardized acute care days obtained from the measures as specified can be

used to distinguish between better and worse quality hospitals."
    - Of the TEP members who responded, 91% agreed (83% moderately or strongly agreed) that the measure will provide an accurate reflection of quality.

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:

- Patients in the following categories are excluded from the measure:
    - Discharged patients without at least 30 days post-discharge information
    - Discharges against medical advice (AMA)
    - Admissions within 30 days of a prior index admission
- The developer notes that all exclusions were determined by careful clinical review and have been made based on clinically relevant decisions and to ensure accurate calculation of the measure
- To determine the impact of exclusions, the developer examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion.
- The number and percentage of patients excluded for each criterion are as follows:
    - Without at least 30 days post-discharge enrollment in FFS Medicare for index admissions: **6,237 (0.6%)**
    - Discharged against medical advice (AMA): **2,636 (.3%)**
    - Admissions within 30 days of a prior index admission: **46,485 (4.5%)**
- The developer also provides the distribution across hospitals for each exclusion criterion.
- The developer notes that the first exclusion criterion, is needed since the outcome cannot be assessed in this group since claims data are used to determine whether a patient returned to the hospital for an ED visit, was placed under observation care, or was readmitted.
- The developer states that the second exclusion criterion is needed for acceptability of the measure to hospitals, who do not have the opportunity to adequately deliver full care and prepare the patient for discharge.
- The developer notes that exclusion criterion 3 is needed to prevent admissions from being counted as both an index admission and a readmission, consistent with the approach taken in the pneumonia readmission measure.

Questions for the Committee:

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

Risk adjustment summary

- For this measure the developer adopted the risk factors from the existing CMS 30-day pneumonia readmission measure. These risk factors are comprised of age, sex, and condition categories (CCs) for prior 12-month and current claims.
- The developer notes these risk factors had been systematically chosen as predictors of any readmission for the same patient cohort as the current measure; the outcome of this measure is dominated by the number of days of a readmission, so they judged it unlikely that repeating the original analysis would produce different results.
- The developer confirmed that there were no additional risk factors to consider by comparing the model

estimated using the a priori set of risk factors to a model which included all additional CCs.

- The measure employs a hierarchical generalized linear model [HGLM]) that consists of two parts, a logit model and a truncated Poisson model. The two-part logit/Poisson model (often called a "hurdle" model) assumes that the outcome results from two related processes: an initial dichotomous event – that a patient has at least one acute care event – which is modeled as the logit of the probability of the event, and for patients with an event (those which clear the "hurdle"), the number of days, which is modeled as a Poisson process. The outcome, number of days, is a half-integer count variable (because ED visits count as 0.5 days).
- There are two random effects for each hospital, one for the logit model and one for the truncated Poisson model, as well as a covariance between the two random effects. The developer suggests that the random effects allows for within-hospital correlation of the observed outcome and accommodates the assumption that underlying differences in quality across hospitals lead to systematic differences in outcomes.
- The final set of 41 risk-adjustment variables is included in the testing attachment; the odds ratio associated with each variable is also provided.
- The developers also considered a number of variables related to sociodemographic status (SDS) for potential inclusion in the risk-adjustment model. Candidate SDS variables were selected for examination based on a review of literature and national data sources.
- **Conceptual analysis of the need for SDS adjustment:**
    - The developers note there is a large body of literature linking various SES factors and African-American race to worse health status and higher readmission risk with income, education, and occupational level being the most commonly examined variables. The developers state that the literature directly examining how SES factors or race might influence the likelihood of older, insured, Medicare patient of being readmitted within 30 days of an admission for heart failure is more limited.
    - The developers state that few studies directly address causal pathways for SDS factors to affect 30-day readmission rates or examine the role of the hospital in these pathways.
    - There are at least four potential pathways for SDS factors to affect 30-day readmission rates:
        - One potential pathway is the relationship to health status at the time of admission. SDS factors 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. The developers note that this pathway should be largely accounted for by their clinical risk-adjustment model.
        - The next potential path way is that patients with low income and African-American patient are more likely to be seen in lower quality hospitals, which can contribute to increased risk of readmission.
        - The third major pathway is that a patient's race or SDS status cause them to experience differential, lower quality care or may not receive the differentiated care they require.
        - Finally, some SES risk factors may affect the likelihood of readmission without directly affecting health status at admission or the quality of care received during the hospitalization. Patients may have worse outcomes due to competing economic priorities or a lack of access to care outside the hospital.
- **Empirical analysis of SDS factors:**
    - The developers considered African-American race, and dual-eligible status-i.e. enrolled in both Medicare and Medicaid. The developers assessed the relationship between the SES variables and race with the outcome and examined the incremental effect in a multivariable mode.
    - The developer assessed the relationship between the SDS variables and the days in acute care and

examined the incremental effect of SDS in a multivariable model, evaluating the extent to which the addition of any one of these variables improved model performance or changed hospital results.
    - The developer notes that 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. Therefore, the developers performed a decomposition analysis to assess the independent effects of the SES and race variables at the patient level and the hospital level.
    - The developers' analysis found that the prevalence of SDS factors in the pneumonia cohort does vary across measured entities.
    - With regard to the empirical association of each SDS variable with the outcome (univariate), the analysis found that patient-level observed days in acute care for dual-eligible patients was higher, at 145.57 per 100 discharges compared with 119.55 days in acute care per 100 discharges for all other patients. The readmission rate for African-American patients was also higher at 176.11 days per 100 discharges compared with 119.91 days per 100 discharges for patients of all other races.
    - With regard to the strength and significance of the SDS variables in the context of a multivariable model, the developers' analysis found that the effect size of each of these variables is small. The developers also found that the c-statistics (i.e., predictive value) for the logit part of the model and the deviance R2 values for the Poisson part of the model are similar with and without the addition of either of these variables into the model. Additionally the developers found the addition of these variables has little to no effect on hospital performance.
        - The median absolute change in hospitals' EDAC when adding a dual-eligibility indicator is 0.40 EDAC per 100 discharges (interquartile range [IQR] 0.19-0.69; minimum 0.00-maximum 8.50), with a Spearman correlation coefficient between EDAC for each hospital with and without dual eligibility added of 0.9997.
        - The median absolute change in hospitals' EDAC when adding a race indicator is 0.56 EDAC per 100 discharges (IQR 0.27-0.98; minimum 0.00-maximum 11.69), with a Spearman correlation coefficient between EDAC for each hospital with and without race added of 0.9997.
    - The developers state that both the patient-level and hospital-level dual eligible and race effects were significant in the logistic part of the pneumonia EDAC model, but only the hospital-level effect was significant in the Poisson part of the model. This indicates that a) both the patient- and hospital-level dual eligible and race effects are associated with an increased risk of acute care but b) only the hospital-level effect is associated with the expected duration of that care. The developers note that if the dual eligible or race are 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 developers state that given these findings and complex pathways that could explain any relationship between SDS and readmission, they did not incorporate SDS variables into the measure.
- **Risk Model Diagnostics:**
    - To assess model discrimination the developers computed two different statistics: one for the logit part of the model and one for the Poisson part.
        - For the logit model of zero versus non-zero days, which includes all patients in the cohort, the developers calculated the c-statistic.
            - C-statistic for logit part of model: 0.616
        - For the Poisson model of non-zero days, which includes only patients with some acute care, the developers calculated the deviance R2. The deviance R2 is computed from the difference in the log-likelihoods between the final model and an empty model (no covariates) attributed to each observation, averaged over all observations.

        - Deviance R2 for truncated Poisson part of model: 0.034 (3.4%)
    - The developers interpret these results as good model calibration.
- In a generalization of the calibration statistics for logistic models, the developers calculated the linear prediction Z = XB and W = XC using the coefficients B and C from the development sample and data X from the validation sample. The developers then estimated a model using the same functional form but only two independent variables, Z for the truncated Poisson part and W for the logit part. The intercepts and coefficients of Z and W in these second models are reported as (γ_0, γ_1), , the calibration statistics for each part of the model. The closer they are to (0, 1), the better the model calibration
    - Calibration Statistics (y0, y1):
        - Logit part of model: (-0.05, 0.99)
        - Poisson part of model: (-0.05, 0.97)

Questions for the Committee:

- *Is an appropriate risk-adjustment strategy included in the measure?*
- *Does the Committee agree with the developer's use of current claims data for risk adjustment variables?*
- *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 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)*:

- To categorize hospital performance, the developers estimated each hospital's EDAC and the corresponding 95% credible interval (CI).
- The developers then assigned hospitals to a performance category by comparing each hospital's EDAC interval estimate to zero. Comparative performance for hospitals with 25 or more eligible cases was classified as follows:
    - "Lower than expected" if the entire 95% CI surrounding the hospital's days is below zero.
    - "No different than expected" if the 95% CI surrounding the hospital's days includes zero.
    - "Higher than expected" if the entire 95% CI surrounding the hospital's days is above zero.
- Hospitals with fewer than 25 eligible cases were assigned to a separate category: "The number of cases is too small (fewer than 25) to reliably assess the hospital's EDAC."
- Of 4,674 hospitals in the study cohort (data from July 1, 2010 through June 30, 2012), 619 had EDACs "lower than expected," 2,542 were "no different than expected," and 1,007 had EDACs "higher than expected." 506 were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing. The mean EDAC per 100 discharges for hospitals in the top decile of performance is -29.8, compared to 230.0 for hospitals in the bottom decile.
- The developer states that the variation in hospital-level EDAC suggests there are meaningful differences in the quality of care received across hospitals for the pneumonia EDAC measure.

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

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)

2. Scientific Acceptability of Measure Properties

2a1. & 2b1. Specifications

Comments: **The new measure is significantly correlated with the standard readmission measure.

**The metrics rely on a difference between predicted and expected and thus you get excess days (either a positive or negative number).

There is some challenges with calibration of the predicted model as seen in figure two with predicted being higher than observed in the high risk deciles (and lower than observed int he low risk deciles). It is unclear the ramifications for implementing the measure. If the model counts the predicted as worse than actual for high risk, doesn't this model make the high risk hospitals appear to perform worse than actual (observed). Thus aren't this hospitals at risk for increased penalties if this model is implemented?

2a2. Reliability Testing

Comments: **The method for testing reliability appears sound; however, the specific results of the EDAC rates were not clearly identified.

Inclusion of a diagnosis of sepsis with a comorbid condition of pneumonia was not stated for the numerator, although it was stated as a criteria for the denominator.

**The development and validation samples perform similarly. The agreement between the two EDAC values was estimated to be ICC=0.80.

2b2. Validity Testing

Comments: **Authors state that the prior validity testing was completed on the claims-based measures together with a TEP. However, it was not clear whether the TEP indicated that the addition of ED and observation visits was determined to be a valid measure.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **The exclusions appear appropriate. Similar variables were used to risk-adjust the data as found in the standard readmission measures. SDS factors were found to have an effect at the patient level but did not affect the strength of the model at the facility level.

**The two step logit/Poisson model is quite complex and it is challenging to assess validity of the approach.

The data the developers present compares the excess days to readmissions (as well as other unadjusted ops rates, emergency department visits, etc). The challenge with these comparisons is that these are part of the excess day calculation.

Correlation of risk standardized readmission rates and excess days in acute care has a person's of 0.732

No mention of SDS variables.

Criterion 3. Feasibility

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 based on administrative claims data (e.g., DRG, ICD-9/10), which the developers note are

routinely generated and collected as part of hospitals' billing processes.
- The developer indicates that all data elements are in defined fields in electronic claims.

Questions for the Committee:
- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*
- *If an eMeasure, does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?*

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

Committee pre-evaluation comments
Criteria 3: Feasibility

3. Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **The data is readily available and has been thoroughly studied.

**The two step model is quite involved both conceptually and practically unclear how well this could be implemented.

Criterion 4: Usability and Use

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**

OR

Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details
- This measure may be used in one or more CMS programs, such as the Hospital Inpatient Quality Reporting (IQR) program.

Improvement results
- Since this measure is not in use, there are no performance results to assess improvement at this time.
- The developer states that they expect that "there will be improvement in measure scores over time since publicly reported measure scores can reduce adverse patient outcomes associated with days spent in acute care for heart failure by capturing and making acute care utilization following the index hospitalization more visible to providers and patients."

Potential harms
- The developer noted that there were no unintended consequences during development or testing. They are

committed to ongoing monitoring of potential unintended consequences, such as the inappropriate shifting of care, increased patient morbidity and mortality, and other negative intended consequences over time.

Feedback :

- In its 2015-2016 review, MAP conditionally supported this measure, pending NQF review, endorsement, and examination of SDS factors. MAP had concerns about the risk-adjustment methodology and also stated that the Standing Committee reviewing this measure should consider SDS factors that examine the true hospital vs. community role in readmissions and consider parsimony with regard to multiple pneumonia readmission measures.

***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

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **There are no performance results yet for this measure.

Criterion 5: Related and Competing Measures

Related or competing measures

- 0506: Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization

Harmonization

- The developers note that both measures are harmonized.

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: Excess days in acute care (EDAC) after hospitalization for pneumonia

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: 1/29/2016

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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: Single measure: quality outcome measure

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

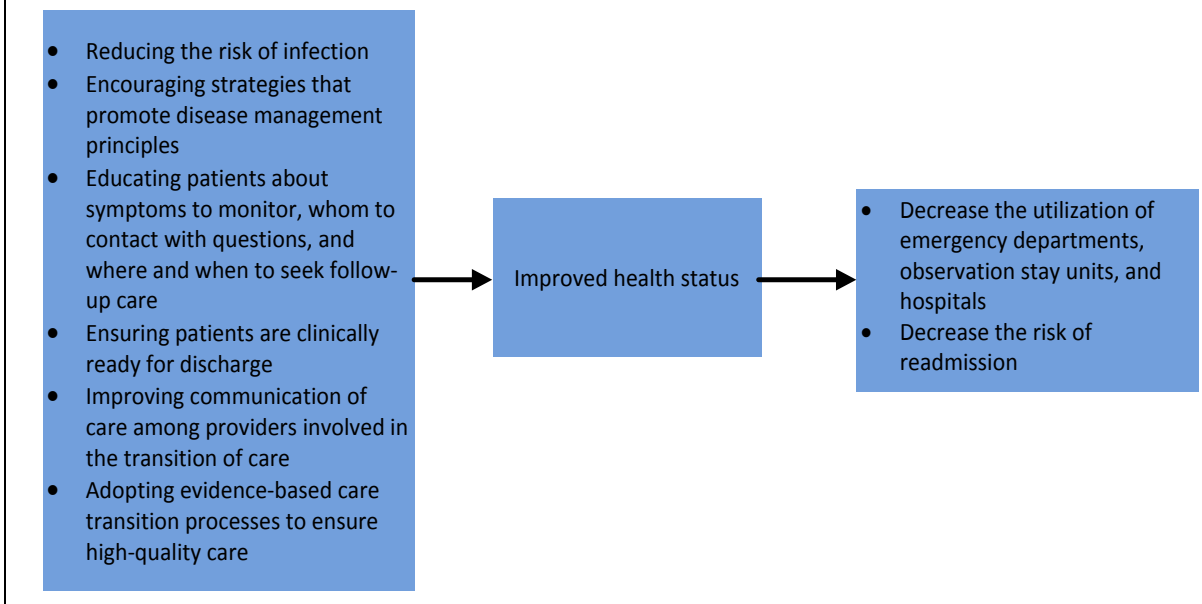
☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.



1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

The diagram above indicates some of the many care processes that can influence post-discharge acute care utilization. These complex and critical aspects of care – such as communication between providers, patient education, patient safety, and coordinated transitions to the outpatient environment – all contribute to patient outcomes but are difficult to measure by individual process measures. Numerous studies have shown that improvement in the following areas can favorably impact utilization rates: communication with patients, caregivers, and other providers; patient education; and quality of care during the initial inpatient admission.

Interventions during and after a hospitalization can be effective in reducing readmission rates in geriatric populations (Benbassat et al., 2000; Naylor et al., 1999; Coleman et al., 2006; Courtney et al., 2009; Koehler et al., 2009) and, particularly, for older patients with pneumonia (Dean et al., 2006). Several randomized trials have reduced 30-day readmission rates by 20-40% (Jack et al., 2009; Coleman et al., 2004; Courtney et al., 2009; Garasen et al., 2007; Koehler et al., 2009; Mistiaen et al., 2007; Naylor et al., 1994; Naylor et al., 1999; van Walraven et al., 2002; Weiss et al., 2010; Krumholz et al., 2012; Balaban et al., 2008). These types of interventions have also been demonstrated to be cost-saving (Naylor et al., 1999; Naylor et al., 2004; Koelling et al., 2005; Krumholz et al., 2002; Stauffer et al., 2011). Outside the randomized controlled trial setting, there is also increasing evidence that hospitals and health plans have been able to reduce readmission rates through more generalizable quality improvement initiatives (Gerhardt et al., 2012; Stauffer et al., 2011; Graham et al., 2012; Harrison et al., 2011; Hernandez et al., 2010).

Studies have also reported reductions in emergency department (ED) visit rates for patients with other conditions after implementation of interventions that focused on the inpatient and outpatient settings (Bondestam et al., 1995).

The current process-based performance measures cannot capture all the ways that care within the hospital might influence outcomes. As a result, many stakeholders, including patient organizations, are interested in outcomes measures that allow patients and providers to assess relative outcomes performance among hospitals (Bratzler et al., 2007).

In the context of the Centers for Medicare and Medicaid Services' (CMS's) publicly reported readmission measures, the increasing use of ED visits and observation stays has raised concerns that current readmission measures do not capture the full range of unplanned acute care in the post-discharge period (Vashi et al., 2013; Rising et al., 2012; Feng et al., 2012). Observation stays can occur in many different parts of the hospital, including dedicated treatment rooms, the ED, or inpatient units. In particular, there is concern that high use of observation stays could in some cases replace readmissions, and that hospitals with high rates of observation stays in the post-discharge period may therefore have low readmission rates that do not accurately reflect the quality of care (Vashi et al., 2013).

Citations

Balaban RB, Weissman JS, Samuel PA, Woolhandler S. Redefining and redesigning hospital discharge to enhance patient care: a randomized controlled study. *J Gen Intern Med.* 2008;23(8):1228-1233.

Benbassat, J., and M. Taragin. 2000. Hospital readmissions as a measure of quality of health care: advantages and limitations. Arch Intern Med 160 (8):1074-81.

Bondestam E, Breikss A, Hartford M. 1995. Effects of early rehabilitation on consumption of medical care during the first year after acute myocardial infarction in patients > or = 65 years of age. Am J Cardiol 75(12):767-771.

Bratzler, DW, Nsa W, Houck PM. Performance measures for pneumonia: are they valuable, and are process measures adequate. Current Opinion in Infectious Diseases. 20(2):182-189, April 2007.

Carlhed R, Bojestig M, Peterson A, et al. Improved clinical outcome after acute myocardial infarction in hospitals participating in a Swedish quality improvement initiative. Circulation. Cardiovascular Quality & Outcomes. 2009;2(5):458-464.

Coleman EA, Parry C, Chalmers S, et al. 2006. The care transitions intervention: results of a randomized controlled trial. Arch Intern Med 166:1822-1828.

Coleman EA, Smith JD, Frank JC, Min SJ, Parry C, Kramer AM. Preparing patients and caregivers to participate in care delivered across settings: the Care Transitions Intervention. J Am Geriatr Soc 2004;52(11):1817-25.

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 2009;57(3):395-402.

Dean NC, Bateman KA, Donnelly SM, Silver MP, Snow GL, Hale D. Improved clinical outcomes with utilization of a community-acquired pneumonia guideline. Chest. 2006;130(3):794-799.

Feng Z, Wright B, Mor V. Sharp rise in Medicare enrollees being held in hospitals for observation raises concerns about causes and consequences. Health affairs (Project Hope). Jun 2012;31(6):1251-1259.

Fisher, ES, JE Wennberg, TA Stukel, et al. 1994. Hospital readmission rates for cohorts of Medicare beneficiaries in Boston and New Haven. N Engl J Med 331 (15):989-95.

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.

Gerhardt G, Yemane A, Hickman P, Oelschlaeger A, Rollins E, Brennan N. Medicare Readmission Rates Showed Meaningful Decline in 2012. *Medicare & Medicaid Research Review.* 2013;3(2):E1-E12.

Gleason PP, Meehan TP, Fine JM, et al. 1999. Associations between initial antimicrobial therapy and medical outcomes for hospitalized elderly patients with pneumonia. Arch Intern Med 159(21):2562-2572.

Graham J, Tomcavage J, Salek D, Sciandra J, Davis DE, Stewart WF. Postdischarge monitoring using interactive voice response system reduces 30-day readmission rates in a case-managed Medicare population. *Medical Care.* 2012;50(1):50-57.

Harrison PL, Hara PA, Pope JE, Young MC, Rula EY. The impact of postdischarge telephonic follow-up on hospital readmissions. *Population Health Management.* 2011;14(1):27-32.

Hernandez AF, Greiner MA, Fonarow GC, et al. Relationship between early physician follow-up and 30-day readmission among Medicare beneficiaries hospitalized for heart failure. *Jama.* 2010;303(17):1716-1722.

Jack BW, Chetty VK, Anthony D, Greenwald JL, Sanchez GM, Johnson AE, et al. A reengineered hospital discharge program to decrease rehospitalization: a randomized trial. Ann Intern Med 2009;150(3):178-87.

Koehler BE, Richter KM, Youngblood L, Cohen BA, Prengler ID, Cheng D, 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. J Hosp Med 2009;4(4):211-218.

Koelling, TM. 2005. Multifaceted outpatient support can improve outcomes for people with heart failure. Commentary. Evid Based Cardiovasc Med 9 Koelling, TM. 2005. Multifaceted outpatient support can improve outcomes for people with heart failure. Commentary. Evid Based Cardiovasc Med 9 (2):138-41.

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.

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

Naylor, MD, Brooten D, Campbell R, et al. 1999. Comprehensive discharge planning and home follow-up of hospitalized elders: a randomized clinical trial. JAMA 281 (7):613-20.

Naylor MD, 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 1994;120(12):999-1006.

Naylor MD, Brooten D, Campbell R, et al. 2004. Transitional care of older adults hospitalized with heart failure: a randomized, controlled trial. J Am Geriatr Soc 52 (5):675-84.

Rising KL, White LF, Fernandez WG, Boutwell AE. Emergency Department Visits After Hospital Discharge: A Missing Part of the Equation. Annals of Emergency Medicine. (0).

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.* 2011;171(14):1238-1243.

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 2002;17(3):186-92.

Vashi AA, Fox JP, Carr BG, et al. Use of hospital-based acute care among patients recently discharged from the hospital. JAMA. Jan 23 2013;309(4):364-371.

Weiss M, Yakusheva O, Bobay K. Nurse and patient perceptions of discharge readiness in relation to postdischarge utilization. Med Care 2010;48(5):482-6.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

N/A. This is an outcome measure.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

N/A. This is an outcome measure.

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

N/A. This is an outcome measure.

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

N/A. This is an outcome measure.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A. This is an outcome measure.

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

N/A. This is an outcome measure.

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

N/A. This is an outcome measure.

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

N/A. This is an outcome measure.

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

N/A. This is an outcome measure.

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

N/A. This is an outcome measure.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A. This is an outcome measure.

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

N/A. This is an outcome measure.

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

N/A. This is an outcome measure.

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

N/A. This is an outcome measure.

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

N/A. This is an outcome measure.

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

N/A. This is an outcome measure.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

N/A. This is an outcome measure.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

N/A. This is an outcome measure.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

N/A. This is an outcome measure.

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

N/A. This is an outcome measure.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

N/A. This is an outcome measure.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

N/A. This is an outcome measure.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

N/A. This is an outcome measure.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

N/A. This is an outcome measure.

1a.8 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.8.1 What process was used to identify the evidence?

N/A. This is an outcome measure.

1a.8.2. Provide the citation and summary for each piece of evidence.

N/A. This is an outcome measure.

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
Pneumonia_Excess_Days_in_Acute_Care_NQF_Measure_Evidence_Form_01-29-2016_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, and hospitals with information about hospital-level, risk-standardized outcomes following hospitalization for pneumonia. Measurement of patient outcomes allows for a broad view of quality of care that cannot be captured entirely by individual process-of-care measures. Safely transitioning patients from hospital to home requires a complex series of tasks which would be cumbersome to capture individually as process measures: timely and effective communication between providers, prevention of and response to complications, patient education about post-discharge care and self-management, timely follow-up, and more. Suboptimal transitions contribute to a variety of adverse events post-discharge, including ED evaluation, need for observation, and readmission. Measures of unplanned readmission already exist, but there are no current measures for ED and observation stay utilization. It is thus difficult for providers and consumers to gain a complete picture of post-discharge outcomes. Moreover, separately reporting each of these outcomes encourages "gaming," such as re-categorizing readmission stays as observation stays to avoid a readmission outcome. By capturing a range of acute care events that are important to patients, we can produce a more complete picture of post-discharge outcomes that better informs consumers about care quality and incentivizes global improvement in transitional care.

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.*
Distribution of EDAC per 100 discharges in the two-year dataset used for measure development:
This analysis includes hospitals that have at least 25 pneumonia index admissions in the two-year period.

Time period//2010-2012
Number of hospitals//3,640
Number of discharges//495,130
Mean EDAC (standard deviation)//8.37 (32.97)
Range (minimum - maximum)//297.58 (-67.59 – 229.99)
Interquartile range//-15.01 – 26.66
Minimum//-67.59
10th percentile//-29.79
20th percentile//-19.74
30th percentile//-11.15
40th percentile//-2.88
50th percentile//4.28
60th percentile//12.54
70th percentile//21.37
80th percentile//32.62
90th percentile//50.56

Maximum//229.99

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

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for 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.*
Distribution of pneumonia EDAC (per 100 discharges) by Proportion of Dual-Eligible Patients:
Dates of Data: July 2010 through June 2012 pneumonia development dataset
Data Source: Medicare FFS claims
Characteristic//Hospitals with a low proportion (=10%) dual-eligible patients//Hospitals with a high proportion (=25.81%) dual-eligible patients
Number of Measured Hospitals//1,187//1,158
Number of Patients//118,183 patients in low-proportion hospitals/87,732 in high-proportion hospitals
Maximum//194.15 //356.31
90th percentile//39.60//59.76
75th percentile//16.02//29.12
Median (50th percentile)//-2.51//4.53
25th percentile//-19.62//-15.02
10th percentile//-30.93//-29.75
Minimum //-67.59//-60.52

Distribution of pneumonia EDAC (per 100 discharges) by Proportion of African-American Patients:
Dates of Data: July 2010 through June 2012 pneumonia development dataset
Data Source: Medicare FFS claims
Characteristic//Hospitals with a low proportion (=0%) African-American patients//Hospitals with a high proportion (=8.05%) African-American patients
Number of Measured Hospitals//1,978 //1,164
Number of Patients//96,720 patients in low-proportion hospitals//144,724 in high-proportion hospitals
Maximum//194.15//356.31
90th percentile//26.94//68.17
75th percentile//6.71//42.18
Median (50th percentile)//-8.27//16.26
25th percentile//-24.52//-5.90
10th percentile//-34.63//-24.02
Minimum//-73.40//-67.59

Low-proportion hospitals are those hospitals whose population of dual-eligible patients or of African-American patients is small enough to place them at or below the 25th percentile among all hospitals; and high proportion are those hospitals whose population of dual eligible patients or African-American patients is large enough to place them at or above the 75th percentile among all hospitals.

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**
High resource use
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.**
Pneumonia results in approximately 1.2 million hospital admissions each year and accounts for more than $10 billion annually in hospital expenditures. Among patients over 65 years of age, it is the second leading cause of hospitalization. Approximately 20% of pneumonia patients were rehospitalized within 30 days, representing the second-highest proportion of all rehospitalizations at 6.3% (Jencks et al., 2009). For the time period of July 2012-June 2013, publicly reported 30-day risk-standardized readmission rates ranged from 13.9% to 22.3% for patients admitted with pneumonia (CMS, 2013).

Rehospitalization, for any reason, is an undesirable outcome, disruptive to patients and caregivers, costly to the healthcare system, and puts patients at additional risk of hospital-acquired infections and complications. Although some readmissions are unavoidable, others may result from poor quality of care or inadequate transitional care. Transitional care includes effective discharge planning, transfer of information at the time of discharge, patient assessment and education, and coordination of care and monitoring in the post-discharge period. Numerous studies have found an association between quality of inpatient or transitional care and early (typically 30-day) readmission rates for a wide range of conditions including pneumonia (Frankl et al., 1991; Corrigan et al., 1992; Oddone et al., 1996; Ashton et al., 1997; Benbassat et al., 2000; Courtney et al., 2003; Halfon et al., 2006; Dean et al., 2006).

Several studies have reported on the relationship between inpatient admissions and other types of hospital care including ED visits and observation stays. ED visits represent a significant proportion of post-discharge acute care utilization. Two recent studies conducted in patients of all ages have shown that 9.5% of patients return to the ED within 30 days of hospital discharge and that about 12.0% of these patients are discharged from the ED and are not captured by current CMS readmissions measures (Rising et al., 2013; Vashi et al., 2013).

Additionally, over the past decade, the use of observation stays has rapidly increased. Specifically, between 2001 and 2008, the use of observation services increased nearly three-fold (Venkatesh et al., 2011) and significant variation has been demonstrated in the use of observation services for conditions such as chest pain (Schuur et al., 2011). These rising rates of observation stays among Medicare beneficiaries have gained the attention of patients, providers, and policymakers (Feng et al., 2012; Hockenberry et al., 2014; Rising et al., 2013; Vashi et al., 2013, Wright B. et al., 2014). A report from the Office of the Inspector General (OIG) notes that in 2012, Medicare beneficiaries had 1.5 million observation stays. Many of these observation stays lasted longer than the intended one day. The OIG report also notes the potential relationship between hospital use of observation stays as an alternative to short-stay inpatient hospitalizations as a response to changing hospital payment incentives (Wright, 2013).

Thus, in the context of CMS's publicly reported readmission measures, the increasing use of ED visits and observation stays has raised concerns that current readmission measures do not capture the full range of unplanned acute care in the post-discharge period. By definition, the readmission measures only assess returns to the hospitals for inpatient stays and not for other acute care services, such as observation stays or ED visits. Stakeholders have expressed concerns about whether observation stays should also be evaluated as markers of the quality of care transitions. In particular, there exists concern that high use of observation stays could in some cases replace readmissions, and hospitals with high rates of observation stays in the post-discharge period may therefore have low readmission rates that do not accurately reflect the quality of care (Carlson, 2013).

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
Ashton CM, Del Junco DJ, Souchek J, Wray NP, Mansyur CL. The association between the quality of inpatient care and early readmission: a meta-analysis of the evidence. Med Care. Oct 1997;35(10):1044-1059.

Benbassat J, Taragin M. Hospital readmissions as a measure of quality of health care: advantages and limitations. Archives of Internal Medicine. Apr 24 2000;160(8):1074-1081.

Carlson J. Faulty Gauge? Readmissions are down, but observational-status patients are up and that could skew Medicare numbers. Modern Healthcare. June 8, 2013 2013.

Centers for Medicare and Medicaid Services (CMS). Medicare Hospital Quality Chartbook Performance Report on Outcome Measures September 2014. September 2014; http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Medicare-Hospital-Quality-Chartbook-2014.

Corrigan JM, Martin JB. Identification of factors associated with hospital readmission and development of a predictive model. Health Serv Res. Apr 1992;27(1):81-101.

Courtney EDJ, Ankrett S, McCollum PT. 28-Day emergency surgical re-admission rates as a clinical indicator of performance. Ann R Coll Surg Engl. Mar 2003;85(2):75-78.

Feng Z, Wright B, Mor V. Sharp rise in Medicare enrollees being held in hospitals for observation raises concerns about causes and consequences. Health affairs (Project Hope). Jun 2012;31(6):1251-1259.

Frankl SE, Breeling JL, Goldman L. Preventability of emergent hospital readmission. Am J Med. Jun 1991;90(6):667-674.

Halfon P, Eggli Y, Pr, et al. Validation of the potentially avoidable hospital readmission rate as a routine indicator of the quality of hospital care. Medical Care. Nov 2006;44(11):972-981.

Hernandez AF, Greiner MA, Fonarow GC, et al. Relationship between early physician follow-up and 30-day readmission among Medicare beneficiaries hospitalized for heart failure. JAMA : the journal of the American Medical Association. May 5 2010;303(17):1716-1722.Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. N Engl J Med. 2009;360(14):1418-28.

Hockenberry JM, Mutter R, Barrett M, Parlato J, Ross MA. Factors Associated with Prolonged Observation Services Stays and the Impact of Long Stays on Patient Cost. Health Serv Res. 2014;49(3):893-909.

Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. N Engl J Med. 2009 Apr 2;360(14):1418-28.

Oddone EZ, Weinberger M, Horner M, et al. Classifying general medicine readmissions. Are they preventable? Veterans Affairs Cooperative Studies in Health Services Group on Primary Care and Hospital Readmissions. Journal of General Internal Medicine. 1996;11(10):597-607.

Rising KL, White LF, Fernandez WG, Boutwell AE. Emergency Department Visits After Hospital Discharge: A Missing Part of the Equation. Annals of Emergency Medicine.

Schuur JD, Baugh CW, Hess EP, Hilton JA, Pines JM, Asplin BR. Critical pathways for post-emergency outpatient diagnosis and treatment: tools to improve the value of emergency care. Academic emergency medicine : official journal of the Society for Academic Emergency Medicine. Jun 2011;18(6):e52-63.

Vashi AA, Fox JP, Carr BG, et al. Use of hospital-based acute care among patients recently discharged from the hospital. JAMA : the journal of the American Medical Association. Jan 23 2013;309(4):364-371.

Venkatesh AK, Geisler BP, Gibson Chambers JJ, Baugh CW, Bohan JS, Schuur JD. Use of observation care in US emergency departments, 2001 to 2008. PloS one. 2011;6(9):e24326.

Wright B., Jung H-Y, Feng Z, Mor V. Hospital, Patient, and Local Health System Characteristics Associated with the Prevalence and Duration of Observation Care. Health Serv Res. 2014;49(4):1088-1107.

Wright S. Hospitals' Use of Observation Stays and Short Inpatient Stays for Medicare Beneficiaries. Washington, DC: OIG; 2013.

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

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):*
Pulmonary/Critical Care : Pneumonia

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Care Coordination : Readmissions, Safety, Safety : Complications, Safety : Readmissions

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:** Pneumonia_Excess_Days_in_Acute_Care_NQF_Data_Dictionary_01-29-16_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 of the measure is a count of the number of days the patient spends in acute care within 30 days of discharge. We define days in acute care as days spent in an ED, admitted to an observation unit, or admitted as an unplanned readmission for any cause within 30 days from the date of discharge from the index pneumonia hospitalization. Each ED treat-and-release visit is counted as one half-day (0.5 days). Observation stays are recorded in terms of hours and are rounded up to the nearest half-day. Each readmission day is counted as one full day (1 day). We count all eligible outcomes occurring in the 30-day period, even if they are repeat occurrences.

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 the measure as within 30 days from the date of discharge of the index pneumonia hospitalization.

Denominator Time Window: The measure was developed and will be reported using two 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.

Outcome Definition

The measure counts ED treat-and-release visits, observation stays, and readmissions to any acute care hospital for any cause within 30 days of the date of discharge of the index pneumonia admission, excluding planned readmissions as defined below.

All events which occur within the 30-day window are counted. For example, if a patient returns to the ED three times on three different days, we count each ED visit as a half-day. Similarly, if a patient has two hospitalizations within 30 days, the days spent in each are counted. Therefore, the measure may include multiple ED visits, observation stays, and/or readmissions per patient.

The measure incorporates "exposure time" (the number of days each patient survives after discharge, up to 30). This exposure time is included to account for differential risk for EDAC after discharge among those patients who do not survive the full post-discharge period. If a hospitalization or observation stay extends beyond the 30-day window, only those days within the 30-day window are counted.

Planned Readmission Algorithm

The Planned Readmission Algorithm is a set of criteria for classifying readmissions as planned among the general Medicare population using Medicare administrative claims data. The algorithm identifies admissions that are typically planned and may occur within 30 days of discharge from the hospital.

The Planned Readmission Algorithm has three fundamental principles:
1. A few specific, limited types of care are always considered planned (obstetric delivery, transplant surgery, maintenance chemotherapy/radiotherapy/ immunotherapy, rehabilitation);
2. Otherwise, a planned readmission is defined as a non-acute readmission for a scheduled procedure; and
3. Admissions for acute illness or for complications of care are never planned.

The algorithm was developed in 2011 as part of the Hospital-Wide Readmission measure. In 2013, CMS applied the algorithm to its other readmission measures. In applying the algorithm to condition- and procedure-specific measures, teams of clinical experts reviewed the algorithm in the context of each measure-specific patient cohort and, where clinically indicated, adapted the content of the algorithm to better reflect the likely clinical experience of each measure's patient cohort. For the CMS 30-day pneumonia EDAC measure, CMS used the Planned Readmission Algorithm without making any changes.

For development of this measure, we used the Planned Readmission Algorithm, Version 3.0. This version and associated code tables are attached in data field S.2b (Data Dictionary or Code Table). For reporting purposes, the measure will use the next version of the Planned Readmission Algorithm, Version 4.0, as will be used in the CMS 30-day pneumonia readmission measure.

Definition of Emergency Department Visit and Observation Stay

We defined ED visits and observation stays using specified billing codes or revenue center codes identified in Medicare hospital outpatient claims and physician Carrier claims. The codes that define ED visits and observation stays are in the attached Data Dictionary.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*

The target population for this measure is Medicare FFS beneficiaries aged 65 years and older hospitalized at non-Federal acute care hospitals for pneumonia.

The cohort includes admissions for patients discharged from the hospital with a principal discharge diagnosis of pneumonia (see codes below in S.9) and with continuous 12 months Medicare enrollment prior to admission. The measure will be publicly reported by CMS for those patients 65 years and older who are Medicare FFS beneficiaries admitted to non-federal hospitals.

Additional details are provided n S.9 Denominator Details.

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

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)*
To be included in the measure cohort used in public reporting, patients must meet the following inclusion criteria:
1. Principal discharge diagnosis of pneumonia, including aspiration pneumonia; or
Principal discharge diagnosis of sepsis (not including severe sepsis), with a secondary discharge diagnosis of pneumonia (including aspiration pneumonia) coded as POA but no secondary discharge diagnosis of severe sepsis.
2. Enrolled in Medicare fee-for-service (FFS) Part A and Part B for the 12 months prior to the date of the admission, and enrolled in Part A during the index admission;
3. Aged 65 or over;
4. Discharged alive from a non-federal short-term acute care hospital; and,
5. Not transferred from another acute care facility.

International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) codes used to define the cohort for the measure are:
480.0 Pneumonia due to adenovirus
480.1 Pneumonia due to respiratory syncytial virus
480.2 Pneumonia due to parainfluenza virus
480.3 Pneumonia due to SARS-associated coronavirus
480.8 Pneumonia due to other virus not elsewhere classified
480.9 Viral pneumonia, unspecified
481 Pneumococcal pneumonia
482.0 Pneumonia due to Klebsiella pneumoniae
482.1 Pneumonia due to Pseudomonas
482.2 Pneumonia due to Hemophilus influenzae
482.30 Pneumonia due to Streptococcus, unspecified
482.31 Pneumonia due to Streptococcus, group A
482.32 Pneumonia due to Streptococcus, group B
482.39 Pneumonia due to other Streptococcus
482.40 Pneumonia due to Staphylococcus, unspecified
482.41 Methicillin susceptible pneumonia due to Staphylococcus aureus
482.42 Methicillin resistant pneumonia due to Staphylococcus aureus
482.49 Other Staphylococcus pneumonia
482.81 Pneumonia due to anaerobes
482.82 Pneumonia due to escherichia coli [E. coli]
482.83 Pneumonia due to other gram-negative bacteria
482.84 Pneumonia due to Legionnaires´ disease
482.89 Pneumonia due to other specified bacteria
482.9 Bacterial pneumonia, unspecified
483.0 Pneumonia due to mycoplasma pneumoniae
483.1 Pneumonia due to chlamydia
483.8 Pneumonia due to other specified organism
485 Bronchopneumonia, organism unspecified
486 Pneumonia, organism unspecified
487.0 Influenza with pneumonia
488.11 Influenza due to identified 2009 H1N1 influenza virus with pneumonia

ICD-9 codes that define patients with aspiration pneumonia:
507.0 Pneumonitis due to inhalation of food or vomitus

ICD-9 codes that define patients with sepsis (not including severe sepsis [995.92 or 785.52]) (Cohort requires principal discharge

diagnosis of sepsis combined with a secondary discharge diagnosis of pneumonia or aspiration pneumonia coded as POA but no secondary discharge diagnosis of severe sepsis):
038.0 Streptococcal septicemia
038.10 Staphylococcal septicemia, unspecified
038.11 Methicillin susceptible Staphylococcus aureus septicemia
038.12 Methicillin resistant Staphylococcus aureus septicemia
038.19 Other staphylococcal septicemia
038.2 Pneumococcal septicemia
038.3 Septicemia due to anaerobes
038.40 Septicemia due to gram-negative organism, unspecified
038.41 Septicemia due to hemophilus influenzae
038.42 Septicemia due to escherichia coli [E. coli]
038.43 Septicemia due to pseudomonas
038.44 Septicemia due to serratia
038.49 Other septicemia due to gram-negative organisms
038.8 Other specified septicemias
038.9 Unspecified septicemia
995.91 Sepsis

An ICD-9 to ICD-10 crosswalk is attached in field S.2b. (Data Dictionary or Code Table).

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
The measure excludes index admissions for patients:
1. Without at least 30 days post-discharge enrollment in FFS Medicare.
2. Discharged against medical advice (AMA);
3. Admitted within 30 days of a prior index discharge;

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)*
1. Admissions without at least 30 days post-discharge enrollment in FFS Medicare are determined by examining the Medicare Enrollment Database (EDB).

2. Discharges against medical advice (AMA) are identified using the discharge disposition indicator in claims data.

3. Pneumonia admissions within 30 days of discharge from a qualifying pneumonia index admission are identified by comparing the discharge date from the index admission with subsequent admission dates.

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)*
Not applicable. This measure is not stratified.

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).

For risk-adjustment, we used a hierarchical generalized linear model (HGLM). This model consists of two parts, a logit model and a

truncated Poisson model. The two-part logit/Poisson model (often called a "hurdle" model) assumes that the outcome results from two related processes: an initial dichotomous event – that a patient has at least one acute care event – which is modeled as the logit of the probability of the event, and for patients with an event (those which clear the "hurdle"), the number of days, which is modeled as a Poisson process. The outcome, number of days, is a half-integer count variable (because ED visits count as 0.5 days). Observation care is counted according to the hours spent in observation care, rounded up to the nearest half-day. For each patient, an exposure variable is defined as the number of survival days post discharge, up to 30. For the hurdle model, exposure time as an offset is included for each part of the model.

There are two random effects for each hospital, one for the logit model and one for the truncated Poisson model, as well as a covariance between the two random effects. The random effects allow us to account for within-hospital correlation of the observed outcome and accommodates the assumption that underlying differences in quality across hospitals lead to systematic differences in outcomes.

We use the current CMS 30-day pneumonia readmission measure final risk-adjustment variables. We verified the adequacy of this risk-adjustment strategy for our new outcome by comparing the discrimination of models with a full set of all comorbidities to the more parsimonious existing risk models. We found no improvement in model discrimination with the full set, indicating that the existing risk models are adequate.

The measures adjust for variables (i.e., age, comorbid diseases, and indicators of patient frailty) that are clinically relevant and have strong relationships with the outcome. For each patient, risk-adjustment variables are obtained from inpatient, outpatient, and physician Medicare administrative claims data extending 12 months prior to, and including, the index admission.

The model adjusts for case-mix differences based on the clinical status of patients 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 (Pope et al., 2000). A file that contains a list of the ICD-9-CM codes and their groupings into CCs is attached in data field S.2b (Data Dictionary or Code Table). In addition, only comorbidities that convey information about the patient at admission or in the 12 months prior, and not complications that arise during the course of the index hospitalization, are included in the risk adjustment. Hence, we do not risk adjust for CCs that may represent adverse events of care and that are only recorded in the index admission.

The final set of risk-adjustment variables includes the following:
Demographics:
1. Male
2. Age (defined as "Age-65" [years above 65, continuous])

Comorbidities:
3. History of Coronary Artery Bypass Graft (CABG) (ICD-9-CM V45.81, 36.10–36.16)
4. History of infection (CC 1, 3-6)
5. Septicemia/shock (CC 2)
6. Metastatic cancer or acute leukemia (CC 7)
7. Lung, upper digestive tract, and other severe cancers (CC 8)
8. Other major cancers (CC 9-10)
9. Diabetes Mellitus (DM) or DM complications (CC 15-20, 119, 120)
10. Protein-calorie malnutrition (CC 21)
11. Disorders of fluid, electrolyte, acid-base (CC 22, 23)
12. Other gastrointestinal disorders (CC 36)
13. Severe hematological disorders (CC 44)
14. Iron deficiency or other unspecified anemias and blood disease (CC 47)
15. Dementia or other specified brain disorders (CC 49, 50)
16. Drug/alcohol abuse/dependence/psychosis (CC 51-53)
17. Major psychiatric disorders (CC 54-56)
18. Other psychiatric disorders (CC 60)
19. Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177, 178)
20. Cardio-respiratory failure or shock (CC 78, 79)
21. Congestive heart failure (CC 80)
22. Acute coronary syndrome (CC 81, 82)

23. Coronary atherosclerosis or angina (CC 83, 84)
24. Valvular or rheumatic heart disease (CC 86)
25. Specified arrhythmias and other heart rhythm disorders (CC 92, 93)
26. Stroke (CC 95, 96)
27. Vascular or circulatory disease (CC 104-106)
28. Chronic obstructive pulmonary disease (CC 108)
29. Fibrosis of lung and other chronic lung disorders (CC 109)
30. Asthma (CC 110)
31. Pneumonia (CC 111-113)
32. Pleural effusion/pneumothorax (CC 114)
33. Other lung disorders (CC 115)
34. End-stage renal disease or dialysis (CC 129, 130)
35. Renal failure (CC 131)
36. Urinary tract infection (CC 135)
37. Other urinary tract disorders (CC 136)
38. Decubitus ulcer or chronic skin ulcer (CC 148, 149)
39. Vertebral fractures (CC 157)
40. Other injuries (CC 162)
41. Respirator dependence/Tracheostomy (CC 77)

References:
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.

Pope GC, et al. 2000. Principal Inpatient Diagnostic Cost Group Models for Medicare Risk Adjustment. Health Care Financing Review 21(3): 93-118.

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*)

S.16. **Type of score:**
Other (specify):
If other: Excess days in acute care (EDAC) per 100 discharges

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.)*
As described above, we used a hierarchical generalized linear model (HGLM). This consists of the two-part logit/truncated Poisson model specifications for days in acute care and includes two random effects for hospitals – one for the logit part and one for the truncated Poisson part – with a non-zero covariance between the two random effects.

This model is used to estimate predicted and expected values for each patient. Predicted values are model predictions that include the hospital random effects, and expected values are model predictions that do not include the hospital random effects. We

describe calculation of the predicted and expected values in the attached Appendix (Section 2.7). The measure reports, for each hospital, the difference ("excess") between each hospital's patients' average days in acute care ("predicted days"), and the number of days in acute care that they would have been expected to spend if discharged from an average performing hospital ("expected days").To be consistent with the reporting of the CMS 30-day pneumonia readmission measure, we have multiplied the final score by 100 so that the reported EDAC represents EDAC per 100 discharges.

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.
Administrative claims

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:
1. Medicare Part A inpatient, Part B hospital outpatient claims and physician carrier claims data: 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.

For development purposes, we obtained the Medicare Part B hospital and physician outpatient claims from the Chronic Condition Data Warehouse (CCW) 100% condition-specific datasets.

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

Reference:
Fleming C, Fisher ES, Chang CH, Bubolz TA, Malenka DJ. 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. Data sources for the all-payer update

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/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

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
Pneumonia_Excess_Days_in_Acute_Care_NQF_Measure_Testing_Form_01-29-2016_v1.1.docx

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

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Excess days in acute care (EDAC) after hospitalization for pneumonia
Date of Submission: 1/29/2016
Type of Measure:

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

The datasets used for testing included Medicare Parts A and B claims, the Medicare Enrollment Database (EDB), and the Chronic Condition Data Warehouse (CCW) 100% condition-specific dataset to capture ED visits and observation stays.

The specific 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 July 1, 2010 through June 30, 2012.

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

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician

☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

For this measure, hospitals are the measured entities. All non-Federal, acute inpatient hospitals in the United States ([US] including territories) with Medicare Fee-for-Service (FFS) beneficiaries over the age of 65 are included. See Section 1.7 for details

1.6. How many and which patients 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 patients and discharges varies by testing type and samples used. See Section 1.7 for the uses of the development sample and validation sample.

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 hospitals, and number of admissions used in each type of testing are as follows:

For reliability testing (Section 2a2):
The reliability of the model was tested by randomly selecting 50% of the Medicare patients aged 65 years or older in a two-year cohort (July 1, 2010-June 30, 2012) and developing a risk-adjusted model for this group (the "development sample"). We then developed a second model for the remaining 50% of patients (the "validation sample") and compared the two.

The development sample consisted of:
Number of discharges: 495,130
Number of hospitals: 4,655
Patient descriptive characteristics: average (standard deviation [SD]) age = 80.7 (8.3); % male = 46.3%

The validation sample consisted of:
Number of discharges: 495,130
Number of hospitals: 4,663
Patient descriptive characteristics: average (SD) age = 80.7 (8.3); % male = 46.2%

We used the development sample for calculation of performance score (Section 1b2), model selection (2b4), testing of disparities (section 1b4), reliability testing (Section 2a2), empirical validity testing (Section 2b2),

testing of measure exclusions (Section 2b3), testing of measure risk adjustment (Section 2b4), and testing to identify meaningful differences in performance (Section 2b5). We also used the development sample to examine 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 (2b4.4b).

We used the validation sample for testing of measure risk adjustment (Section 2b4), and data element and performance measure reliability (Section 2a2).

Data Elements:
• African-American race and dual- eligible status (i.e., enrolled in both Medicare and Medicaid) patient-level data are obtained from Centers for Medicare and Medicaid Services (CMS) enrollment 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. 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 pneumonia is more limited, here too though studies suggest a possible increased risk of readmission in particular with the inclusion of race variables (Calvillo-King et al., 2013; Joynt et al., 2011; Lindenauer et al., 2013; McHugh et al., 2010; Mather et al., 2014; Vidic et al., 2015). 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
• African-American race

In selecting variables, our intent was to be responsive to the National Quality Forum (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.

References

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. J Gen Intern Med. 2013 Feb; 28(2):269-82. doi: 10.1007/s11606-012-2235-x. Epub 2012 Oct 6.

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.

Gilman M, Adams EK, Hockenberry JM, et al. 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 KE, Orav EJ, Jha AK. Thirty-day readmission rates for Medicare beneficiaries by race and site of care. JAMA. 2011 Feb 16; 305(7):675-81. doi: 10.1001/jama.2011.123.

Lindenauer PK, Lagu T, Rothberg MB, et al. Income inequality and 30 day outcomes after acute myocardial infarction, heart failure, and pneumonia: retrospective cohort study. BMJ. 2013 Feb 14; 346:f521. doi: 10.1136/bmj.f521.

Mather JF, Fortunato GJ, Ash JL, et al. Prediction of pneumonia 30-day readmissions: a single-center attempt to increase model performance. Respir Care. 2014 Feb; 59(2):199-208. doi: 10.4187/respcare.02563. Epub 2013 Aug 13.

McHugh MD, Carthon JM, Kang XL. Medicare readmissions policies and racial and ethnic health disparities: a cautionary tale. Policy Polit Nurs Pract. 2010 Nov; 11(4):309-16. doi: 10.1177/1527154411398490.

Vidic A, Chibnall JT, Hauptman PJ. Heart failure is a major contributor to hospital readmission penalties. J Card Fail. 2015 Feb; 21(2):134-7. doi: 10.1016/j.cardfail.2014.12.002. Epub 2014 Dec 9.

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 aimed to utilize only those data elements from the claims that have both face validity and reliability. We used the final risk-adjustment variables in the current CMS 30-day pneumonia readmission measure.

We avoided the use of fields that are thought to be coded inconsistently across facilities. Specifically, we used fields that are consequential for payment and which are audited. We identified such variables through empiric analyses and our understanding of the CMS auditing and billing policies. We sought to avoid variables which do not meet these standards.

In addition, CMS has in place several hospital auditing programs used to assess overall accuracy of claims-based coding, to ensure appropriate billing, and for overpayment recoupment. CMS routinely conducts data analysis to identify potential problem areas and to 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 assessed the reliability of the data elements by comparing variable frequencies between our development sample and validation sample.

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, is measured again using a second random subset exclusive of the first, and then the agreement between the two resulting performance measures across hospitals is calculated (Rousson et al., 2002).

For test-retest reliability, we calculated the EDAC for each hospital using first the development sample, then the validation sample. Thus, we measured each hospital twice, each time 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) as defined by ICC[2,1] by Shrout and Fleiss (1979) and assessed the values according to conventional standards (Landis and Koch, 1977). We restricted this calculation to hospitals with at least 8 discharges in both samples to approximate the set of hospitals that would have at least 24 discharges over three years and are thus likely to be included in public reporting.

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. In addition, using our split-sample datasets underestimates the test-retest reliability that would be achieved if the measure were reported using three years of data, because the smaller samples for each hospital in one year of data are less reliable. To correct for this underestimate, we used the Spearman-Brown prophecy formula (Spearman 1910, Brown 1910) to adjust the ICC[2,1] to represent three years of data.

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

Risk variable	Development sample (N=495,130)		Validation sample (N=495,130)	
	n	%	n	%
Age, continuous (mean [SD])	80.7		80.7	
Male	228,999	46.3	228,940	46.2
History of Coronary Artery Bypass Graft (CABG) surgery (ICD-9 codes V45.81, 36.10-36.16)	38,126	7.7	38,187	7.7
History of infection (CC 1, 3-6)	207,779	42.0	207,928	42.0
Septicemia/sepsis (CC 2)	55,074	11.1	55,134	11.1
Metastatic cancer or acute leukemia (CC 7)	24,416	4.93	24,539	5.0
Lung, upper digestive tract, and other severe cancers (CC 8)	31,894	6.4	32,522	6.6
Other major cancers (CC 9-10)	86,889	17.5	86,549	17.5
Diabetes mellitus (DM) or DM complications (CC 15-19, 119-120)	208,227	42.1	209,026	42.2
Protein-calorie malnutrition (CC 21)	82,422	16.7	82,356	16.6
Disorders of fluid/electrolyte/acid-base (CC 22-23)	213,630	43.2	213,633	43.2
Other gastrointestinal disorders (CC 36)	322,899	65.2	323,103	65.3
Severe hematological disorders (CC 44	19,744	4.0	19,925	4.0
Iron deficiency or other unspecified anemias and blood disease (CC 47)	292,853	59.2	293,047	59.2

Risk variable	Development sample (N=495,130)		Validation sample (N=495,130)	
	n	%	n	%
Dementia or other specified brain disorders (CC 49-50)	184,947	37.4	184,316	37.2
Drug/alcohol abuse/dependence/psychosis (CC 51-53)	68,875	13.9	68,340	13.8
Major psychiatric disorders (CC 54-56)	77,228	15.6	77,203	15.6
Other psychiatric disorders (CC 60)	87,425	17.7	87,898	17.8
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)	57,011	11.5	57,197	11.6
Respiratory dependence/tracheostomy (CC 77)	6,481	1.3	6,539	1.3
Cardio-respiratory failure or shock (CC 78-79)	113,939	23.0	113,568	22.9
Congestive heart failure (CC 80)	198,870	40.2	198,355	40.1
Acute coronary syndrome (CC 81-82)	38,953	7.9	38,666	7.8
Coronary atherosclerosis or angina (CC 83-84)	243,961	49.3	242,428	49.0
Valvular or rheumatic heart disease (CC 86)	121,938	24.6	121,575	24.6
Specified arrhythmias and other heart rhythm disorders (CC 92-93)	220,658	44.6	219,802	44.4
Stroke (CC 95-96)	58,683	11.9	58,753	11.9
Vascular or circulatory disease (CC 104-106)	219,450	44.3	219,381	44.3
Chronic obstructive pulmonary disease (COPD) (CC 108)	263,957	53.3	264,218	53.4
Fibrosis of lung or other chronic lung disorders (CC 109)	80,512	16.3	80,706	16.3
Asthma (CC 110)	53,542	10.8	53,738	10.9
Pneumonia (CC 111-113)	256,784	51.9	256,518	51.8
Pleural effusion/pneumothorax (CC 114)	86,950	17.6	87,147	17.6
Other lung disorders (CC 115)	234,450	47.4	234,737	47.4
End-stage renal disease or dialysis (CC 129-130)	14,783	3.0	14,634	3.0
Renal failure (CC 131)	146,689	29.6	147,192	29.7
Urinary tract infection (CC 135)	162,238	32.8	162,059	32.7
Other urinary tract disorders (CC 136)	126,286	25.5	126,398	25.5
Decubitus ulcer or chronic skin	69,781	14.1	69,657	14.1

Risk variable	Development sample (N=495,130)		Validation sample (N=495,130)	
	n	%	n	%
ulcer (CC 148-149)				
Vertebral fractures (CC 157)	25,867	5.2	26,173	5.3
Other injuries (CC 162)	202,696	40.9	201,401	40.7

Measure Score Reliability Results
The agreement between the two EDAC values for each hospital was estimated for three years to be ICC[2,1] = 0.80 which according to the conventional interpretation is "substantial" (Landis & Koch, 1977).

Reference
Landis J, Koch G. The measurement of observer agreement for categorical data, Biometric. 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 results are consistent with existing hospital-level measures of patient outcomes. Compared to the development sample, the mean age of patients and the frequencies of the risk-adjustment variables were very similar in the validation sample; this indicates that the data elements are reliable. The ICC [2,1] score of 0.80, estimated for three years of data, demonstrates substantial agreement between samples across the full range of measure values. We interpret this to mean that when used with a full three years of data, the measure will be reliable by the standards of hospital measurement.

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)*

We demonstrated measure validity through relevant prior validity testing that we conducted for other claims-based measures, through use of established measure development guidelines, through assessment by external groups, and through systematic assessment of measure face validity by technical expert panel (TEP) of national experts and stakeholder organizations.

Empirical Validity Testing
This measure is closely related in design to the existing readmission measure for patients with pneumonia. While the current measure includes additional endpoints and measures them in a different metric (days rather than rates), we would expect that hospitals would have similar – though not identical – performance rankings on

the two measures. Thus, as one assessment of validity, we compared the rankings of all hospitals using the two measures to assess the consistency of hospital performance on closely related outcomes. We calculated the Pearson correlation, and graphed the readmission measure against the EDAC measure to determine if there were outliers.

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 six NQF-endorsed measures currently in public reporting (acute myocardial infarction [AMI], heart failure, 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) (Krumholz et al. 2006; Keenan et al. 2008), AMI patients (Cooperative Cardiovascular Project data) (Krumholz, Wang, et al. 2006), and pneumonia patients (National Pneumonia Project dataset) (Bratzler et al. 2011). 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, supporting the use of the claims-based models for public reporting. This measure uses the same risk-adjustment variables that were previously validated in the chart review studies.

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 (National Quality Forum, 2010), CMS Measure Management System guidance, and the guidance articulated in the American Heart Association scientific statement, “Standards for Statistical Models Used for Public Reporting of Health Outcomes” (Krumholz, Brindis, et al. 2006).

Validity as Assessed by External Groups

Throughout measure development, we obtained expert and stakeholder input via three mechanisms in the initial, early phase of development: a discussion with an advisory Methodology Workgroup, discussions with a national Technical Expert Panel (TEP), and a 30-day public comment period in order to increase transparency and to gain broader input on the measure.

The Methodology Workgroup meeting addressed key issues related to measure methodology, including weighing the pros and cons of and measure specifications, modeling, and use (e.g., defining the measure cohort and outcome) to ensure the measure is meaningful, useful, and well-designed. The group provided a forum for focused expert review and discussion of technical issues during measure development.

List of Methodology Workgroup Members:
1) Arlene Ash, PhD; University of Massachusetts Medical School (Professor and Division Chief)
2) Jeremiah Brown, MS, PhD; The Dartmouth Institute for Health Policy and Clinical Practice (Assistant Professor of Health Policy and Clinical Practice)
3) Grant Ritter, PhD, MS, MA; Schneider Institute for Health Policy & Heller Graduate School (Senior Scientist)
4) Patrick Romano, MD, MPH; University of California Davis School of Medicine (Professor of Medicine and Pediatrics)

In alignment with the CMS Measures 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 physicians, consumers, purchasers, as well as individuals with experience in quality improvement, performance measurement, and health care disparities. We held two structured TEP conference calls consisting of a presentation of key issues, our proposed approach, and relevant data, followed by open discussion among TEP members. We solicited additional input and comments from the TEP via e-mail between meetings.

Following completion of the preliminary model, we solicited public comment on the measure through the CMS site link http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/CallforPublicComment.html. The public comments were then posted publicly for 30 days. The resulting input was taken into consideration during the final stages of measure development, and led to supplementary analyses reported in the application (1b.4).

Face Validity as Determined by Technical Expert Panel

One means of confirming the validity of this measure was face validity assessed by our TEP, which included 16 members, including patient representatives, expert clinicians, researchers, providers, and purchasers.

List of TEP members:
1) Kevin E. Driesen, PhD, MPH, MA; Center for Rural Health Mel and Enid Zuckerman College of Public Health, University of Arizona (Assistant Professor & Director of the Arizona Rural Hospital Flexibility Program)
2) David Engler, PhD; America's Essential Hospitals (Senior Vice President for Leadership and Innovation)
3) Timothy Farrell, MD; University of Utah School of Medicine (Assistant Professor of Medicine, Geriatrics; Adjunct Professor of Family Medicine)
4) Karen Farris, PhD; University of Michigan College of Pharmacy (Charles R. Walgreen III Professor of Pharmacy Administration; Director of the Social and Administrative Pharmacy Graduate Program)
5) Maura C. Feldman, MSW; Blue Cross Blue Shield of Massachusetts, Inc (Director for Hospital Performance Measurement and Improvement)
6) Jay A. Gold, MD, JD, MPH; Meta Star, Inc. (Vice President & Chief Medical Officer)
7) Sally Hinkle, DNP, MPA, RN; Temple University Hospital (Director of Performance Improvement & Clinical Value)
8) Amy J.H. Kind, MD, PhD; University of Wisconsin School of Medicine and Public Health (Assistant Professor of Geriatrics)
9) Marjorie King, MD, FACC, MAACVPR; Helen Hayes Hospital (Director of Cardiac Services)
10) Eugene Kroch, PhD; University of Pennsylvania (Adjunct Faculty at the Health Care Systems Department); Premier, Inc. (Vice President & Chief Scientist) University of Pennsylvania; Philadelphia, PA
11) Keith D. Lind, JD, MS, BSN; American Association of Retired Persons (AARP) Public Policy Institute (Senior Policy Advisor)
12) Grace McConnell, PhD; Patient representative
13) Michael A. Ross, MD, FACEP; Emory University School of Medicine (Medical Director of Observation Medicine and Chest Pain Center; Professor of Emergency Medicine)
14) Mark Louis Sanz, MDl; International Heart Institute of Montana (Interventional Cardiologist)
15) Paul Takahashi, MD; Mayo Clinic College of Medicine (Associate Professor of Medicine)
16) Patient representative

We systematically assessed the face validity of the measure score as an indicator of quality by soliciting the TEP members' agreement with the following statement: "*The risk-standardized acute care days obtained from the measures as specified can be used to distinguish between better and worse quality hospitals.*"

We measured agreement on a six-point scale: 1=Strongly disagree, 2=Moderately disagree, 3=Somewhat disagree, 4=Somewhat agree, 5=Moderately agree, 6=Strongly agree.

Process Used to Identify International Classification of Diseases, Tenth Revision (ICD-10) Codes Statement of Intent

[X] Goal was to convert this measure to a new code set, fully consistent with the intent of the original measure.
[ ] Goal was to take advantage of the more specific code set to form a new version of the measure, but fully consistent with the original intent.
[ ] The intent of the measure has changed.

Process of Conversion

ICD-10 codes were initially identified using 2015 General Equivalence Mappings (GEM) 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. An ICD-9 to ICD-10 crosswalk is attached in field S.2b. (Data Dictionary or Code Table).

Citations

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

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

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

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

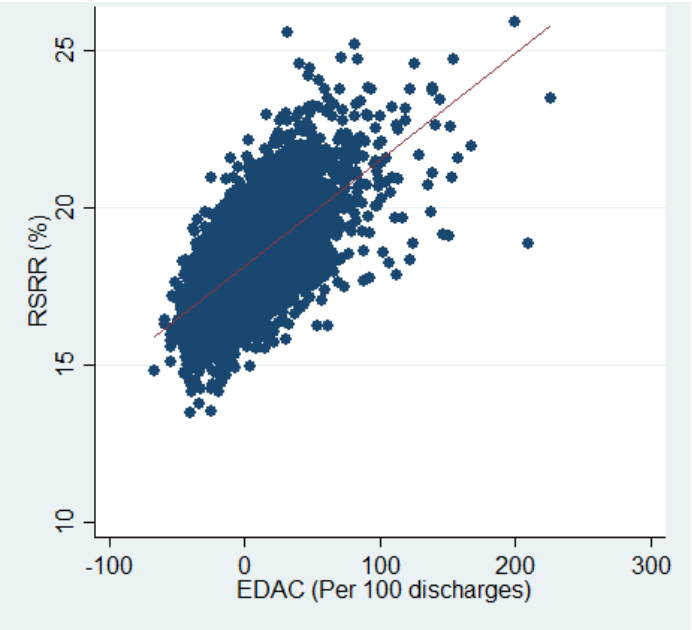
National Quality Forum. National voluntary consensus standards for patient outcomes, first report for phases 1 and 2: A consensus report http://www.qualityforum.org/projects/Patient_Outcome_Measures_Phases1-2.aspx. Accessed August 19, 2010.

Krumholz HM, Brindis RG,Brush JE, et al. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. *Circulation.* January 24, 2006 2006;113(3):456-462.

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

Empirical Validity Testing

Comparison of the new measure with the existing CMS 30-day pneumonia readmission measure found a Pearson Correlation of 0.732 ($P < 0.0001$). The following figure shows the relationship between risk-standardized readmission rate (RSRR) and EDAC for pneumonia:



Systematic Assessment of Face Validity

The results of the TEP rating of agreement with the validity statement were as follows:

N=12

Mean rating=5

Rating	# of Responses	Percent (%)	Cumulative Percent (%)
6 (Strongly agree)	4	33.3%	33.3%
5 (Moderately agree)	6	50.0%	83.3%
4 (Somewhat agree)	1	8.3%	91.7%
3 (Somewhat disagree)	0	0.0%	91.7%

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?*)

Empirical Validity Testing

There was substantial correlation between the two hospital measures, indicating that the proposed measure and the existing readmission measure share underlying properties. This result, and the notable lack of outliers in the figure, provide external empirical validity.

Validity as Assessed by External Groups

The face validity testing results demonstrated TEP agreement with overall face validity of the measure as specified.

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*)

All exclusions were determined by careful clinical review and have been made based on clinically relevant considerations. To ascertain impact of exclusions on the cohort, we examined overall frequencies and proportions of the total cohort excluded for all exclusions, and examined distributions 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. Rationales for the exclusions are detailed in data field S.10 (Denominator Exclusions).

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*)

We examined overall frequencies and proportions of the admissions excluded for each criterion in all pneumonia admissions from July 1, 2010 to June 30, 2012.

The exclusion categories are not mutually exclusive.
1. Discharged patients without at least 30 days post-discharge information (0.6%)
2. Discharges against medical advice (AMA) (0.3%)
3. Admissions within 30 days of a prior index admission (4.5%)

Exclusion	N	%	Distribution across hospitals with ≥ 25 discharges (N=4,196): Minimum, 25th percentile, 50th percentile, 75th percentile, maximum
1. Without at least 30 days post-discharge enrollment in FFS Medicare for index admissions	6,237	0.6	(0.0, 0.0, 0.4, 0.9, 14.7)
2. Discharged against medical advice (AMA)	2,636	0.3	(0.0, 0.0, 0.0, 0.3, 8.3)
3. Admission within 30 days of a prior pneumonia index admission	46,485	4.5	(0.0, 2.9, 4.2, 5.4, 24.1)

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*)

Exclusion 1 (patients without at least 30 days post-discharge enrollment in FFS Medicare for index admissions in non-VA hospitals) accounts for 0.6% of all index admissions excluded from the initial index cohort. This exclusion is needed since the outcome cannot be assessed in this group since claims data are used to determine whether a patient returned to the hospital for an ED visit, was placed under observation care, or was readmitted.

Exclusion 2 (patients who are discharged AMA) accounts for 0.3% of all index admissions excluded from the initial index cohort. This exclusion is needed for acceptability of the measure to hospitals, who do not have the opportunity to deliver full care and prepare the patient for discharge. Because of a very small percent of patients being excluded it is unlikely to affect measure score.

Exclusion 3 (patients with admissions within 30 days of a prior index admission) accounts for 4.5% of all index admissions excluded from the initial index cohort. This criterion varies more substantially among hospitals. Since risk of readmission is much higher after a hospitalization that is itself a readmission than after a hospitalization with no recent prior admission, including readmissions as index cases could distort the performance results. Accordingly, we restrict the sample only to hospitalizations with no recent prior admissions to ensure apples-to-apples comparisons across hospitals.

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 41 risk factors**

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

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

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (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 is tailored to, and appropriate for, a publicly reported outcome measure as articulated in published scientific guidelines (Krumholz et al. 2006, Normand et al. 2007). We adopted the risk factors from the existing CMS 30-day pneumonia readmission measure (Lindenauer et al. 2015). These risk factors are comprised of age, sex, and condition categories (CCs) for prior 12-month and current claims. These risk factors had been systematically chosen as predictors of any readmission for the same patient cohort as the current measure; the outcome of this measure is dominated by the number of days of a readmission, so we judged it unlikely that repeating the original analysis would produce different results. We confirmed that there were no additional risk factors to consider by comparing the model estimated using the a priori set of risk factors to a model which included all additional CCs.

For risk adjustment, we used a hierarchical generalized linear model (HGLM). This model consists two parts, a logit model and a truncated Poisson model. The two-part logit/Poisson model (often called a "hurdle" model) assumes that the outcome results from two related processes: an initial dichotomous event – that a patient has at least one acute care event – which is modeled as the logit of the probability of the event, and for patients with an event (those which clear the "hurdle"), the number of days, which is modeled as a Poisson process. The outcome, number of days, is a half-integer count variable (because ED visits count as 0.5 days). Observation care is counted according to the hours spent in observation care, rounded up to the nearest half day. For each patient, an exposure variable is defined as the number of survival days post discharge, up to 30. For the hurdle model, exposure time as an offset is included for each part of the model.

There are two random effects for each hospital, one for the logit model and one for the truncated Poisson model, as well as a covariance between the two random effects. The random effects allow us to account for within-hospital correlation of the observed outcome and accommodates the assumption that underlying differences in quality across hospitals lead to systematic differences in outcomes.

Socioeconomic Status Factors and Race

We selected variables representing 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 days in acute care in the 30-days after discharge.

Our conceptualization of the pathways by which patient SES or race affects days in acute care in the 30-days is informed by the literature on the association of SES and race with pneumonia readmissions, since the majority of the EDAC outcome is composed of readmission days, and since there is a much more robust literature about readmission than about observation care and ED visits.

Literature Review of Socioeconomic Status and Race Variables and Pneumonia Excess Days in Acute Care

To examine the relationship between SES and race variables and hospital 30-day, all-cause EDAC following pneumonia hospitalization, a literature search was performed with the following exclusion criteria: international studies, articles published more than 10 years ago, articles using Veterans Affairs databases as the primary data source, and articles not explicitly focused on SES or race and pneumonia readmission. Seventeen studies were reviewed by title and abstract, and eleven studies were excluded from full-text review. Among studies reviewed, there was evidence that SES and race increased the risk of pneumonia readmission (Lindenauer et al., 2013; Mather et al., 2014), with a noted risk associated with race in particular (Joynt et al., 2011; McHugh et al., 2010). However, other studies including a systematic review showed that results have been inconclusive (Calvillo-King et al., 2013; Vidic et al., 2015).

Causal Pathways for Socioeconomic Status 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 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 in 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 sufficient 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

We assessed the relationship between the dual-eligible status and race with the outcome and examined the incremental effect of each 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

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 2014; 7:391-7.

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. J Gen Intern Med. 2013 Feb; 28(2):269-82. doi: 10.1007/s11606-012-2235-x. Epub 2012 Oct 6.

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.

Gilman M, Adams EK, Hockenberry JM, et al. 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 KE, Orav EJ, Jha AK. Thirty-day readmission rates for Medicare beneficiaries by race and site of care. JAMA. 2011 Feb 16; 305(7):675-81. doi: 10.1001/jama.2011.123.

Krumholz H, Normand S, Galusha D, et al. Risk-Adjustment Methodology for Hospital Monitoring/Surveillance and Public Reporting Supplement #1: 30-Day Mortality Model for Pneumonia. 2006. Available at:

https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1163010421830. Accessed November 19, 2015.

Lindenauer PK, Lagu T, Rothberg MB, et al. Income inequality and 30 day outcomes after acute myocardial infarction, heart failure, and pneumonia: retrospective cohort study. BMJ. 2013 Feb 14; 346:f521. doi: 10.1136/bmj.f521.

Mather JF, Fortunato GJ, Ash JL, et al. Prediction of pneumonia 30-day readmissions: a single-center attempt to increase model performance. Respir Care. 2014 Feb; 59(2):199-208. doi: 10.4187/respcare.02563. Epub 2013 Aug 13.

McHugh MD, Carthon JM, Kang XL. Medicare readmissions policies and racial and ethnic health disparities: a cautionary tale. Policy Polit Nurs Pract. 2010 Nov; 11(4):309-16. doi: 10.1177/1527154411398490.

Normand S-LT, Shahian DM. Statistical and Clinical Aspects of Hospital Outcomes Profiling. 2007/05 2007:206-226.

Reames BN, Birkmeyer NJ, Dimick JB, et al. Socioeconomic disparities in mortality after cancer surgery: failure to rescue. JAMA surgery 2014; 149:475-81.

Skinner J, Chandra A, Staiger D, et al. 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.

Vidic A, Chibnall JT, Hauptman PJ. Heart failure is a major contributor to hospital readmission penalties. J Card Fail. 2015 Feb; 21(2):134-7. doi: 10.1016/j.cardfail.2014.12.002. Epub 2014 Dec 9.

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

Below is a table showing the final variables in the model with associated parameter estimates.

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

Risk variable	Part 1: Logit model		Part 2: Poisson model	
	Estimate	CI	Estimate	CI
Age minus 65 (years above 65, continuous)	0.001	(0.000, 0.002)	0.004	(-0.004, -0.003)
Male	0.090	(0.078, 0.102)	0.045	(0.039, 0.050)
History of Coronary Artery Bypass Graft (CABG) surgery (ICD-9 codes V45.81, 36.10-	-0.085	(-0.110, -0.059)	-0.028	(-0.038, -0.019)

Risk variable	Part 1: Logit model		Part 2: Poisson model	
	Estimate	CI	Estimate	CI
36.16)				
History of infection (CC 1, 3-6	0.039	(0.026, 0.053)	0.024	(0.019, 0.029)
Septicemia/sepsis (CC 2)	0.007	(-0.015, 0.029)	0.051	(0.044, 0.059)
Metastatic cancer or acute leukemia (CC 7)	0.264	(0.228, 0.297)	0.031	(0.018, 0.044)
Lung, upper digestive tract, and other severe cancers (CC 8)	0.181	(0.153, 0.209)	0.015	(0.002, 0.026)
Other major cancers (CC 9-10)	0.015	(-0.003, 0.031)	0.026	(0.020, 0.033)
Diabetes mellitus (DM) or DM complications (CC 15-19, 119-120)	0.074	(0.060, 0.089)	0.027	(0.023, 0.033)
Protein-calorie malnutrition (CC 21)	0.177	(0.159, 0.196)	0.122	(0.115, 0.130)
Disorders of fluid/electrolyte/acid-base (CC 22-23)	0.145	(0.129, 0.161)	0.022	(0.015, 0.028)
Other gastrointestinal disorders (CC 36)	0.079	(0.064, 0.093)	-0.016	(-0.023, -0.010)
Severe hematological disorders (CC 44)	0.165	(0.133, 0.196)	0.063	(0.053, 0.074)
Iron deficiency or other unspecified anemias and blood disease (CC 47)	0.108	(0.095, 0.012	0.064	(0.058, 0.070)
Dementia or other specified brain disorders (CC 49-50)	0.083	(0.070, 0.097)	-0.022	(-0.029, -0.016)
Drug/alcohol abuse/dependence/psychosis (CC 51-53)	0.112	(0.095, 0.131)	-0.045	(-0.052, -0.035)
Major psychiatric disorders (CC 54-56)	0.063	(0.047, 0.080)	-0.005	(-0.012, 0.003)
Other psychiatric disorders (CC 60)	0.120	(0.104, 0.136)	-0.028	(0.020, 0.037)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)	0.087	(0.064, 0.109)	0.028	(0.020, 0.037)
Respiratory dependence/tracheostomy (CC 77)	0.141	(0.090, 0.199)	0.003	(-0.013, 0.019)
Cardio-respiratory failure or	0.125	(0.104, 0.144)	0.061	(0.054, 0.067)

Risk variable	Part 1: Logit model		Part 2: Poisson model	
	Estimate	CI	Estimate	CI
shock (CC 78-79)				
Congestive heart failure (CC 80)	0.130	(0.112, 0.146)	0.043	(0.037, 0.049)
Acute coronary syndrome (CC 81-82)	0.093	(0.070, 0.118)	0.000	(-0.009, 0.008)
Coronary atherosclerosis or angina (CC 83-84)	0.064	(0.048, 0.078)	-0.025	(-0.031, -0.019)
Valvular or rheumatic heart disease (CC 86)	0.057	(0.041, 0.071)	0.019	(0.013, 0.025)
Specified arrhythmias and other heart rhythm disorders (CC 92-93)	0.086	(0.072, 0.100)	0.005	(-0.002, 0.012)
Stroke (CC 95-96)	0.082	(0.060, 0.109)	0.015	(0.007, 0.022)
Vascular or circulatory disease (CC 104-106)	0.043	(0.028, 0.057)	0.001	(-0.005, 0.007)
Chronic obstructive pulmonary disease (COPD) (CC 108)	0.102	(0.086, 0.116)	0.044	(0.039, 0.051)
Fibrosis of lung or other chronic lung disorders (CC 109)	0.056	(0.039, 0.073)	0.038	(0.030, 0.045)
Asthma (CC 110)	-0.004	(-0.027, 0.023)	-0.041	(-0.050, -0.033)
Pneumonia (CC 111-113)	0.032	(0.017, 0.047)	0.026	(0.019, 0.032)
Pleural effusion/pneumothorax (CC 114)	0.069	(0.050, 0.087)	0.068	(0.062, 0.074)
Other lung disorders (CC 115)	0.045	(0.030, 0.058)	-0.003	(-0.008, 0.004)
End-stage renal disease or dialysis (CC 129-130)	0.301	(0.266, 0.337)	-0.080	(-0.091, -0.067)
Renal failure (CC 131)	0.129	(0.112, 0.147)	0.058	(0.051, 0.064)
Urinary tract infection (CC 135)	0.089	(0.072, 0.104)	0.005	(-0.001, 0.011)
Other urinary tract disorders (CC 136)	0.046	(0.033, 0.062)	-0.005	(-0.012, 0.001)
Decubitus ulcer or chronic skin ulcer (CC 148-149)	0.085	(0.064, 0.106)	0.070	(0.063, 0.076)
Vertebral fractures (CC 157)	0.093	(0.064, 0.125)	0.056	(0.047, 0.065)
Other injuries (CC 162)	0.090	(0.075, 0.106)	-0.039	(-0.044, -0.033)
Respirator dependence/Tracheostomy (CC 77)	0.141	(0.090, 0.199)	0.003	(-0.013, 0.019)

References

Lindenauer PK RJ, Strait KM, et al. Reevaluation and Re-Specification Report of the Hospital-Level 30-Day Risk-Standardized Measures Following Hospitalization for Pneumonia-Pneumonia Readmission–Version 9.2, Pneumonia Readmission – Version 8.2. 2015; https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Measure-Methodology.html. Accessed December 3, 2015.

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*. 113 (3): 456-462.

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

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 dual-eligible and African-American patients in the pneumonia cohort varies across hospitals (number of hospitals=4,655). The median percentage of dual-eligible patients is 16.7% (interquartile range [IQR] 10.0%-25.8%). The median percentage of black patients is 1.4% (IQR 0%-8.05%).

Empirical Association with the Outcome (Univariate)

The mean patient-level observed days in acute care is higher for dual-eligible patients, 145.57 days in acute care per 100 discharges, compared with 119.55 days in acute care per 100 discharges for all other patients. The mean observed days in acute care for African-American patients was also higher at 176.11 days per 100 discharges compared with 119.91 days per 100 discharges for patients of all other races.

Incremental Effect of Socioeconomic Status Variables and Race in a Multivariable Model

We then examined the strength and significance of the dual-eligible status and race variables in the context of a multivariable model. When we include either of these variables in a multivariate model that includes all of the claims-based clinical variables, the effect size of the variable is small. We also find that the c-statistics for the logit part of the model and the deviance R^2 values for the Poisson part of the model are similar with and without the addition of either of these variables into the model. The c-statistics for the logit model with and without the dual-eligibility indicator in the model are 0.616. The c-statistic for the logit model without the race indicator is 0.616 and with the race indicator is 0.617. The deviance R^2 values for the Poisson model with and without dual-eligibility indicator are 0.034. The deviance R^2 values for the Poisson model with and without the race indicator are 0.034. Furthermore, we find that the addition of any of these variables into the model has little to no effect on hospital performance. We examined the change in hospitals' EDAC with the addition of either of these variables. The median absolute

change in hospitals' EDAC when adding a dual-eligibility indicator is 0.40 EDAC per 100 discharges (interquartile range [IQR] 0.19-0.69; minimum 0.00-maximum 8.50), with a Spearman correlation coefficient between EDAC for each hospital with and without dual eligibility added of 0.9997. The median absolute change in hospitals' EDAC when adding a race indicator is 0.56 EDAC per 100 discharges (IQR 0.27-0.98; minimum 0.00-maximum 11.69), with a Spearman correlation coefficient between EDAC for each hospital with and without race added of 0.9997.

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

Both the hospital-level and patient-level dual-eligible effects were significant in the logistic part of the pneumonia EDAC model, but only the hospital-level effect was significant in the Poisson part of the model. This indicates that a) both the patient- and hospital-level dual eligible effects are associated with an increased risk of acute care but b) only the hospital-level effect is associated with the expected duration of that care.

Both the patient-level and hospital-level race effects were significant in the logistic and Poisson parts of the pneumonia EDAC model. This indicates that a) both the hospital- and patient-level effects are associated with an increased risk of acute care and with the duration of that care following discharge from a pneumonia admission.

Because both the hospital- and patient-level effects contribute to the increased risk, if the dual-eligible or race variables were used in the model to adjust for patient-level differences, then some of the differences in both risk of acute care and expected duration of care between hospitals would also be adjusted for, potentially obscuring a signal of hospital quality.

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

Pneumonia EDAC Decomposition Analysis

Parameter	Logistic model estimate (standard error)	Logistic model p-value	Poisson model estimate (standard error)	Poisson model p-value
Dual Eligible – Patient-Level	0.0719 (0.0091)	<.0001	-0.0056 (0.0035)	0.1119
Dual Eligible – Hospital-Level	0.2428 (0.0395)	<.0001	0.2191 (0.0437)	<.0001
African American – Patient-Level	0.1809 (0.0137)	<.0001	0.0163 (0.0049)	0.0010
African American – Hospital-Level	0.1421 (0.0349)	<.0001	0.5912 (0.0386)	<.0001

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*)

Dataset

This model selection process was performed using one half (the development sample) of the random three-year split sample.

Approach to Determining Model Specifications

Because the outcome, number of days of acute care, is novel not only for quality measurement, but also in the literature as a measure of utilization, we considered a range of model specifications. We performed a number of analyses to determine the best model specification for the number of days in acute care. This is a pseudo-count variable (similar to a count variable, but taking half-integer values for half days of acute care), and we therefore considered models that were generalized count models. All model development was performed using the development sample.

Inspection of the distribution of the outcome determined that the number of event days was highly skewed, with a large number of zeroes. Thus, we considered models appropriate for skewed data, including approaches that modeled the zero-day outcomes and non-zero day outcomes separately. We only considered approaches that allowed us to incorporate exposure time to account for differential risk.

First, using only patients with non-zero days, we estimated a generalized linear model (GLM) using a Poisson specification, and applied a Park test (Manning and Mullahy, 2001); the Park test indicated that Poisson was the best fit for our outcome. The Poisson model is commonly used for modeling count data and can be generalized to dependent variables that take non-integer values, such as ours.

We then considered three different model specifications for the full set of outcomes (zero and non-zero days): Poisson, zero-inflated Poisson (ZIP), and two-part logit/Poisson ("hurdle" model). For each model, we included an offset for the number of days the patient survived discharge, up to 30 (i.e., the exposure time). For the hurdle model, we included exposure time as an offset for each part because the Poisson part included only observations with non-zero days, it was technically a 'truncated' Poisson model.

For each of the three specifications listed above, we estimated (non-hierarchical) generalized linear models with days in acute care as the outcome. We compared the three different model specifications for the outcome using the following criteria: Akaike information criterion (AIC), Baysian information criterion (BIC), and Log-likelihood.

Criterion	Poisson	Zero-inflated Poisson	Two-part logit/Poisson
Akaike information criterion (AIC)	2,420,000	1,450,000	1,460,000
Bayesian	2,420,000	1,450,000	1,460,000

Criterion	Poisson	Zero-inflated Poisson	Two-part logit/Poisson
information criterion (BIC)			
Log-likelihood	-1,210,000	-725,000	-730,000

We selected the best model based on these statistics and judgment regarding the technical challenges of extending each to a random effects model for the measure. The AIC is a measure of the relative quality of statistical models for a given set of data. The best performing model was the two-part logit/Poisson model. This model also made the most sense conceptually, with the likelihood of returning for acute care being modelled separately from the number of days of acute care received.

Assessing Model Discrimination and Calibration

Discrimination: We computed two different statistics – one for the logit part of the model and one for the Poisson part – using the development sample. For the logit model of zero versus non-zero days, which includes all patients in the cohort, we calculated the c-statistic. For the Poisson model of non-zero days, which includes only patients with some acute care, we calculated the deviance R^2. The deviance R^2 is computed from the difference in the log-likelihoods between the final model and an empty model (no covariates) attributed to each observation, averaged over all observations (Cameron, Windmeijer, 1996).

Calibration Statistics

In a generalization of the calibration statistics for logistic models, we calculated the linear prediction Z = XB and W = XC using the coefficients B and C from the development sample and data X from the validation sample. We then estimated a model using the same functional form but only two independent variables, Z for the truncated Poisson part and W for the logit part. The intercepts and coefficients of Z and W in these second models are reported as (γ_0, γ_1), the calibration statistics for each part of the model. The closer they are to (0, 1), the better the model calibration (Harrell, 2013).

Calibration Plot

To further assess model calibration we constructed calibration plots with mean predicted and mean observed days in acute care plotted against decile of predicted utilization rate (predicted days/exposure days).

References

Cameron AC and Windmeijer FAG. R-Squared Measures for Count Data Regression Models with Applications to Health-Care Utilization. Journal of Business & Economic Statistics, Vol. 14, No. 2 (Apr., 1996), pp. 209-220.

Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis.* Springer New York; 2013.

Manning WG, Mullahy J. Estimating log models: to transform or not to transform? *Journal of health economics.* 2001;20(4):461-494.

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*)**:**

Dataset:
The model discrimination statistics were calculated using the development sample:

Discrimination Statistics:
C-statistic for logit part of model: 0.616
Deviance R^2 for truncated Poisson part of model: 0.034 (3.4%)

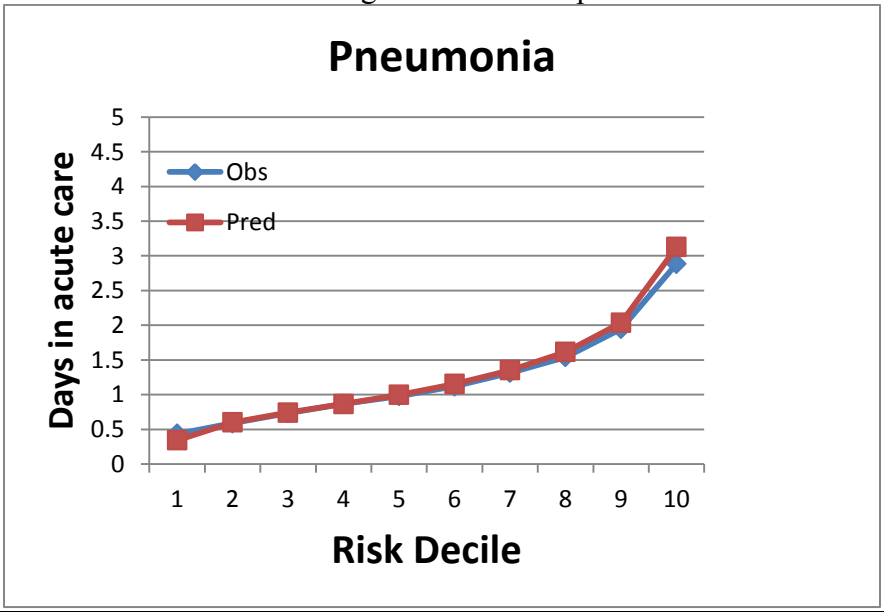
2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Dataset:
The model discrimination statistics were calculated using both the development and validation samples; see section 1.7.

Calibration Statistics* (y0, y1)*:
Logit part of model: (-0.05, 0.99)
Poisson part of model: (-0.05, 0.97)

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

Calibration Plot:
The plot below shows that the model underestimates risk for the lowest risk decile patients and slightly overestimates risk for the highest risk decile patients.



2b4.9. Results of Risk Stratification Analysis:

N/A. This measure is not risk stratified.

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 for the logit part of the model was 0.62; the deviance R^2 for the Poisson part of 0.034 is consistent with deviance R^2 for other count data models, indicating good model calibration.

Calibration Statistics
Over-fitting (Calibration γ0, γ1)
If the γ_0 in the validation sample is substantially far from zero and the γ_1 is substantially far from one, there is potential evidence of over-fitting. The calibration value of close to 0 at one end and close to 1 to the other end indicates good calibration of the model.

Calibration Plot
The calibration plot shows very good agreement between the mean of predicted days and the mean of observed days within same risk decile.

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*)

N/A

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)*

To categorize hospital performance, we estimated each hospital's EDAC and the corresponding 95% credible interval (CI) described in the attached Appendix (Section 2.7.2). We assigned hospitals to a performance category by comparing each hospital's EDAC interval estimate to zero. Comparative performance for hospitals with 25 or more eligible cases was classified as follows:

- "Lower than expected" if the entire 95% CI surrounding the hospital's days is below zero.
- "No different than expected" if the 95% CI surrounding the hospital's days includes zero.
- "Higher than expected" if the entire 95% CI surrounding the hospital's days is above zero.

If a hospital has fewer than 25 eligible cases for a measure, we assigned the hospital to a separate category: "The number of cases is too small (fewer than 25) to reliably assess the hospital's EDAC."

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)

Of 4,674 hospitals in the study cohort (data from July 1, 2010 through June 30, 2012), 619 had EDACs "lower than expected," 2,542 were "no different than expected," and 1,007 had EDACs "higher than expected." 506 were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing. The mean EDAC per 100 discharges for hospitals in the top decile of performance is -29.8, compared to 230.0 for hospitals in the bottom decile.

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 hospital-level EDAC suggests there are meaningful differences in the quality of care received across hospitals for the pneumonia EDAC measure.

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 criterion is directed 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).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of 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 demonstrating comparability of 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.**

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.**
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)
Public Reporting Not in use	

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

N/A. The measure is not yet in use.

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?*)

This measure is not currently publicly reported or used in an accountability application because it only recently completed development and is being submitted to NQF for initial endorsement.

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 may ultimately be used in one or more CMS programs, such as the:
-Hospital Inpatient Quality Reporting Program.

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**

N/A

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.**

Since this measure is not yet in use, there are no performance results to assess improvement.

We expect there will be improvement in measure scores over time since publicly reported measure scores can reduce adverse patient outcomes associated with days spent in acute care for pneumonia by capturing and making acute care utilization following

the index hospitalization more visible to providers and patients.

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 or model testing. 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)**

0229 : Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following heart failure (HF) hospitalization for patients 18 and older

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 all-cause risk-standardized readmission rate (RSRR) following elective primary total hip arthroplasty (THA) and total knee arthroplasty (TKA)

1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

1891 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization

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?

No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**

We developed the measure in the Medicare Fee-for-Service (FFS) population and completely harmonized the cohort definition and risk-adjustment strategy with those of the existing CMS 30-day pneumonia readmission measure. However, while the existing

measure counts readmissions as a dichotomous outcome, the proposed measure counts the number of days for all readmissions during the follow-up period, as well as the number of days of observation stays and ED visits. This difference in the outcome measure imposes differences on the statistical modeling and reporting format. There are no differences in data collection burden.

5b. Competing Measures

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

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
N/A

Appendix

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

Attachment **Attachment:** Pneumonia_Excess_Days_in_Acute_Care_Appendix_to_NQF_Application_01-29-16_v1.0.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services (CMS)
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 (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.
Yale New Haven Health Services Corporation/Center for Outcomes Research (YNHHSC/CORE) Measure Reevaluation Team Members
1. Faseeha K. Altaf, MPH- Lead Project Coordinator. Provided experience relevant to performance measurement.
2. Susannah Bernheim, MD, MHS- Project Director. Provided experience relevant to clinical content and performance measurement.
3. Nihar Desai, MD, MPH- Clinical Consultant. Provided experience relevant to clinical content and performance measurement.
4. Jacqueline Grady, MS- Supporting Analyst. Provided experience relevant to performance measurement.
5. Jeph Herrin, PhD- Statistician. Provided experience relevant to performance measurement.
6. Yongfei Wang, PhD- Supporting Analyst. Provided experience relevant to performance measurement.
7. Leora Horwitz, MD, MHS- Project Lead. Provided experience relevant to clinical content and performance measurement.
8. Zhenqiu Lin, PhD- Director of Analytics. Provided experience relevant to performance measurement.
9. Shuling Liu, PhD- Statistical consultant. Provided experience relevant to performance measurement.
10. Steven Susaña-Castillo, BA- Research Assistant. Provided support relevant to performance measurement.
11. Arjun Venkatesh, MD, MBA- Clinical Consultant. Provided experience relevant to clinical content and performance measurement.
12. Changqin Wang, MD, MS- Lead Analyst. Provided experience relevant to performance measurement.
13. Yongfei Wang- Supporting Analyst. Provided experience relevant to performance measurement.
14. Sharon-Lise Normand, Ph.D.* - Statistical Consultant. Provided statistical expertise for the project.

*Harvard Medical School

Technical Expert Panel (TEP) Members
1. Anonymous Patient- Patient Representative. Provided patient perspective.
2. Kevin E. Driesen, PhD, MPH, MA- Assistant Professor, Mel and Enid Zuckerman College of Public Health; Director, Arizona Rural Hospital Flexibility Program. Provided experience relevant to performance measurement.
3. David Engler, PhD- Senior Vice President for Leadership and Innovation, America's Essential Hospitals. Provided experience relevant to clinical content, performance measurement, and coding and informatics.
4. Timothy Farrell, MD- Assistant Professor of Medicine, Adjunct Professor of Family Medicine, Physician Investigator; University of Utah School of Medicine. Provided experience relevant to clinical content and performance measurement.
5. Karen Farris, PhD- Charles R. Walgreen III Professor of Pharmacy Administration, Director of the Social and Administrative Pharmacy Graduate Program; University of Michigan College of Pharmacy. Provided experience relevant to performance measurement.
6. Maura C. Feldman, MSW- Director for Hospital Performance Measurement and Improvement, Blue Cross Blue Shield of Massachusetts. Provided consumer perspective.
7. Jay A. Gold, MD, JD, MPH- Senior Vice President and Chief Medical Officer, MetaStar. Provided experience relevant to clinical content and performance measurement.
8. Sally Hinkle, DNP, MPA, RN- Director of Performance Improvement and Clinical Value, Temple University Hospital. Provided experience relevant to performance measurement.
9. Amy Jo Haavisto Kind, MD, PhD - Assistant Professor of Geriatrics, University of Wisconsin School of Medicine and Public Health; Attending Physician, William S. Middleton VA. Provided experience relevant to clinical content and performance measurement.
10. Marjorie King, MD, FACC, MAACVPR- Director of Cardiac Services, Helen Hayes Hospital. Provided experience relevant to clinical content and performance measurement.
11. Eugene Kroch, PhD- Vice President and Chief Scientist, Premier. Provided experience relevant to performance measurement.
12. Keith D. Lind, JD, MS, BSN- Senior Policy Advisor, American Association of Retired Persons (AARP) Public Policy Institute. Provided consumer perspective.
13. Grace McConnell, PhD- Patient Representative. Provided patient perspective.
14. Michael A. Ross, MD, FACEP- Medical Director, Professor of Emergency Medicine; Emory University School of Medicine. Provided experience relevant to clinical content and performance measurement.
15. Mark Louis Sanz, MD- Interventional Cardiologist, International Heart Institute of Montana. Provided experience relevant to clinical content and performance measurement.
16. Paul Takahashi, MD- Associate Professor of Medicine, Mayo Clinic College of Medicine. Provided experience relevant to performance measurement.

Methodology Work Group Members
1. Arlene Ash, PhD- Professor and Division Chief, University of Massachusetts Medical School. Provided experience relevant to performance measurement.
2. Jeremiah Brown, PhD, MS- Assistant Professor of Health Policy and Clinical Practice, The Dartmouth Institute for Health Policy and Clinical Practice. Provided experience relevant to performance measurement.
4. Grant Ritter, PhD, MS, MA- Senior Scientist, Schneider Institute for Health Policy & Heller Graduate School. Provided experience relevant to performance measurement.
5. Patrick Romano, MD, MPH- Professor of Medicine and Pediatrics, University of California Davis School of Medicine. Provided experience relevant to performance measurement.

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:**
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

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 #: 2827
De.2. **Measure Title:** PointRight® Pro Long Stay(TM) Hospitalization Measure
Co.1.1. **Measure Steward:** American Health Care Association
De.3. **Brief Description of Measure:** The PointRight Pro Long Stay Hospitalization Measure is an MDS-based, risk-adjusted measure of the rate of hospitalization of long-stay patients (aka "residents") of skilled nursing facilities (SNFs) averaged across the year, weighted by the number of stays in each quarter.
1b.1. **Developer Rationale:** In November 2013 the HHS Office of the Inspector General published a document entitled "Medicare Nursing Home Hospitalization Rates Merit Additional Monitoring" (HHS Document OEI-06-11-00040). The OIG report noted that one-quarter of Medicare nursing home residents had hospitalizations (i.e., direct discharges to acute care hospitals of Medicare residents, whether post-acute or long stay), and that these hospitalizations cost $14.3 billion – and this is for Medicare Fee for Service only. The rates of hospitalization varied significantly between states and between SNFs with different five-star ratings, suggesting that rates could be improved substantially if facilities rendered higher-quality care. The report details reasons for hospitalization and associates hospitalization costs with these reasons. For example, hospitalizations for pneumonia cost Medicare $844 million in one year, those for urinary tract infections without sepsis cost $422 million, and those related to aspiration of food or vomitus cost $618 million. These three conditions alone are obvious opportunities for quality improvement: Pneumococcal pneumonia can be prevented by immunization; catheter-associated UTIs can be prevented by high quality catheter care, avoidance of unnecessary indwelling catheters, and prophylactic antibiotics where appropriate; aspiration rates can be reduced by dietary modifications, supervised eating, and therapy for addressable swallowing problems. Even when infections develop many can be safely and effectively treated in the facility if the diagnosis is timely – reducing hospitalization rates both for the specific infection and for sepsis. Review of the OIG report suggests that reducing hospitalization costs by over $1 billion per year – for FFS Medicare beneficiaries alone – is a modest and attainable target. A 2010 report, showed that one third of the dually eligible population in SNFs are hospitalized at least once and over a third of them can be avoidable (Walsh et al., 2010). The same study stated that in 2005, the Medicare program paid $3 billion for potentially avoidable hospitalizations, and Medicaid paid $463 million. Again, these numbers demonstrate the high cost associated with hospitalizations.

CMS through its contractor RTI has developed a 30-day hospitalization rate quality measure for SNFs based on Medicare claims, and PointRight has developed one based on the MDS; both are endorsed by the NQF. However, to date no corresponding measure has been developed for long-stay residents. According to the national MDS data from CMS, there were 437,356 long nursing home stays discharged to an acute hospital in the year ending 2015 Q1. This demonstrates the importance of needing a hospitalization measure for long-stay residents,

In addition to their costs, it is known that hospitalizations are risky and potentially traumatic events for frail elderly patients, frequently associated with a declines in independent function, delirium and/or cognitive decline that may not be reversible, worsening of nutritional status and physical conditioning, and a risk of falls with injury, new pressure ulcers, and hospital-acquired infections They have also been tied to other risks associated with transitions of care such as the increased risk of medication errors. This offers additional motivation for reducing hospitalization rates of SNF residents, further establishing the need for a comprehensive set of performance measures related to this problem, and thus for a measure focusing on long-stay residents and including all payers.

Other published studies confirm the observations and the conclusions reported by the OIG in 2013, e.g., ones from the Kaiser Foundation (Jacobson, 2010), the Commonwealth Fund (Schoen, 2013), MedPAC (MedPAC, 2012) and CMS (Walsh, 2010). Studies by Ouslander have shown that structural and process issues within SNFs have a high impact on the rate of hospitalizations (Ouslander, 2012; Ouslander, 2011), further supporting the hypothesis that hospitalization rates could be reduced by feasible changes in

facilities' operations

Citations can be found at 1c.4.

S.4. **Numerator Statement:** The numerator for the measure is the sum over four quarters of the counts of hospitalizations of the quarterly denominator populations, where hospitalizations comprise discharges directly from the SNF to an acute care hospital.

The count of hospitalizations excludes discharges from the SNF to LTACHs, IRFs, and psychiatric hospitals, and excludes admissions to acute care hospitals that directly follow a discharge from the SNF to a setting other than an acute care hospital.

However, if a patient is discharged from a SNF directly to an acute care hospital during a quarter at risk, the hospitalization will be counted in the numerator even if the patient was discharged to a setting other than an acute care hospital earlier in that quarter.

Hospitalizations are counted over at-risk intervals of 3 months at a time because this period is long enough to yield nonzero numerators even for SNFs with low rates of hospitalization, yet short enough so that almost all of the denominator population will be present in the facility for all, or almost all, of the period. The latter feature makes the calculation simpler than if the risk exposure was calculated by days or weeks.Four quarters of denominators and four quarters of numerators are summed to yield the values for the full measure period.

S.7. **Denominator Statement:** The quarterly denominator population consists exactly of those patients present in the SNF on the first day of the quarter (the "snapshot date") who meet the criterion for long stay on that date. The denominator for a quarter is the number of patients in the quarterly denominator population. The denominator for the measure is the sum of the quarterly denominators for the four quarters in the 12 month measure period.

The criterion for a patient's having a long stay is a cumulative length of stay in the facility of more than 100 days as of the snapshot date. The cumulative length of stay of a patient is the length of the current stay as of the snapshot date and plus the full lengths of stay of any previous stays that are linked to it. According to the criteria for linkage of stays used in the present measure, a stay in a SNF is linked to a subsequent stay in the SNF if the patient was discharged from the SNF to the community and was readmitted to the SNF within 10 days or fewer. All stays in a sequence of linked stays are included in the sum of days used to determine a patient's cumulative length of stay. In these criteria the term "community" comprises private residences and all organized settings that are primarily residential in character, including senior housing, independent living facilities, board and care homes, and assisted living facilities.

A patient can contribute multiple times to the denominator for a 12 month measure period. For example, a resident continuously present in the facility for a full year would contribute four to the denominator.

S.10. **Denominator Exclusions:** There are no exclusions from the denominator; all patients in the facility on the snapshot date who meet the long stay criterion on that date are included. However, the measure will not be reported for a SNF if the annual unknown outcome rate is greater than 10%.The definition of the annual unknown outcome rate is provided in S.11.

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Electronic Clinical Data
S.26. **Level of Analysis:** Facility

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

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

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

De.4. **IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** N/A

New Measure -- Preliminary Analysis

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.

Summary of evidence provided by the developer:

- As a rationale for measuring this health outcome, the developer suggests that skilled nursing facilities are able to influence rates of hospitalizations for long term care residents in an number of ways including structural interventions such as high staffing levels and nurse practitioner availability as well as process interventions such as early detection of signs and symptoms of impending infections (pneumonia, urinary tract infection, etc.) and chronic disease exacerbation (e.g. congestive heart failure, diabetes mellitus, etc.)

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

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 developer cited a 2010 study showing that 33% of SNFs hospitalization can be avoidable, and in 2005 (according to the same 2010 study), avoidable hospitalizations cost Medicare $3 billion and Medicaid $463 million. Additionally, the developer presented data obtained from the national MDS data from CMS, citing 437,356 long nursing home stays discharged to an acute hospital from the first quarter of 2015.
- The developer compared the distributions of the PointRight Pro Long Stay Hospitalization Measure over three consecutive one-year measure periods, for a sample of 1,639 facilities that consistently submitted data to PointRight over all three periods and had known outcome rates of 90% or greater.
  - From 7/1/2014 to 6/30/2015, this included a sample of 448,642 patients.
  - The mean long stay adjusted hospitalization rate was 14.3%

Disparities

- To help in the assessment of potential disparities, the developers performed a univariate analysis to determine the relationship of sociodemographic factors.

Sociodemographic Factors	Percentage of Patient Quarters	Person Correlation Coefficient: Sociodemographic Factor with First Hospitalization During Quarter of Risk	p-value
Age			
<65	17%	0.03	<.0001
65-69	6%	0.02	<.0001
70-74	9%	0.02	<.0001
75-79	12%	0.01	<.0001
80-84	17%	-0.01	<.0001
85-89	19%	-0.02	<.0001
>=90	20%	-0.03	<.0001
Race/Ethnicity			
Asian	2%	0.00	0.3346
Black or African American	16%	0.03	<.0001

Hispanic Latino	6%	0.01	<.0001
White	73%	-0.03	<.0001
Other	3%	0.00	0.0095
Medicaid Beneficiary			
Yes	82%	0.02	<.0001
Gender			
Male	32%	0.05	<.0001

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

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

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

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):

Specifications:

- This measure calculates the rate of hospitalization of residents of skilled nursing facilities (SNFs), averaged across the year, weighted by the number of stays in each quarter.
- The measure produces a for a facility's adjusted PointRight Pro Hospitalization Rate which is calculated as: [Observed rate of all hospitalizations]/[Expected rate of all hospitalizations]*[National average rate of all hospitalizations].
- The numerator for the measure is the sum over four quarters of the counts of hospitalizations of the quarterly denominator populations, where hospitalizations comprise discharges directly from the SNF to an acute care hospital.
- The quarterly denominator population consists exactly of those patients present in the SNF on the first day of the quarter (the "snapshot date") who meet the criterion for long stay on that date. The denominator for a quarter is the number of patients in the quarterly denominator population. The denominator for the measure is the sum of the quarterly denominators for the four quarters in the 12 month measure period.
- The data source for this measure is the SNF-Minimum Data Set (MDS) version 3.0.
- This measure is calculated over twelve months, comprising four consecutive calendar quarters. Each quarter has its own denominator population and its own numerator; the quarterly numerators and quarterly denominators are each summed to create the numerator and the denominator for the 12 month measure period. The measure is updated quarterly.
- The measure is risk-adjusted using a statistical risk model (see details below).

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

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

The developers performed three types of reliability testing:

1) Agreement of Model Independent Variables
    a. The developers compared the prevalence of the risk adjustment covariates between a testing sample of 2,096 SNFs and the national population.
2) Reliability of Rates over Time
    a. The developer analyzed change from quarter to quarter in the observed and adjusted long-stay hospitalization rates. The developer explained that their reasoning was that the underlying probability of a SNF's long-stay patients hospitalizing and the characteristics of its long-stay patient population were unlikely to change greatly in a three month period so that most of the change from quarter to quarter would be due to limitations on measure reliability.
3) Stability of Facility Level Adjusted Rate Bootstrapping
    a. The developer recalculated adjusted rates for the measure for CY 2014 using a random sample of stays. The developer then reviewed the distribution of differences between facilities' original adjusted rates and the rates calculated with the new sample. The developer interpreted a distribution of differences with a small variance and a mean of zero as acceptable measure stability or reliability.

Results of reliability testing

1) Agreement of Model Independent Variables
    a. 48% of the comparable risk adjustment model covariates were found to have prevalence within 5% of the prevalence found in the national sample. 66% were found to have prevalence within 10% of the prevalence found in the national sample.
    b. The developer interpreted these findings that the sample sufficiently represents the SNF population.
2) Reliability of Rates over Time
    a. Correlations from one quarter to the next ranged between .884 to .894 for the parametric statistic and .877 to .886 for the rank order statistic.
    b. The developers note that this suggests that the measure is adequately stable over short periods, but sufficiently variable to reflect clinically meaningful changes.
3) Stability of Facility Level Adjusted Rate Bootstrapping
    a. 65.6% of the PointRight sample had a difference in adjusted rates of less than 2% and only 2.1% of facilities had a difference greater than 5%. The mean difference was .0008%.

Guidance from the Reliability Algorithm

Question 1: Submitted specifications are precise, unambiguous, and complete.
Question 2: Empirical reliability testing was conducted using a adjusted rate bootstrapping.
Question 3: Empirical validity testing of patient-level data was conducted.
Question 4: Reliability testing was conducted with computed performance measure scores for each measure
Question 5: The adjusted rate bootstrapping method can be considered a method for assessing the proportion of variability due to real differences among measured entities.
Question 6: The bootstrapping procedure shows that a. 65.6% of the PointRight sample had a difference in adjusted rates of less than 2% and only 2.1% of facilities had a difference greater than 5%. The mean difference was .0008%.

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?*
- *Does the Committee agree with the developer's approach to assessing measure score reliability?*

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

2b. Validity

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**

- This measure calculates the sum over four quarters of the counts of hospitalizations of the quarterly denominator population.
- As a rationale for measuring this health outcome, the developers that there are a number of interventions such as staffing levels, nurse practitioner availability, early detection of infections, management of chronic disease exacerbations that facilities can undertake to reduce the incidence of hospitalizations.

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.

Validity testing method:

The developer performed two methods of validity testing:

- Agreement of Model Dependent Variables
  - The developer compared the identification of hospitalizations of Medicare Fee for Service beneficiaries between the MDS and Medicare FFS claims.
  - The developer used 2012 MDS data, claims data, and enrollment data because it was the most recent available.
- Performance Measure Score- Correlation with SNF Industry Measures of Quality
  - The developer tested the relationship of the measure with various components of the CSM Five-Star ratings for SNFs and its correlation with CMS's long-stay quality measure.
  - The developer hypothesized that that facilities with higher star ratings would have lower adjusted long-stay hospitalization rates, and specifically that the relationship would be stronger for the long-stay quality measures and RN staffing stars, as opposed to survey (i.e., compliance) stars or overall staffing stars.

Validity testing results:

- Agreement of Model Dependent Variables
  - The test was based on 241, 857 long stay discharges for patients enrolled in Medicare Part A. This sample covered 15,091 SNFs.
  - The comparison showed that that 86% of hospitalizations of Medicare FFS patients identified by the MDS are confirmed by Medicare FFS claims; in the other direction, 98% (208,891 of 213,772) of acute inpatient claims found near an MDS discharge have an MDS discharge code of acute hospital.
  - The developer interprets this finding that MDS discharge assessments appear to be overstating the rate of acute hospitalizations to a moderate degree but that the overall high level of agreement between MDS discharge coding and claims supports the validity of the measure.
- Performance Measure Score- Correlation with SNF Industry Measures of Quality

- The developer found that the PointRight Pro Long Stay Hospitalization Measure was correlated with other measures of quality.
- The Pro Long Stay Adjusted Hospitalization rates had a statistically significant positive relationship with Pro30 Adjusted Rehospitalization rates at $p < .01$. The correlation coefficient was .47.
- The developer notes that higher star ratings are associated with lower adjusted long-stay hospitalization rates, and the relationship is strongest for the Quality and RN Staffing stars.

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:

- There are no exclusions from the denominator; all patients in the facility on the snapshot date who meet the long stay criterion on that date are included. However, the measure will not be reported for a SNF if the denominator population over the measure period's 4 snapshot dates is less than 30.

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

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

- This measure employs four logistic regression models applied to four discrete subgroups of the denominator population.
  - Calculation of a patient's risk of any hospitalization (or equivalently, the risk of a first hospitalization) during a quarter at risk begins by assigning the patient to one of four subgroups of the denominator population based on the duration of the patient's current stay in the SNF as of the snapshot date.
  - For each group the risk of one or more discharges from the SNF directly to an acute care hospital during the quarter is estimated by a logistic regression.
- The developer notes that the selection of risk factors (independent variables) involved an iterative process.
  - A panel of clinicians with extensive SNF experience recommended potential risk adjusters. The developer's overall approach was to begin with reliable and rarely-missing patient-level SDS variables nominated by the clinical experts: Medicaid status (as a proxy for financial assets and income), black versus non-black, Hispanic/Latino versus non-Hispanic/Latino, and the interactions of Medicaid status and race. The significance of these variables in predicting hospitalization rates was tested in fixed-effects logistic regression models. We reasoned that patient-level effects that were significant in models that included facility-specific constant terms probably reflected otherwise-unmeasured differences in baseline health status. These, and a full set of sociodemographic and contextual factors were tested for univariate relationships with hospitalizations.
  - The variables with the strongest univariate correlations were then used to build multivariate models.
  - The multivariate models (logistic regressions for each stratum of LOS) were reviewed by a larger panel of clinicians and potential users of the measure.
  - Variables were rejected and replaced if their coefficients were opposite to their univariate correlation with the hospitalization, or if they were viewed as potentially under the control of the SNF – i.e., creating a risk of over-adjustment.

Conceptual analysis of the need for SDS adjustment

- The developer did not include a conceptual analysis of the need for SDS factors.

Empirical analysis of SDS factors:

- The developer tested SDS factors that were included on the MDS assessment that were consistently and reliably filled out.
- The factors available were age, gender, race/ethnicity, and Medicaid beneficiary status.
- The developer performed a univariate analysis of the above factors. The developer found that the strength of the effects supports the inclusion of the factors in the risk adjustment models.

Sociodemographic Factors	Percentage of Patient Quarters	Person Correlation Coefficient: Sociodemographic Factor with First Hospitalization During Quarter of Risk	p-value
Age			
<65	17%	0.03	<.0001
65-69	6%	0.02	<.0001
70-74	9%	0.02	<.0001
75-79	12%	0.01	<.0001
80-84	17%	-0.01	<.0001
85-89	19%	-0.02	<.0001
>=90	20%	-0.03	<.0001
Race/Ethnicity			
Asian	2%	0.00	0.3346
Black or African American	16%	0.03	<.0001
Hispanic Latino	6%	0.01	<.0001
White	73%	-0.03	<.0001
Other	3%	0.00	0.0095
Medicaid Beneficiary			
Yes	82%	0.02	<.0001
Gender			
Male	32%	0.05	<.0001

Risk model diagnostics

- To assess the overall performance of their risk-adjustment model, the developers compared their model coefficients to the mean coefficients from bootstrap analysis, expressed as actual values, standard deviation (S.D.) and percentage.
- The developer performed a Hosmer-Lemeshow test for the goodness of fit of the logistic regression models. The test assesses whether or not the observed event rates match expected event rates in subgroups of the model population.
- Risk-Model Discrimination Statistics:
    - Logistic Regression Model Long Stay Group 1 c-statistic = .64
    - Logistic Regression Model Long Stay Group 2, c-statistic = .63
    - Logistic Regression Model Long Stay Group 3, c-statistic = .62
    - Logistic Regression Model Long Stay Group 4, c-statistic = .63
    - Linear Regression Model Rate of all Hospitalizations, R-squared = .96
- Risk-Model Calibration Statistics:

- Hosmer-Lemeshow Statistic Long Stay Group 1 (current LOS <= 100 days but cumulative days in SNF >100 days)

Partition for the Hosmer and Lemeshow Test			
Group	Total Patients at Risk	One or More Hospitalizations	
		Observed	Expected
1	6113	641	625
2	6655	1036	1082
3	6748	1309	1284
4	6712	1424	1450
5	6700	1606	1632
6	6682	1735	1801
7	6636	1984	1946
8	6641	2176	2135
9	6655	2438	2358
10	7447	3072	3107

- Homer-Lemeshow Statistic Long Stay Group 2 (100 days < LOS <= 181 days)

Partition for the Hosmer and Lemeshow Test			
Group	Total Patients at Risk	One or More Hospitalizations	
		Observed	Expected
1	9860	711	691
2	9853	864	1040
3	9845	1071	1174
4	9874	1313	1312
5	9860	1540	1433
6	9864	1651	1582
7	9861	1855	1741
8	9858	1965	1947
9	9857	2311	2243
10	9841	2839	2958

- Homer-Lemeshow Statistic Long Stay Group 3 (181 days < LOS <= 364 days)

Partition for the Hosmer and Lemeshow Test			
Group	Total Patients at Risk	One or More Hospitalizations	
		Observed	Expected
1	14404	721	745
2	14441	1064	1119
3	14410	1180	1255
4	14420	1322	1387
5	14410	1547	1514
6	14400	1733	1667
7	14410	1906	1837
8	14409	2109	2061

9	14413	2415	2387
10	14387	3167	3192

- Hosmer-Lemeshow Statistic Long Stay Group 4 (LOS > 364 days)

Partition for the Hosmer and Lemeshow Test			
Group	Total Patients at Risk	One or More Hospitalizations	
		Observed	Expected
1	38837	1150	1151
2	38834	1615	1679
3	38834	1973	1974
4	38833	2165	2229
5	38829	2444	2479
6	38843	2794	2750
7	38833	3128	3070
8	38837	3577	3471
9	38834	4054	4046
10	38825	5451	5504

- Linear Regression of All Hospitalizations by Decile of Expected Rates

Group	Observed Rate of Total Hospitalizations	Expected Rate of Total Hospitalizations
1	2.2%	2.5%
2	6.1%	6.6%
3	8.0%	8.7%
4	9.9%	10.4%
5	11.4%	11.9%
6	13.1%	13.4%
7	14.8%	15.0%
8	16.8%	16.9%
9	19.3%	19.4%
10	26.9%	25.5%

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.*

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

- To determine the change in the risk adjusted rate that will be considered meaningful the developer observed the distribution of change in the testing sample.
- The developer started with risk adjusted rates covering the 12 month measurement period of October 1st, 2013 to September 30th, 2014 and observed the quarterly changes up until the 12 month measurement period of

January 1st, 2014 to December 31st, 2014.

- The developer bucked the sample into deciles of change in adjusted rates and calculated the average change for each bucket. The same analysis was performed on subsets of the sample, which was divided into 3 groups based on denominator size:
  - Large – Denominator > 400 patient quarters
  - Medium – 400 >= Denominator > 200 patient quarters
  - Small – 200 >= Denominator > 30 patient quarters
- The distribution of change in adjusted rates was similar across all four quarters where for each quarter the average change for deciles 2 through 8 was less than +/- 3%. Deciles 1 and 10 had average changes greater than +/-3.5%.
- The distribution of differences was larger for facilities with smaller denominators and this indicated that recommendations of clinically meaningful difference should be dependent upon facility size.
- The developer made the following recommendations to identify changes in adjusted rates that would move a facility several deciles in the sample's distribution.
  - Large Facilities - 4%
  - Medium Facilities - 3%
  - Small Facilities - 2%

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

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

Criterion 3. Feasibility

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.

The developer states:

- The required data elements are routinely generated and used during care delivery. They are collected and used by healthcare personnel during the provision of care.
- Implementing the measure will not result to additional charges for healthcare organizations. Measure scores are computed by large-scale data management software (SAS, SPSS, Stata, R) already typically in-place in such facilities.
- Measure reporting requires including the measure's trademark, indicating that measure specifications are copyrighted by PointRight.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*

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

Criterion 4: Usability and Use

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**
OR
Planned use in an accountability program? ☒ **Yes** ☐ **No**

- Planned use for public reporting – for CMS' evaluation of SNF's clinical performance; AHCA plans to publish this measure on its website for free use by AHCA members and other selected stakeholders
- Quality improvement with external benchmarking to multiple organizations
- Quality improvement with internal benchmarking to specific organization

Accountability program details N/A

Improvement results

- PointRight Pro long stay rates from 2013 to 2015 show increase from 13.2% in 2013, to 14% in 2014 and 14.3% in 2015.
- Short stay rates have consistently decreased over the same period. 17.5% in 2013 to 17.4% in 2014 and 17.3% in 2015.
- Increase in long stay rates reinforces the need for a long stay hospitalization measure that nursing homes can use in their quality improvement programs.

Potential harms

- The developer states that "no unintended consequences have been identified or are anticipated to occur as a result of this measure."

Feedback : N/A

***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**

Criterion 5: Related and Competing Measures

Related or competing measures

- No related or competing measures.

Harmonization

- N/A

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** N/A

Measure Title: **PointRight® Pro Long Stay™**

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: N/A

Date of Submission: 1/29/2016

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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: Long Stay Hospitalizations

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

☐ Process: Click here to name the process

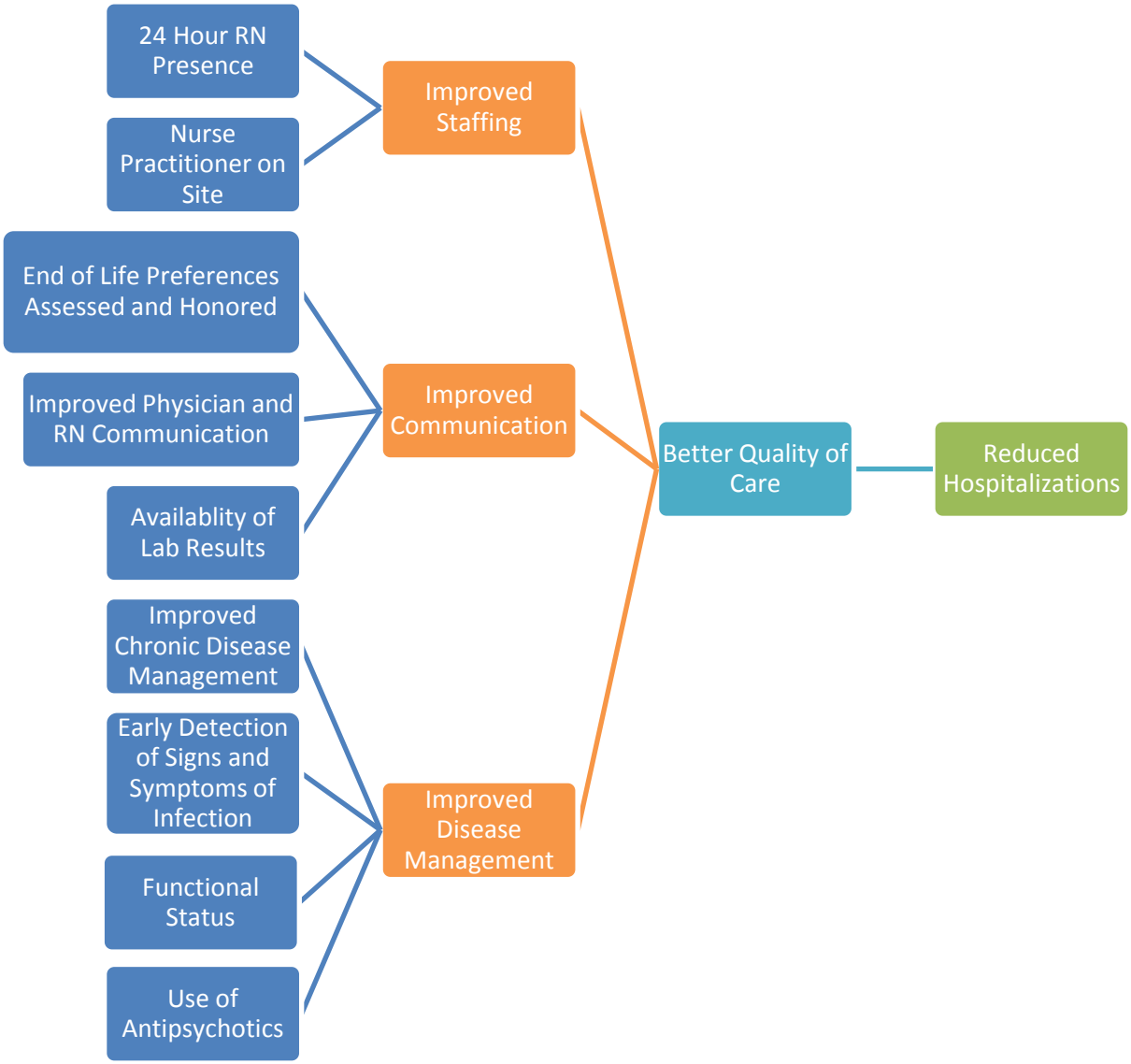
☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

Hospitalizations of any cause among individuals admitted to a skilled nursing facility (SNF) are the result of numerous clinical and non-clinical situations (Ouslander, 2012). However, a combination of structure, process and interventions influence the likelihood of hospitalizations more than patient acuity and condition. Structural interventions such as high staffing levels and nurse practitioner availability and processes interventions such as early detection of signs and symptoms of impending infections (pneumonia, UTI, etc.) and chronic disease exacerbation (e.g. CHF, DM, etc.) can all work to decrease the incidence of hospitalizations (Ouslander, 2012; Young et al., 2011). The diagram below provides an overview of the structures and processes that can ultimately influence hospitalizations in long term care residents.



1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Quality of life and quality of care are two areas that past research initiatives have utilized to measure quality. Quality of life focuses on issues surrounding the resident's autonomy while quality of care examines the technical aspects of health care that affect the resident's quality of health outcomes such as pressure ulcer prevalence (Spilsbury et al., 2011). Previous evidence supports the theory that quality measures are beneficial to determine the rate of hospitalization among long-stay residents. Facilities that do not have a high standard on their quality measures are more likely to have higher rates of hospitalization among long-stay residents. The evidence presented below contains past findings that describe the causal relationship between clinical outcomes and quality measures among long-stay nursing home residents and the ultimate influence on long-stay resident hospitalizations.

Improving Staffing

Staffing levels within skilled nursing facilities have the ability to affect residents' quality of care and quality of life. Measures such as staffing ratios and licensed nursing staff availability have previously been determined to have a causal relationship between quality and health outcomes for long-stay residents. Horn et al. (2005) evaluated staffing levels in relationship to residents' health outcomes. Their study provided evidence that

facilities where registered nurses (RN) provided 30 to 40 minutes of care per resident per day had positive health outcomes. Higher ratios among licensed practical nurses and certified nurses' aides also provided better health outcomes. Newly admitted residents in the study were less likely to remain in the study (71.2 versus 80.8 days, $P<0.001$), and to develop pressure ulcers. In addition, in centers where registered nurses provided 30 to 40 minutes of direct care showed a decreased in adverse outcomes while improvements in care processes increased. In addition, evidence showed a reduction in catheterization, pressure ulcers, and the development of UTI (Horn et al., 2005). Quality measures for long-stay residents showed improvement when residents received 4.1 hours of direct care per day and 1.35 hours of care from licensed staff per day (Collier & Harrington, 2008).

Indicators such as quality of care deficiencies, quality of life deficiencies, in-bed time, and resident satisfaction have also been examined (Spilsbury et al., 2011). The relationship among RNs and residents typically assume a linear relationship where higher staffing numbers provides better quality of care and fewer deficiencies (Spilsbury et al., 2011). Resident outcomes included fewer resident care deficiencies within the first year of admission and reduced mortality (Collier & Harrington, 2008). Outcome measures for long-stay residents were measured by a comparison of two MDS assessments (separated by 90 days) which focused on variables such as functional improvement and weight loss (Collier & Harrington, 2008).

Facilities that have an on-site physician report lower hospitalization rates compared to facilities without a physician on-site. Young et al. (2011) reported a decrease in hospitalizations where facilities employed on-site physician assistants, nurse practitioners, and a training program for nurses' aides. However, the majority of facilities do not have an on-site physicians and rely on nursing assessments to observe residents' health and function-related problems as a strategy for management of care (Young et al., 2011). It is essential to have licensed staff members that can perform a proper assessment on residents' conditions while determining if a hospital transfer is necessary. Facilities that cannot perform the proper medical assessment and communicate their findings to a physician have a higher risk of hospitalization rates (Young et al., 2011).

Improving Communication

Effective communication between physicians and nursing staff leads to a reduction in hospitalization among long-stay residents. Both physicians and nursing staff must be trained on effective communication to provide better information about a patient's condition regarding acute conditions and end-of-life care. Effective communication reduces the number of hospitalizations and encourages physicians to treat patients in the nursing home, thus avoiding unnecessary transfers (Young et al., 2011). At the same time, physicians must be provided resources to direct care within a nursing home such as the patients' medical history, a lab, and lab results within a four hour timeframe during non-business hours. In addition, nursing staff must be trained to provide accurate assessments of a resident's condition so that the physician may determine if a hospital transfer is necessary. Proper protocols must be put in place in order to provide the correct level of care to the resident and avoid hospitalizations (Young et al., 2011). The protocols for patient transfers must include resources for non-business hours (6 p.m.-6 a.m. and weekends) such as licensed staff on-site, in order to avoid improper hospitalization. Saliba et al. (2000) found that inappropriate transfers were more likely to occur during non-business hours because the facilities did not have the proper resources to treat the resident.

It is necessary to have conversations on advanced care planning with the resident and family members. These conversations are centered on noting the resident's preferences while they are cognitively and physically able to share their wishes. Research has demonstrated that advanced care planning improves end of life care, decreases life-sustaining treatment and prevents hospitalizations. At the same time, advanced care planning leads to an increase in the use of hospice and palliative care (Brinkman-Stoppelenburg et at., 2014). It has further been shown that having an advanced care directive can lower the rate of hospitalizations and death in a hospital

(Detering & Silveira, 2015). All staff must be aware of healthcare advanced directives when discussing patient transfers to avoid inappropriate transfers and respect end-of-life wishes (Saliba et al., 2000).

Improving Disease Management

INTERACT

The Interventions to Reduce Acute Care Transfers (INTERACT) is a set of evidence-based clinical practice tools and strategies initially developed as a demonstration program to reduce hospitalization rates. The program reduced avoidable hospitalization in rates among nursing homes during the six-month implementation period (Ouslander et al., 2011). Overall, the program saw fewer complications and less morbidity from hospitalizations and reductions in Medicare expenditures (Ouslander et al., 2011). In combination with the INTERACT tools, nursing homes can employ best practices to avoid or mitigate risk factors for hospitalization among long-stay residents with chronic conditions. Potentially avoidable hospitalizations among nursing home residents are considered to be hospital admissions based on acute exacerbation of a chronic condition where preventative care could have been provided (Spector et al., 2013). Chronic conditions can be effectively managed in nursing homes if preventative measures or best practices are put into place. For example, infection control, falls prevention, and proper hygiene for residents with open sores are measures that could be utilized to reduce unnecessary hospitalization (Spector et al., 2013).

Functional Status/ADLs

Long-stay residents are more likely to demonstrate functional and behavioral impairment throughout their length of stay. Functional status is a practical outcome measure for this population, specifically in the physical and self-care domain, as long-stay residents are likely to demonstrate functional limitations (Gillen et al., 1996). Change of functional status is most likely to occur within the three month period after admission. Long-stay residents are more likely to remain stable at the same functional level and therefore, less likely to be discharged from a facility. Gillen et al. (1996) found a positive relationship between higher levels of functional impairment and higher probabilities of hospitalization and death among the long-stay population. Over half of their sample experienced a functional status change and/or two or more transitions. Activities of daily living (ADLs) dependency level is another risk factor for hospitalization and post discharge mortality (Ponzetto et al. 2003). It is essential that nursing homes provide the appropriate level of care in order for residents to maintain the same functional status. Maintaining functional status will prevent the deteriorating of health and reduce hospitalization.

Antipsychotics

Antipsychotics have been utilized to treat behavioral and psychotic symptoms in dementia patients. However, recent initiatives have warned against the adverse effects of these drugs on the elderly. The use of antipsychotics among long-stay residents shows evidence of mixed reviews with caution for adverse effects. The typical approach to treating a health condition is a combination of pharmacological and nonpharmacological methods. There is evidence that side effects from antipsychotic use have the ability to reduce a resident's functional status and quality of life. For example, drugs with anticholinergic burden (ACB) have shown to increase cognitive and physical impairments which can lead to a rapid functional decline (Kolanowski et al., 2009).

Typically, adverse effects due to antipsychotic use may affect a resident's quality of life including depression, cognitive impairment and hospitalization. Older adults are more sensitive to adverse effects from antipsychotics, therefore, caution must be used (Frenchman, 2005). Long-stay residents who ingest high levels of ACB are more likely to be socially withdrawn from activities that require high social engagement. Sedation and confusion are common side effects associated with ACB (Kolanowski et al., 2009). Atypical antipsychotics have also been proven to have negative results on individuals with dementia. For instance, a study by Gareri et. al

(2010), found that the drugs risperidone and olanzapine have been shown in increase adverse cardiovascular events in the elderly.

Brinkman-Stoppelenburg, A., Rietjens, J.A., Van Der Heide, A. (2014). The effects of advance care planning on end-of-life care: A systematic review. Palliat Med. 28: 1000-1025.

Collier, E. & Harrington, C. (2008). Staffing characteristics, turnover rates, and quality of resident care in nursing facilities. Research in Gerontological Nursing. 1(3): 157-170.

Detering, K. & Silveira, M.J. (2015). Advanced Care Planning and Advanced Directives. UpToDate. http://www.uptodate.com/contents/advance-care-planning-and-advance-directives.

Frenchman, I.B. (2005). Atypical antipsychotics for nursing home patients. Drugs and Aging. 22(3): 257-264.

Gareri, P., Marigliano, N. M., De Fazio, S., Lacava, R., Castagna, A., Costantino, D. S., & De Sarro, G. (2010). Antipsychotics and dementia. BMC Geriatrics, 10(Suppl 1), A93.

Gillen, P., Spore, D., Mor, V., Freiberger, W. (1996). Functional and residential status transitions among nursing home residents. Journal of Gerontology: Medical Sciences. 51A(1): M29-M36.

Hamer, S., Haddad, P.M. (2007). Adverse effects of antipsychotics as outcome measures. British Journal of Psychiatry. 191(50): s64-s70.

Horn, S.D., Buerhaus, P., Bergstrom, N., Smout, R.J. (2005). RN staffing time and outcomes of long-stay nursing home residents. Journal of Nursing. 105(11): 58-70.

Intrator, O., Zinn, J., Mor, V. (2004). Nursing home characteristics and potentially preventable hospitalizations of long-stay residents. Journal of American Geriatrics Society. 52: 1730-1736.

Kahn, J.M., Werner, R.M., David, G., Have, T.R.T., Benson, N.M., Asch, D.A. (2013). Effectiveness of long-term acute care hospitalization in elderly patients with chronic critical illness. MedCare. 51(1): 4-10.

Kolanowski, A., Fick, D.M., Campbell, J., Litaker, M., Boustani, M. (2009). A preliminary study of anticholinergic burden and relationship to a quality of life indicator, engagement in activities, in nursing home residents with dementia. Journal of the American Medical Directors Association. 10(4): 252-257.

of Psychiatry. 191(59): 264-279.

Oslander, J.G., & Maslow, K. (2012). Geriatrics and the triple aim: Defining preventable hospitalizations in the long-term care population. J Am Geriatr Soc., 60(12): 2313-2318.

Ouslander, J.G., Lamb, G., Tappen, R., Herndon, L., Diaz, S., Roos, B.A., Grabowski, D.C., Bonner, A. (2011). Interventions to reduce hospitalizations from nursing homes: Evaluation of the INTERACT II collaborative quality improvement project. Journal of American Geriatrics Society 59: 745-753.

Ponzetto, M., Maero, B., Maina, P., Rosato, R., Ciccone, G., Merletti, F., Rubenstein, L.Z., Fabris, F. (2003). Risk factors for early and late morality in hospitalized older patients: The continuing importance of functional status. Journal of Gerontology: Medical Sciences. 58A(11): 1049-1054.

Saliba, D., Kington, R., Buchanan, J., Bell, R., Wang, M., Lee, M., Herbst, M., Lee, D., Sur, D., Rubenstein, L. (2000). Appropriateness of the decision to transfer nursing facility residents to the hospital. Journal of American Geriatrics Society. 48: 154-163.

Spector, W.D., Limcangco, R., Williams, C., Rhodes, W., Hurd, D. (2013). Potentially avoidable hospitalizations for elderly long-stay residents in nursing homes. Medical Care 51(8): 673-681.

Spilsbury, K., Hewitt, C., Stirk, L., Bowman, C. (2011). The relationship between nurse staffing and quality of care in nursing homes: A systematic review. International Journal of Nursing Studies. 48: 732-750.

Werner, R.M., Konetzka, R.T., Kim, M.M. (2013). Quality improvement under nursing home compare: The association between changes in process and outcome measures. MedCare. 51(7): 582-588.

Young, Y. Inandar, S., Dichter, B.S., Kilburn, H., Hannan, E.L. (2011). Clinical and nonclinical factors associated with potentially preventable hospitalizations among nursing home residents in New York state. Journal of American Medical Directors Association. 12: 364-371.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of 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
PointRight_Pro_Long_StayTM_Evidence_Final.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)*
In November 2013 the HHS Office of the Inspector General published a document entitled "Medicare Nursing Home Hospitalization Rates Merit Additional Monitoring" (HHS Document OEI-06-11-00040). The OIG report noted that one-quarter of Medicare nursing home residents had hospitalizations (i.e., direct discharges to acute care hospitals of Medicare residents, whether post-acute or long stay), and that these hospitalizations cost $14.3 billion – and this is for Medicare Fee for Service only. The rates of hospitalization varied significantly between states and between SNFs with different five-star ratings, suggesting that rates could be improved substantially if facilities rendered higher-quality care. The report details reasons for hospitalization and associates hospitalization costs with these reasons. For example, hospitalizations for pneumonia cost Medicare $844 million in one year, those for urinary tract infections without sepsis cost $422 million, and those related to aspiration of food or vomitus cost $618 million. These three conditions alone are obvious opportunities for quality improvement: Pneumococcal pneumonia can be prevented by immunization; catheter-associated UTIs can be prevented by high quality catheter care, avoidance of unnecessary indwelling catheters, and prophylactic antibiotics where appropriate; aspiration rates can be reduced by dietary modifications, supervised eating, and therapy for addressable swallowing problems. Even when infections develop many can be safely and effectively treated in the facility if the diagnosis is timely – reducing hospitalization rates both for the specific infection and for sepsis. Review of the OIG report suggests that reducing hospitalization costs by over $1 billion per year – for FFS Medicare beneficiaries alone – is a modest and attainable target. A 2010 report, showed that one third of the dually eligible population in SNFs are hospitalized at least once and over a third of them can be avoidable (Walsh et al., 2010). The same study stated that in 2005, the Medicare program paid $3 billion for potentially avoidable hospitalizations, and Medicaid paid $463 million. Again, these numbers demonstrate the high cost associated with hospitalizations.

CMS through its contractor RTI has developed a 30-day hospitalization rate quality measure for SNFs based on Medicare claims, and PointRight has developed one based on the MDS; both are endorsed by the NQF. However, to date no corresponding measure has been developed for long-stay residents. According to the national MDS data from CMS, there were 437,356 long nursing home stays discharged to an acute hospital in the year ending 2015 Q1. This demonstrates the importance of needing a hospitalization measure for long-stay residents,

In addition to their costs, it is known that hospitalizations are risky and potentially traumatic events for frail elderly patients, frequently associated with a declines in independent function, delirium and/or cognitive decline that may not be reversible, worsening of nutritional status and physical conditioning, and a risk of falls with injury, new pressure ulcers, and hospital-acquired infections They have also been tied to other risks associated with transitions of care such as the increased risk of medication errors. This offers additional motivation for reducing hospitalization rates of SNF residents, further establishing the need for a comprehensive set of performance measures related to this problem, and thus for a measure focusing on long-stay residents and including all payers.

Other published studies confirm the observations and the conclusions reported by the OIG in 2013, e.g., ones from the Kaiser Foundation (Jacobson, 2010), the Commonwealth Fund (Schoen, 2013), MedPAC (MedPAC, 2012) and CMS (Walsh, 2010). Studies by Ouslander have shown that structural and process issues within SNFs have a high impact on the rate of hospitalizations (Ouslander, 2012; Ouslander, 2011), further supporting the hypothesis that hospitalization rates could be reduced by feasible changes in

facilities' operations

Citations can be found at 1c.4.

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.*
PointRight compared the distributions of the PointRight Pro Long Stay Hospitalization Measure over three consecutive one-year measure periods, for a sample of 1,639 facilities that consistently submitted data to PointRight over all three periods and had known outcome rates of 90% or greater. This sample included facilities of various bed counts, various proportions of post-acute versus long-term care, chain and independent facilities, hospital-based and freestanding facilities, and for-profit and nonprofit facilities. The sample was similar to the one used to develop the PointRight Pro 30 measure of post-acute rehospitalizations, an NQF-endorsed measure (measure #: 2375). In the validation of that measure risk adjustment models and measure performance statistics were tested on a full national sample of MDS data obtained from CMS: the performance of the risk adjustment model and of the risk-adjusted measure were similar when tested on the larger samples

The distribution (and the stability) of the adjusted PointRight Pro Long Stay Hospitalization Rate as tested on the 1639-facility sample are shown in the appendix table A.2. For this analysis a single national benchmark rate was used, derived from the measure period July 1, 2013 to June 30, 2014, to avoid confounding the analysis of measure stability with secular changes in national benchmark rates.

To assess the stability of the measure we calculated the Pearson and Spearman correlation statistics (see appendix tables A.3-A.4) for the 12-month periods one quarter apart, for measure periods ending March 31, 2014 to December 31, 2014. Correlations from one quarter to the next ranged between .884 to .894 for the parametric statistic and .877 to .886 for the rank order statistic. The correlations suggest that the measure is adequately stable over short periods, but sufficiently variable to reflect clinically meaningful changes. Stability of the measure is important because large short-term fluctuations in facilities' measured performance would make the measure hard to utilize either for clinical decision-making or for systemic quality improvement.

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

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for 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.*
The MDS allows us to test the impact of race/ethnicity (A1000), gender (A0800), age (A0900) and Medicaid status (A0700). All four of these variables are significant factors in explaining the risk of a long stay hospitalization and are included in one or more of the 4 logistic regression models used to calculate the expected rate of first hospitalizations.

In our selection of sociodemographic factors to test we required that the factor be included on the MDS assessment and that it be consistently and reliably filled out. These factors were age, gender, race/ethnicity and Medicaid beneficiary status as indicated by the patient's having a Medicaid number. (This does not mean that Medicaid was necessarily the payer for every day of the patient's stay – it is Medicaid eligibility that is the indicator of socioeconomic status as this means the patient has low income and few assets.) The items for occupation and education on admission MDS assessments often were missing. We rejected the option of using community-level (e.g., ZIP code based or census based) socioeconomic variables to impute socioeconomic status of individuals, both because of the high error variance implicit in that approach, and because we thought this would make the risk adjustment less acceptable to providers. In our view it would make little sense to them that the adjusted rates of hospitalization for otherwise identical facilities one city block apart would differ because one was in a different census tract or ZIP code from the other.

In testing the above mentioned risk factors we compared their effects in both multi-level fixed effects models and in simple logistic regression models. If a multi-level fixed effects model including a given risk factor candidate explained significantly more variance in hospitalization rates than a simple logistic regression we inferred that part of its effect was via disparities in facility performance correlated with the makeup of the facility's resident population – disparities we did not want to mitigate by risk adjustment. We attributed variance at the individual level to otherwise-unmeasured differences in baseline health status for which risk adjustment

would be appropriate.

Specific risk factors were tested and utilized as follows:

a)Age. This was tested using binary variables for age ranges: <65, 65-69, 70-74, …90 or higher. Of these, only the variable indicating age of 90 or over added significant explained variance to the predictive models.

b)Race/Ethnicity. Individual ethnicities (black/African American, Latino, etc.) that are listed on the MDS were tested as binary variables, as was the constructed variable White/Nonwhite. Only black/African American as a binary variable added significantly to the explained variance of logistic regression models. To determine an appropriate coefficient for the black/African American variable we tested it in a two-level fixed effects model with both facility and individual effects. In this model most of the variance due to black race was associated with the facility level – i.e., facilities with a high proportion of black residents showed worse performance after adjustment for other risk factors - but the variable remained significant at the individual level. We adopted the coefficients from the fixed effects models and forced them into the simple logistic regression models used in our final risk adjustment model.

c)Medicaid status. As with race, Medicaid was associated with higher hospitalization rates, with most of the effect at the facility level – i.e., facilities with high proportions of Medicaid residents had worse outcomes. Nonetheless, even in the two-level fixed-effects model there was an effect of Medicaid status at the patient level. This was most effectively included through the interaction terms white-Medicaid and white-non-Medicaid. The coefficients of these terms in the simple logistic regression models were determined in fixed effects models and forced into the logistic regressions.

d)Gender. There was a strong effect of gender at the patient level. We interpreted this as totally due to health status differences associated with gender and not under the control of the facility.

Univariate effects of the above sociodemographic factors are shown in the appendix table A.5. The strength of the effects supports the inclusion of the factors in the risk adjustment models. The contrast between the fixed effects and the simple logistic regressions for black race, white-Medicaid, and white-non-Medicaid also is shown in appendix table A.6.

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, High resource use
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.**
See section 1.b.1 on the rationale of the measure for a detailed discussion of the epidemiological and resource use factors that make this measure a high priority measure.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
Department of Health and Human Services Office of the Inspector General (2013): Medicare aursing home hospitalization rates merit additional monitoring. (HHS Document OEI-06-11-00040).
http://www.commonwealthfund.org/Publications/Fund-Reports/2013/Sep/Low-Income-Scorecard.aspx

Jacobson, G., Neuman, T., & Damico, A. (2010). Medicare spending and use of medical services for beneficiaries in nursing homes and other long term care facilities: A potential for achieving Medicare saving and improving the quality of care. The Henry J. Kaiser Family Foundation.

Krumholz, H.M. (2013). Post-hospital syndrome- an acquired, transient condition of generalized risk. NEJM, 386(2): 100-102.

MedPAC. (2012) Report to congress: Payment policy. http://medpac.gov/documents/mar12_entirereport.pdf

Mor, V., Intrator, O., Feng, Z., & Grabowski, D.C. (2010). The revolving door of rehospitalizations from skilled nursing facilities. Health Affairs, 29(1): 57-64.

Oslander, J.G., & Maslow, K. (2012). Geriatrics and the triple aim: Defining preventable hospitalizations in the long-term care population. J Am Geriatr Soc., 60(12): 2313-2318.

Ouslander, J.G. & Berenson, R.A. (2011). Reducing unnecessary hospitalizations of nursing home residents. New England Journal of Medicine. 365(13): 1165-1167.

Ouslander, J.G., Lamb, G., Perloe, M., Givens, J.H., Kluge, L., Rutland, T, … Saliba, D. (2010). Potentially avoidable hospitalizations of nursing home residents: Frequency, causes, and costs. J Am Geriatr Soc., 58(4): 627-635.

Ouslander, J.G., Lamb,G., Tappen, R., Herndon, L., Diaz, S., Roos, B.A., … Bonner, A. (2011). Interventions to reduce hospitalizations from nursing homes: Evaluation of the INTERACT II collaborative quality improvement project. J Am Geriatr Soc., 59(4): 745-753.

Polniaszek, S., Walsh, E.G., & Wiener, J.M. (2011). Hospitalizations of nursing home residents: Background and options. Office of the Assistant Secretary for Planning and Evaluation, http://aspe.hhs.gov/daltcp/reports/2011/NHResHosp.pdf

Schoen, C., Radley, D., Riley, P., Lippa, J., Berenson, J., Dermody, C., & Shih A. (2013). Health Care in the two Americas: Findings from the scorecard on the state health system performance for low-income populations. The Commonwealth Fund.

Walsh, E.D., Freiman, M., Haber, S., Bragg, A., Ouslander, J., & Wiener, J.M. (2010) Cost drivers for dually eligible beneficiaries: Potentially avoidable hospitalization from nursing facility, skilled nursing facility, and home and community-based services waiver programs, final task 2 report. RTI International.

Walsh, E.G. & Wiener, J.M. (2011). Hospitalizations of nursing home residents: Background and options. RTI International. CMS Contract Number: HHSP23320095651WC.

Young, H.M., Kurtzman, E., Roes, M., Toles, M., Ammerman, A., & Pace, D. (2011). Measurement opportunities & gaps: Transitional care processes and outcomes among adult recipients of long-term services and supports. Long Term Quality Alliance, Quality Measurement Workgroup.

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

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):*

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination : Readmissions, Safety : Readmissions

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)*
No data dictionary **Attachment:**

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 numerator for the measure is the sum over four quarters of the counts of hospitalizations of the quarterly denominator populations, where hospitalizations comprise discharges directly from the SNF to an acute care hospital.

The count of hospitalizations excludes discharges from the SNF to LTACHs, IRFs, and psychiatric hospitals, and excludes admissions to acute care hospitals that directly follow a discharge from the SNF to a setting other than an acute care hospital.

However, if a patient is discharged from a SNF directly to an acute care hospital during a quarter at risk, the hospitalization will be counted in the numerator even if the patient was discharged to a setting other than an acute care hospital earlier in that quarter.

Hospitalizations are counted over at-risk intervals of 3 months at a time because this period is long enough to yield nonzero numerators even for SNFs with low rates of hospitalization, yet short enough so that almost all of the denominator population will be present in the facility for all, or almost all, of the period. The latter feature makes the calculation simpler than if the risk exposure was calculated by days or weeks.Four quarters of denominators and four quarters of numerators are summed to yield the values for the full measure period.

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.)*
This measure is calculated over twelve months, comprising four consecutive calendar quarters. Each quarter has its own denominator population and its own numerator; the quarterly numerators and quarterly denominators are each summed to create the numerator and the denominator for the 12 month measure period. The measure is updated quarterly.

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.
The numerator for a quarter is the number, during the quarter, of discharges from the SNF directly to an acute care hospital of patients in the denominator population for that quarter as indicated by MDS item A2100=03 'discharge status = acute hospital'. A patient in the quarterly denominator population can contribute multiple times to the quarterly numerator.

Discharges to LTACHs, IRFs, and mental hospitals are not included in the numerator, nor are acute hospital admissions directly following a discharge from the SNF to a setting other than an acute care hospital. As noted above, if a patient is discharged from a SNF directly to an acute care hospital during a quarter at risk, the hospitalization will be counted in the numerator even if the

patient was discharged to a setting other than an acute care hospital earlier in that quarter.

The numerator for the measure is the sum of the quarterly numerators for the four quarters in the 12 month measure period.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The quarterly denominator population consists exactly of those patients present in the SNF on the first day of the quarter (the "snapshot date") who meet the criterion for long stay on that date. The denominator for a quarter is the number of patients in the quarterly denominator population. The denominator for the measure is the sum of the quarterly denominators for the four quarters in the 12 month measure period.

The criterion for a patient's having a long stay is a cumulative length of stay in the facility of more than 100 days as of the snapshot date. The cumulative length of stay of a patient is the length of the current stay as of the snapshot date and plus the full lengths of stay of any previous stays that are linked to it. According to the criteria for linkage of stays used in the present measure, a stay in a SNF is linked to a subsequent stay in the SNF if the patient was discharged from the SNF to the community and was readmitted to the SNF within 10 days or fewer. All stays in a sequence of linked stays are included in the sum of days used to determine a patient's cumulative length of stay. In these criteria the term "community" comprises private residences and all organized settings that are primarily residential in character, including senior housing, independent living facilities, board and care homes, and assisted living facilities.

A patient can contribute multiple times to the denominator for a 12 month measure period. For example, a resident continuously present in the facility for a full year would contribute four to the denominator.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk : Dual eligible beneficiaries, Populations at Risk : Individuals with multiple chronic conditions, Senior Care

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 denominator population for a quarter is a subset of the patients present in the SNF on the snapshot date (the first day of the quarter). A patient is in that subset if his or her cumulative length of stay as of the snapshot date is more than 100 days.

The cumulative length of stay is calculated by taking the length of stay of the current admission as of the snapshot date and adding the lengths of stay of any linked stays at the same SNF. The length of the current admission as of the snapshot date is the snapshot date minus the entry date for the current admission, which is MDS item A1600. A stay is linked to a subsequent stay if the patient is discharged to the community (A2100=01) and admitted to the same SNF within 10 days or less (i.e., A1600 for the second stay minus A2100 for the first stay is less than or equal to 10 days).

The denominator for a quarter is the number of residents in the denominator population for that quarter. The denominator for the measure, which reports on a full year's performance, is the sum of the denominators for the four quarters that constitute that year.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
There are no exclusions from the denominator; all patients in the facility on the snapshot date who meet the long stay criterion on that date are included. However, the measure will not be reported for a SNF if the annual unknown outcome rate is greater than 10%.The definition of the annual unknown outcome rate is provided in S.11.

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)*
The denominator of the annual unknown outcome rate is the sum of the four quarterly denominators. The numerator of the annual unknown outcome rate is the sum over the four quarters of the numbers of quarterly denominator patients with an unknown outcome in the quarter at risk. An outcome is regarded as unknown if it cannot be reasonably inferred or conservatively imputed. The numerator of the unknown outcome rate is the sum of the quarterly unknown outcome counts for the four quarters in the year. The quarterly unknown outcome count is the number of patients in the quarterly denominator for whom it is not known and cannot be reasonably inferred or imputed that the patient was or was not hospitalized during the quarter (e.g. they did not have an MDS discharge assessment completed or a subsequent regularly scheduled MDS assessment completed indicating they resided in the SNF the entire time). It would be known that a patient was hospitalized during the quarter if he or she had a discharge MDS with an

acute care hospital as a discharge disposition. It would be known that a patient was not hospitalized during the quarter if he or she had an MDS assessment with an assessment reference date (item A2300) following the end of the quarter at risk and had an admission date (item A1600) on or prior to the snapshot date. If the patient has a discharge MDS during the quarter at risk and is subsequently readmitted to the same SNF within the same quarter it is assumed that there was a second discharge during that quarter (whether to an acute care hospital or elsewhere) if and only if there is a discharge MDS with an assessment reference date within that quarter. If there is an admission to the SNF from an acute care hospital during the quarter at risk but no preceding discharge MDS, we then make the inference that the preceding discharge was directly to an acute care hospital and the inferred discharge is counted in the numerator of the measure. If a patient has no MDS assessment of any kind with an assessment reference date 100 days or fewer after the latest MDS in the interval starting 10 days before the snapshot date and ending one day before the end of the quarter the patient's outcome is regarded as unknown. If the count N of patients with unknown outcomes is 10% or less of the denominator, N*0.8 is added to the numerator (see S.22). If N is more than 10% of the denominator the measure is not reported.

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)*
The risk adjustment model for PointRight Pro Long Stay Hospitalization Rate begins by segmenting the quarterly denominator population for each quarter into four groups based on the duration of the patient's current stay in the SNF. The denominator population is segmented into these four groups because even after controlling for the other risk adjusters, significant variation by length of stay remains and the coefficients within the length of stay groups are different. For each group the risk of one or more discharges from the SNF directly to an acute care hospital during the quarter is estimated by a logistic regression. (Note that the dependent variable is a binary variable rather than the count of hospitalizations of the patient during the quarter.) The independent variables in each logistic regression model come from the patient's most recent MDS 3.0 assessment prior to the snapshot date that has the variable. (Not all of the independent variables in the logistic regressions are present on every type of MDS assessment; this implies that it is sometimes necessary to extract independent variables from two or more discrete MDS assessments.)

The four logistic regression models use subsets of the following set of independent variables. In S.18 below, MDS items corresponding to each listed variable are provided.

Active Diagnoses (A diagnosis is "active" if it affects the patient's current clinical status or treatment plan. An active diagnosis must be documented in the medical record by a physician or physician extender to be checked off in the MDS. Diagnoses are used in the model only if they are indicated in check boxes on Section I of the MDS; if they are indicated by write-in codes in MDS item I8000 they are not utilized in determining the values of the independent variables.):
-Anemia
-Chronic Lung Disease (including Asthma and COPD) -Chronic Lung Disease receiving oxygen therapy at least one time in the 14 days prior to the MDS date
-Diabetes Mellitus receiving insulin at least once in the 7 days prior to the MDS assessment reference date
-Gastroesophageal Reflux Disease (GERD) or Ulcer (esophageal, gastric, or duodenal)
-Heart Failure
-Hypertension
-Viral Hepatitis
-Neurogenic Bladder
-Renal Insufficiency, Renal Failure, or End-Stage Renal Disease

Incontinence:
-Total bowel incontinence

Demographics:
-Age 90 or over
-Male

Medications received at least once within the 7 days prior to the MDS assessment reference date:
-Anticoagulant
-Antibiotic

Context of Care:
-Current stay began with admission from an acute care hospital
-In this SNF 6 months before the snapshot date (whether or not in the facility continuously for the 6 months preceding the snapshot date
-In this SNF 12 months before the snapshot date (whether or not in the facility continuously for the 12 months preceding the snapshot date
-Natural log of (the length of the current stay as of the snapshot date minus 100 days). (Linked stays are not included in this calculation.)

Symptoms:
-Dyspnea (shortness of breath or trouble breathing) on exertion

Skin condition:
-Surgical wound(s)

Hospice Status:
-Receiving hospice care while resident in the facility, at some time during the 14 days prior to the MDS assessment reference date

Treatments (given in the facility at least once in the 14 days preceding the MDS assessment reference date):
-IV fluid or medication
-Oxygen therapy

Socioeconomic Status:
- Medicaid beneficiary (as indicated by having a Medicaid number or having a Medicaid number pending)
- Black or African-American race/ethnicity (as described the patient or family, either as a sole identity or one of several, e.g., black and Caucasian, black and Latino)

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.
Provided in response box S.15a

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
The risk-adjusted hospitalization rate is the observed (unadjusted) rate, divided by the expected rate, multiplied by the national benchmark rate, where the latter is the (observed) national numerator over the (observed) national denominator, and not the average of all SNF-level observed rates. That is, [Adjusted rate of (direct) hospitalizations] = [Observed rate of hospitalizations] / [Expected rate of hospitalizations] × [National benchmark rate].

The calculation of a facility's expected rate of hospitalizations makes use of a predictive model for the first hospitalization of a member of the denominator population during the quarter at risk. As always throughout this measure description, "hospitalization" means discharge of a patient directly to an acute care hospital, and excludes hospitalizations taking place after a patient is discharged to the community, and hospitalizations at psychiatric hospitals, IRFs and LTACHs. The predictive model comprises four logistic regression models, of which one is applied to a patient after he or she classified into one of four discrete groups based on the length of the current stay in the SNF.

The estimated probability of a first hospitalization (i.e., of any hospitalization) for an individual patient in the denominator population is calculated as follows:

1.Categorize the patient as belonging to one of the following four groups according to the history of the current stay in the SNF as of the snapshot date.

a.Long Stay Group 1 - Interrupted Long Stay – Patients who meet the long stay criterion but whose length of stay of the current admission as of the snapshot date is 100 days or fewer.

b.Long Stay Group 2 – Patients whose length of stay of the current admission as of the snapshot date is between 101 days and 181 days, inclusive.

c.Long Stay Group 3 – Patients whose length of stay of the current admission as of the snapshot date is between 182 and 364 days, inclusive.

d.Long Stay Group 4 – Patients whose length of stay of the current admission as of the snapshot date is 365 days or more.

2.Calculate the log odds of a first hospitalization (i.e., the log odds of any hospitalization) for the resident using the linear equation corresponding to the applicable long stay group. The values of the independent variables in the calculation come from the most recent MDS assessment prior to the start of the quarter that has a value for the variable. (For example, some independent variables are scored on admission, annual and significant change MDS assessments, but not on routine quarterly assessments. For a resident in Long Stay Group 2 the most recent MDS before the start of the quarter at risk would almost always be a routine quarterly MDS assessment, and in that case the values of those variables would almost always be drawn from the admission MDS.)

3.Convert the log odds, L, into a probability estimate, P, using the inverse logit function: P=1/(1+exp(-L)).

The expected number of first hospitalizations for a quarterly denominator population is the sum of the estimated probabilities of a first hospitalization over all the patients in that population. The expected number of total hospitalizations for the quarterly denominator population is the expected number of first hospitalizations multiplied by 1.2528. The expected number of total hospitalizations for the entire measure period is the sum of the expected numbers of total hospitalizations for the four quarters in the measure period. Equivalently, the expected number of first hospitalizations for the full 12-month denominator population is the sum of the expected numbers of first hospitalizations for the four quarters in the measure period, and the expected number of total hospitalizations for the full 12-month period is this sum times 1.2528.

The expected rate of total hospitalizations for the measure period is the expected number of total hospitalizations divided by the total denominator, which is the sum of the quarterly denominators for the four quarters in the measure period.

The conversion of the expected number of first hospitalizations to the expected number of all hospitalizations is based on an analysis that compared the rates of first hospitalizations (F) with the rates of all hospitalizations (A) for a national sample of 1,029 SNFs over the measure period July 1, 2013 to June 30, 2014, in which each SNF in the sample had a 100% known outcome rate. The curve approximating the relationship between F and A was defined by the linear equation A=1.2528. The R-square statistic for this approximation was 0.96. The statistic was not improved significantly if the linear equation was replaced with a log-linear, exponential or polynomial equation.

Each of the predictors in the equations defined above is defined as follows:

Active Diagnoses:
Anemia: If I0200 =1 then Variable=1; else Variable=0.
Chronic Lung Disease: If I6200 >= 1 then Variable=1; else Variable=0.
Chronic Lung Disease on oxygen: If I6200 = 1 & O0100C2 = 1 then Variable=1; else Variable=0.
Diabetes Mellitus on insulin: If N0350A >= 1 and I2900 = 1 then Variable=1; else Variable=0.
Gastroesophageal Reflux Disease (GERD) or Ulcer: If I1200=1 then Variable=1; else Variable=0.
Heart Failure: If I0600 =1 then Variable=1; else Variable=0.
Hypertension: If I0700=1 then Variable=1; else Variable=0.
Viral Hepatitis: If I2400 =1 then Variable=1; else Variable=0.
Neurogenic Bladder: If I1550 =1 then Variable=1; else Variable=0.
Renal Insufficiency, Renal Failure or End-Stage Renal Disease: If I1500=1 then Variable=1; else Variable=0.

Incontinence:
Total bowel incontinence: If H0400=3 then Variable=1; else Variable=0.

Demographics:
Age 90 or over: If date of birth (A0900) is present and age (snapshot date – date of birth) >= 90 years then Variable=1; else Variable=0.Male: If A0800=1 then Variable=1; else Variable=0.

Medications received at least once in the 7 days preceding the MDS assessment reference date:
Anticoagulant: If N0410E ? 0 then Variable=1; else Variable=0.
Antibiotic: If N0410F=1 then Variable=1; else Variable=0.

Context of Care:
Current stay began with admission from an acute care hospital: If A1800 = 3 then Variable=1; else Variable=0.
In this SNF 6 months before snapshot date (any stay): If for any stay (snapshot date – 182) lies between the entry date (A1600) and the discharge date (A2000), inclusive, then Variable=1; else Variable=0.
In this SNF 12 months before snapshot date (any stay): If for any stay (snapshot date – 365) lies between the entry date (A1600) and the discharge date (A2000), inclusive then Variable=1; else Variable=0.
Natural log of (Length of current stay minus 100): natural log ((snapshot date – current stay entry date (A1600 for the current stay) – 100 days).

Symptoms:
Dyspnea on exertion: If J1100A = 1 then Variable=1; else Variable=0.

Skin Condition:
Surgical wound(s): If M1040E=1 then Variable=1; else Variable=0.

Hospice Status:
Receiving hospice care (in the SNF at some time during the 14 days preceding the MDS assessment reference date): If O0100K2 = 1 then Variable=1; else Variable=0.

Treatments Received in the SNF at Some Time During the 14 Days Preceding the MDS Assessment Reference Date:
IV fluid or medications: If O0100H2 = 1 then Variable=1; else Variable=0.
Oxygen therapy: If O0100C2=1 then Variable=1; else Variable=0.

Socioeconomic Status:
Medicaid Status: if A0700 ? "N" or A0700 is missing then Variable = 1; else Variable = 0.
Black or African-American Race/Ethnicity: If A1000c = 1 then Variable = 1; else Variable = 0;

A table containing model coefficients (Table A.1) and a scatter plot of the independent variable prevalence (Figure A.1) is found in appendix at S.15.

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.)*
The formula for a facility's adjusted PointRight Pro Hospitalization Rate is: [Observed rate of all hospitalizations]/[Expected rate of all hospitalizations]*[National average rate of all hospitalizations].

The observed and expected rates are updated quarterly and the national benchmark rate is updated annually; the national benchmark rate used in the calculation is the most recently calculated benchmark rate at the time the observed and expected rates are calculated.

The procedure for calculating the adjusted rate is (a numeric example can be found in the appendix at Figure A.3):

1) Calculate the observed rate.
   - The denominator for a quarter is the number of residents present in the facility on the first day of a calendar quarter who

qualify as long stay on that day

•The numerator for a quarter is number of hospitalizations of residents in the denominator population for that quarter, where hospitalization means discharge from the SNF directly to an acute care hospital, either with no return to the SNF or with return to the SNF after at least one midnight outside the SNF. The numerator excludes: (1) hospitalizations occurring after a patient has been discharged somewhere other than an acute care hospital and (2) hospitalizations at psychiatric hospitals, rehabilitation hospitals, or LTACHs. The numerator includes: (1) "observations stays" if these involve at least one midnight away from the SNF and (2) "planned" hospitalizations.

•The observed PointRight Pro Long Stay Hospitalization Rate is the sum of the four quarterly numerators divided by the sum of the four quarterly denominators.

2. Calculate the expected rate.

•Calculate the expected number of first hospitalizations of the quarterly denominator population for each of the four quarters in the measure period and sum them; multiply the sum by 1.2528 to obtain the expected number of total hospitalizations for the 12-month measure period. Divide this number by the sum of the quarterly denominators to get the expected rate for the measure period.

3. Calculate the national benchmark rate (this will be updated annually, while the observed and expected rates will be updated quarterly).

•The national benchmark rate is the observed PointRight Pro Long Stay Hospitalization Rate for a denominator population consisting of the denominator populations for all SNFs in the largest available national sample that have complete non-discharge MDS data for all of their patients for all four quarters in the measure period and have 100% known outcomes for all patients in their denominator populations for all four quarters in the measure period. For a given member of a quarterly denominator population a known outcome means either that the patient had a discharge MDS submitted with a discharge date within the quarter and a discharge destination filled in, that the patient was readmitted from an acute care hospital during the quarter, or that the patient had a quarterly or other MDS submitted in the 100 days following the end of the quarter that gave an admission date prior to the snapshot date for the given quarter.

Procedure for Calculating the Measure:

1.Establish a 12-month measure period comprising of four calendar quarters (each three months in length). For each quarter, the (quarterly) denominator is the number of residents who qualify as long stay for that quarter, i.e. whose cumulative length of stay as of the snapshot date (the first day of the quarter) is more than 100 days. (Cumulative length of stay is defined as the sum of the lengths of stay of the current stay and all stays linked to it.) The sum of the quarterly denominators for the four quarters constitutes the denominator for the measure period.

2.For the quarterly denominator population determine the number of (direct) acute care hospitalizations of the residents in that quarter (the quarterly numerator). The count of the hospitalizations is the quarterly numerator. The sum of the quarterly numerators for the four quarters constitutes the numerator for the measure. As noted above the count includes only admissions to acute care hospitals directly from the SNF. Planned (or presumptively planned) hospitalizations are included, as are observation stays. Hospitalizations subsequent to a discharge somewhere other than an acute care hospital, and hospitalizations at LTACHs and specialty hospitals are excluded.

3.Divide the total numerator by the total denominator to obtain the observed rate for the SNF.

4.Calculate the estimated probability of a first hospitalization for each member of each quarterly denominator population using the predictive model described above, and sum these probabilities to get the expected number of first hospitalizations per quarter for the total 12 month denominator population. Sum these expected numbers over the four quarters of the measure period to get the expected number of first hospitalizations for the measure period. Multiply this result by 1.2528 to get the expected number of total hospitalizations for the total measure period denominator population, and divide this by the total measure period denominator to get the expected PointRight Pro Long Stay Hospitalization Rate for the measure period.

5.Divide the observed rate by the expected rate and multiply by the most recent national benchmark rate to obtain the Adjusted PointRight Pro Long Stay Hospitalization Rate.

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

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

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
The measure will not be reported for a SNF if the annual unknown outcome rate is greater than 10%.The definition of the annual unknown outcome rate is provided in S.11.

On occasion facilities fail to submit MDS assessments adhering to the MDS submission schedule specified by regulation. This can result in a patient being included in a quarterly denominator population but not having a known outcome during the quarter following the snapshot date. The unknown outcome rate for the measure period is the sum of the counts of patients over the four quarterly denominator populations that have unknown outcomes, divided by the measure period denominator.

In the above definition "known outcome" includes both explicitly known outcomes and outcomes that can be imputed with high (though not perfect) accuracy. Specifically:

If a patient in the quarterly denominator population has no discharge MDS assessment but has an admission MDS assessment during the quarter indicating admission from an acute care hospital, it is assumed (imputed) that the patient had been discharged from the SNF to an acute care hospital.This is the most common explanation for this pattern of MDS data. However, a small percentage of such MDS submission patterns arise from long stay patients being discharged to the community, failing to thrive in the community, admitted to an acute care hospital, and then sent back to the SNF from which they were discharged.

If a patient in the quarterly denominator population has no MDS assessment of any kind during the quarter at risk and has no subsequent MDS assessment indicating an admission date prior to the quarter's snapshot date, it is assumed that the patient was discharged during the quarter.Nationally, approximately 80% of discharges of long-stay SNF patients with known outcomes are to acute care hospitals.For this reason 0.8 is added to the quarterly numerator for each member of the quarterly denominator with an unknown outcome.

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

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.
SNF-Minimum Data Set (MDS) version 3.0.

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)*
Available in attached appendix at A.1

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)*
Post Acute/Long Term Care Facility : Nursing Home/Skilled Nursing 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
2a. Reliability – See attached Measure Testing Submission Form **2b. Validity – See attached Measure Testing Submission Form** PointRight_Pro_Long_StayTM_Measure_Testing_Final.docx

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

Measure Number (*if previously endorsed*)**:** N/A
Measure Title: PointRight® Pro Long Stay™ Hospitalization Measure
Date of Submission: 1/29/2016z
Type of Measure:

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

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 decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator

exclusion category computed separately). [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 that influence the measured outcome (but not factors related to disparities in care or the quality of care) 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: Skilled Nursing Facility (SNF) Minimum Data Set (MDS) 3.0	☒ other: Skilled Nursing Facility (SNF) Minimum Data Set (MDS) 3.0

SNF MDS 3.0 data came from a national sample of 3,539 SNFs; these SNFs all were subscribers to online analytics provided as a commercial service by the measure developer. SNFs came from 47 states, and included both nonprofit and for-profit facilities; both independent and chain-affiliated facilities; and both hospital-based and freestanding facilities. Facilities included in the sample consistently submitted 100% of their MDS 3.0 assessments for at least 2 years. Data used in estimating the risk adjustment models comprised of patients who came from 2,516 SNFs that had a minimum known outcome rate of 90%. Data used in testing measure validity included all residents (including those with imputed outcomes) 2,096 SNFs that had a known outcome rate of at least 90%. Data used in establishing the relationship between the rate of first hospitalizations and the rate of total hospitalizations came from 1,029 SNFs that had a known outcome rate of 100%.

To confirm the representativeness of the PointRight sample of MDS assessments, we analyzed the agreement between MDS discharge status codes, and the presence of inpatient admissions from the Medicare Part A claims data or dates of death from the enrollment data; this sample contained 1,087,766 MDS discharges from CY2012, which covered 14,620 facilities. Additionally, for certain facility-level characteristics, we compared the PointRight sample against the 15,643 SNFs in the December 2014 release of Nursing Home Compare; and compared certain patient characteristics using MDS assessments for long stay residents (in the facility for >100 days) from the 3rd quarter of 2014, which included 1,087,766 stays spanning 14,620 SNFs.

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*).

Skilled Nursing Facility (SNF) Minimum Data Set (MDS) 3.0

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

The PointRight® Pro Long Stay™ Hospitalization Measure risk adjustment models were fit on data from April 1, 2013 to November 30, 2014, from which covariates and the dependent variable were ascertained for a 12 month risk period between July 1, 2013 and June 30, 2014. Data covering the 12 month risk period January 1st, 2014 to December 31, 2014

were utilized in various reliability and validity tests. MDS discharge status codes and Medicare Part A claims data used to demonstrate the representativeness of the PointRight® sample of MDS assessments came from CY 2013.

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

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

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

The PointRight Pro Long Stay Hospitalization Measure was developed on MDS 3.0 assessments collected from skilled nursing facilities that purchased analytics services from PointRight. Modeling and estimation of risk adjustment were performed on 2,516 SNFs using four snapshots dates: Q3 2013, Q4 2013, Q1 2014 and Q2 2014. Testing and analysis were performed on 2,096 SNFs using calendar year 2014. In table 1 below, find facility level descriptive statics on these SNFs and how they compare to the national population of skilled nursing facilities.

TABLE 1. Testing Sample Facility Level Descriptive Statistics

Metric	**Nation (N=15,643)**		**PointRight (N=2,096)**	
	N	%	N	%
Part of chain	8,748	56.4%	1,555	80.3%
Not part of chain	6,755	43.6%	382	19.7%
For profit	10,916	69.8%	1,666	86.0%
Metric	**Nation (N=15,643)**		**PointRight (N=2,096)**	
	N	%	N	%
Government	958	6.1%	36	1.9%
Hospital-based	871	5.6%	10	.5%
Not hospital-based	14,772	94.4%	1,928	99.5%
Medicare certified facilities	15,169	97.0%	1,937	99.9%
Non Medicare certified facilities	474	3.0%	1	0.1%
Resident count less than 50	3,765	24.1%	196	10.1%
Resident count greater than 50, less than 110	7,994	51.1%	1,078	55.6%
Resident count greater than 110	3,884	24.8%	664	34.3%

The PointRight sample contained facilities of various bed counts, chain vs. independent ownership, hospital based vs. non-hospital based affiliation and for-profit vs. nonprofit designation. The PointRight sample had greater proportions of

large for-profit chain facilities than the national SNF population; all provider types were sufficiently represented in the PointRight sample.

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 PointRight Pro Long Stay Hospitalization Measure measures the rate of hospitalization for a SNF's Long Stay population over a 12-month measure period. The measure sums numerators and denominators from four snapshot dates – the first day of each calendar quarter within the measure period. Using MDS from 2,096 SNFs, our dataset contained more than 150 thousand patients, present in the facility, on each of the 4 snapshot dates contained in the CY 2014 measure period. The table below presents the demographics of the denominator sample from one representative snapshot date, 10/1/2014.

TABLE 2. Characteristics of Patients

MDS Variables	Prevalence of PointRight Client Sample (N=2,096)	Prevalence of Nation (N=14,620)
Age		
<65	19.4%	15.4%
65-74	16.6%	15.5%
75-84	28.6%	26.4%
85-89	16.7%	19.1%
90 or over	18.7%	23.6%
Gender		
Female	68.6%	67.5%
Male	31.4%	32.6%
MDS Variables	**Prevalence of PointRight Client Sample (N=2,096)**	**Prevalence of Nation (N=14,620)**
Race		
Asian	1.8%	1.9%
Black or African American	16.4%	14.1%
Hispanic Latino	5.1%	5.2%
White	73.6%	76.2%
Other	3.1%	3.0%
Medicaid Beneficiary		
Yes	80.9%	N/A
No	19.1%	N/A
Admission Setting		
Acute Hospital	82.9%	74.7%
Active Diagnosis		
Anemia	31.1%	29.2%
Asthma, COPD, or Chronic Lung Disease whether or not on oxygen	21.1%	20.4%

Asthma, COPD, or Chronic Lung Disease on oxygen	6.2%	N/A
Diabetes on insulin	20.7%	32.4%
Gastroesophageal Reflux Disease (GERD) or ulcer	31.5%	33.9%
Heart Failure	20.2%	19.3%
Hypertension	76.6%	75.1%
Viral Hepatitis	0.6%	0.6%
Neurogenic bladder	3.0%	2.7%
Renal failure or insufficiency	9.8%	10.0%
Incontinence		
Total bowel incontinence	34.7%	31.4%
Medications Received		
Anticoagulant within 7 days prior to ARD	14.6%	12.3%
Antibiotics within 7 days prior to ARD	1.1%	11.0%
Symptoms		
Dyspnea on exertion	7.1%	7.5%
Skin		
Surgical wound(s)	1.8%	2.1%
Hospice Status		
Receiving hospice care	5.1%	4.7%
Recent Treatments		
IV fluid or meds within 7 days before last MDS	1.3%	1.3%
Oxygen in 7 days before last MDS	10.8%	11.2%

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 majority of measure reliability and validity testing was conducted on the measure development sample of 2,096 SNFs described above, which as noted above is national and provides a good representation of all major demographic categories and provider types, though it is not a random sample of all U.S. SNFs.

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 tested black/non-black, Medicaid/non-Medicaid, and the interaction between these binary variables for their relationships with long-stay hospitalization rates. We did not utilize the patient-level variable concerning occupation, because it often is not completed on the admission MDS or the admission MDS is not available for analysis; also, it is an unreliable indicator of the patient's primary lifetime occupation. We did not use the two language-related items on the MDS, because they often are missing or unreliable.

Race/ethnicity items other than black/non-black either did not have significant patient-level effects in a fixed effects model (Hispanic/Latino) or our sample was insufficient (American Indian or Alaska Native; Native Hawaiian or Pacific Islander).

We rejected geographically-based proxy variables for two reasons:

1. We did not want to take the risk of adjusting away true disparities in care quality that might be found comparing SNFs in poorer neighborhoods with those in richer ones.

2. Particularly for the long-stay SNF population, most of whom have the SNF as their primary residence, the ZIP code or census tract of the patient is simply that of the SNF. This may not be indicative of the patients' socioeconomic status as it is for community-dwelling patients.

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*)

Agreement of Model Independent Variables-

For the MDS data items utilized as risk adjustment covariates a comparison of the variables prevalence between the testing sample of 2,096 SNFs and the national population was performed. Table 2, displayed above demonstrates that the PointRight client database although not a random sample of patients, is representative of the national population.

Reliability of Rates over Time -

To assess the reliability of the overall measure we analyzed change from quarter to quarter in the observed and adjusted long-stay hospitalization rates. We reasoned that a SNF's underlying probability of its long-stay patients hospitalizing, and the characteristics of its long-stay patient population, are unlikely to change greatly over a 3 month period, so that most of the change from quarter to quarter will be related to limitations on measurement reliability. Some of the reliability limitations will reflect error in the measurement of the dependent variable or risk adjustment covariates on the MDS, but most is likely to reflect changes in the characteristics of the long-stay population from one snapshot date to the next. If correlation coefficients – both parametric and non-parametric – are relatively high when consecutive quarters are compared, we infer that the combination of measurement-related variability and sampling-related variability is acceptable.

Stability of Facility Level Adjusted Rate Bootstrapping –

To further test the reliability of the measure, adjusted rates for the measure period CY 2014 were recalculated for our testing sample, where a random sample of stays was drawn with replacement for each facility. We then reviewed the distribution of differences between facilities' original adjusted rates and the rates calculated with resampling. If the distribution of differences has a small variance and a mean of 0 we can assume the measure is acceptably stable.

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*)

Agreement of Model Independent Variables -

Results found in Table 2.

Reliability of Rates over Time –

TABLE 3. Pearson Correlation Coefficients of PointRight® Pro Long Stay™ Hospitalization Risk Adjusted Rates by Quarter

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0				
	Adjusted Hospitalization Rates Measure Period Jan 2014 to Dec 2014	Adjusted Hospitalization Rates Measure Period Oct 2013 to Sept 2014	Adjusted Hospitalization Rates Measure Period July 2013 to June 2014	Adjusted Hospitalization Rates Measure Period April 2013 to Mar 2014
Adjusted Hospitalization Rates Measure Period -Jan 2014 to Dec 2014	1.000 2096	0.888 <.0001 2062	0.743 <.0001 2051	0.583 <.0001 2033
Adjusted Hospitalization Rates Measure Period -Oct 2013 to Sept 2014	0.888 <.0001 2062	1.000 2062	0.884 <.0001 2047	0.744 <.0001 2028
Adjusted Hospitalization Rates Measure Period -July 2013 to June 2014	0.743 <.0001 2051	0.884 <.0001 2047	1.000 2051	0.894 <.0001 2031
Adjusted Hospitalization Rates April 2013 to Mar 2014	0.583 <.0001 2033	0.744 <.0001 2028	0.894 <.0001 2031	1.000 2033

TABLE 4. Pearson Correlation Coefficients of PointRight® Pro Long Stay™ Hospitalization Observed Rates by Quarter

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations				
	Observed Hospitalization Rates Measure Period Jan 2014 to Dec 2014	Observed Hospitalization Rates Measure Period Oct 2013 to Sept 2014	Observed Hospitalization Rates Measure Period July 2013 to June 2014	Observed Hospitalization Rates Measure Period April 2013 to Mar 2014
Observed Hospitalization Rates Measure Period -Jan 2014 to Dec 2014	1 2096	0.9304 <.0001 2062	0.83829 <.0001 2051	0.72504 <.0001 2033
Observed Hospitalization Rates Measure Period -Oct 2013 to Sept 2014	0.9304 <.0001 2062	1 2062	0.93729 <.0001 2047	0.84734 <.0001 2028
Observed Hospitalization Rates Measure Period -July 2013 to June 2014	0.83829 <.0001 2051	0.93729 <.0001 2047	1 2051	0.94118 <.0001 2031
Observed Hospitalization Rates April 2013 to Mar 2014	0.72504 <.0001 2033	0.84734 <.0001 2028	0.94118 <.0001 2031	1 2033

TABLE 5. Spearman Correlation Coefficients of PointRight® Pro Long Stay™ Hospitalization Risk Adjusted Rates by Quarter

Spearman Correlation Coefficients Prob > |r| under H0: Rho=0				
	Adjusted Hospitalization Rates Measure Period Jan 2014 to Dec 2014	Adjusted Hospitalization Rates Measure Period Oct 2013 to Sept 2014	Adjusted Hospitalization Rates Measure Period July 2013 to June 2014	Adjusted Hospitalization Rates Measure Period April 2013 to Mar 2014
Adjusted Hospitalization Rates Measure Period -Jan 2014 to Dec 2014	1.000 2096	0.877 <.0001 2062	0.738 <.0001 2051	0.588 <.0001 2033
Adjusted Hospitalization Rates Measure Period -Oct 2013 to Sept 2014	0.877 <.0001 2062	1.000 2062	0.881 <.0001 2047	0.740 <.0001 2028
Adjusted Hospitalization Rates Measure Period -July 2013 to June 2014	0.738 <.0001 2051	0.881 <.0001 2047	1.000 2051	0.886 <.0001 2031
Adjusted Hospitalization Rates April 2013 to Mar 2014	0.588 <.0001 2033	0.740 <.0001 2028	0.886 <.0001 2031	1.000 2033

TABLE 6. Spearman Correlation Coefficients of PointRight® Pro Long Stay™ Hospitalization Observed Rates by Quarter

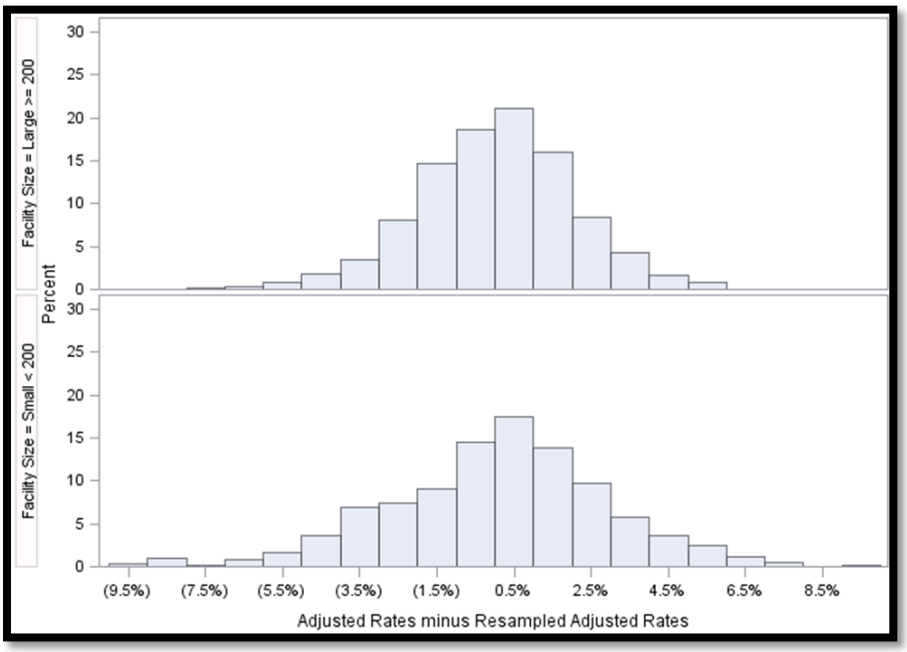
Spearman Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations				
	Observed Hospitalization Rates Measure Period Jan 2014 to Dec 2014	Observed Hospitalization Rates Measure Period Oct 2013 to Sept 2014	Observed Hospitalization Rates Measure Period July 2013 to June 2014	Observed Hospitalization Rates Measure Period April 2013 to Mar 2014
Observed Hospitalization Rates Measure Period -Jan 2014 to Dec 2014	1 2096	0.91975 <.0001 2062	0.81484 <.0001 2051	0.69665 <.0001 2033
Observed Hospitalization Rates Measure Period -Oct 2013 to Sept 2014	0.91975 <.0001 2062	1 2062	0.92667 <.0001 2047	0.82625 <.0001 2028
Observed Hospitalization Rates Measure Period -July 2013 to June 2014	0.81484 <.0001 2051	0.92667 <.0001 2047	1 2051	0.93137 <.0001 2031
Observed Hospitalization Rates April 2013 to Mar 2014	0.69665 <.0001 2033	0.82625 <.0001 2028	0.93137 <.0001 2031	1 2033

Table 7. Distribution of Differences between Facility Adjusted Rates and Resampled Adjusted Rates

Quantiles	
Quantile	**Difference in Rates: (Adj Rates -Resampled Adj Rates)**
100% Max	10.4%

99%	5.6%
95%	3.6%
90%	2.7%
75% Q3	1.2%
50% Median	-0.2%
25% Q1	-1.7%
10%	-3.3%
5%	-4.4%
1%	-6.8%
0% Min	-12.2%

Figure 1. Histogram of Differences between Facility Adjusted Rates and Resampled Adjusted Rates by Facility Size



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?*)

Agreement of Model Independent Variables -

48% of the risk adjustment model covariates, that were comparable, were found to have prevalence within 5% of the prevalence found in the national sample. Two third (14 out of 21) risk adjustment model covariates, that were comparable, were found to have prevalence within 10% of the prevalence found in the national sample. Although the measure testing sample is not a random sample of all U.S. SNF patients, all the model IV cohorts are sufficiently represented in our sample.

Reliability of Rates Over Time –

Correlations from one quarter to the next ranged between .884 to .894 for the parametric statistic and .877 to .886 for the rank order statistic. The correlations suggest that the measure is adequately stable over short periods, but sufficiently variable to reflect clinically meaningful changes.

Stability of Facility Level Adjusted Rate Bootstrapping -

Reviewing the distribution of facility level differences between adjusted hospitalization rates and resampled adjusted rates illustrates the PointRight Pro Long Stay Hospitalization Measure has a high level of precision. 65.6% of the PointRight sample had a difference in adjusted rates of less than 2% and only 2.1% of facilities had a difference greater than 5%. The mean difference was .0008%.

1,445(70%) of the facilities in our sample had a denominator greater than 200 patient quarters. For these larger faculties we noticed, as expected, the variance of the distribution in differences shrinks. Smaller facilities will have less measure reliability, but we found the variance acceptable even for facilities with denominators between 30 and 200.

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)*

Agreement of Model Dependent Variables -

For the dependent variable of acute care hospitalization we compared the identification of hospitalization events of Medicare FFS beneficiaries based on the MDS and hospitalization events based on Medicare FFS claims. Because MDS data include all payers, not just Part A Medicare patients, we restricted the MDS discharges to those where the patient was enrolled in Part A Medicare when he/she was discharged, and who had data in our extract of inpatient and other claims data for patients who had a SNF claim in calendar year 2012. We used 2012 MDS data, claims data and enrollment data (for Part A enrollment and dates of death) because it was the most recent in our database. We would not expect significant differences in the match rates for newer data, and those differences would likely be improvements, given CMS's MDS focus surveys and other MDS data quality initiatives.

Performance Measure Score- Correlation with SNF Industry Measures of Quality:

To test construct validity of the PointRight Pro Long Stay Hospitalization measure we tested the relationship of the measure with the various components of the CMS Five-Star ratings for SNFs and its correlation with CMS's long-stay Quality Measures. We hypothesized that facilities with higher star ratings would have lower adjusted long-stay hospitalization rates, and specifically that the relationship would be stronger for the long-stay quality measures and RN staffing stars, as opposed to survey (i.e., compliance) stars or overall staffing stars. Our reasoning was that high long-stay hospitalization rates reflect some combination of more adverse events that could lead to hospitalization (e.g., new pressure ulcers) and lesser capacity to respond to new medical issues without resorting to the ER or hospital, as might be associated with lesser RN staffing.

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

Agreement of Model Dependent Variables –

TABLE 8. Agreement between MDS discharge status codes for long stays, and inpatient claims and death records

			According to claims and the Medicare enrollment record…							
	Total		STACH/CAH		IRF, LTCH, Psych Hospital, or Other IP		Alive but No IP Claim		Died	
	N	Col %	N	Row %	N	Row %	N	Row %	N	Row %
*All Long Stay MDS Discharges^*	*332,919*	*100%*	*213,772*	*64%*	*14,762*	*4%*	*70,756*	*21%*	*33,629*	*10%*
Acute hospital	241,857	73%	208,891	86%	6,381	3%	25,066	10%	1,519	1%
IRF, LTCH or Psych Hospital	9,957	3%	*	*	7,967	80%	*	*	*	*
Other setting	48,956	15%	3,851	8%	*	*	44,128	90%	*	*
Died	32,149	10%	*	*	*	*	*	*	31,545	98%

* Positive patient counts less than 11 must be blinded due to our CMS data use agreements.
^ A long stay discharge is defined as the patient having been in the facility for 100 days from admission to discharge.

Performance Measure Score- Correlation with SNF Industry Measures of Quality:

TABLE 9. Average Hospitalization Rate by Overall Five-Star Rating

Overall Five-Star Rating December 2014	Number of Facilities	PointRight® Pro Long Stay™ Adjusted Hospitalization Rate CY 2014
1	178	15.8%
2	403	14.3%
3	361	14.1%
4	512	13.5%
5	480	12.9%

*Person Correlation Coefficient = -.1510, significant at $p < .05$

TABLE 10. Average Hospitalization Rate by Survey Five-Star Rating

Survey Five-Star Rating December 2014	Number of Facilities	PointRight® Pro Long Stay™ Adjusted Hospitalization Rate CY 2014
1	408	14.7%
2	494	14.1%
3	460	13.5%
4	417	13.2%
5	155	13.8%

*Person Correlation Coefficient = -.0852 significant at $p < .05$

TABLE 11. Average Hospitalization Rate by Quality Five-Star Rating

Quality Five-Star Rating December 2014	Number of Facilities	PointRight® Pro Long Stay™ Adjusted Hospitalization Rate CY 2014
1	13	17.1%
2	60	15.9%
3	189	14.7%
4	667	14.5%
5	1005	13.1%

*Person Correlation Coefficient = -.1494 significant at $p < .05$

TABLE 12. Average Hospitalization Rate by Staffing Five-Star Rating

Staffing Five-Star Rating December 2014	Number of Facilities	PointRight® Pro Long Stay™ Adjusted Hospitalization Rate CY 2014
1	200	14.8%
2	366	14.0%
3	459	14.2%
4	810	13.6%
5	81	11.5%

*Person Correlation Coefficient = -.0950 significant at p < .05

TABLE 13. Average Hospitalization Rate by RN Staffing Five-Star Rating

RN Staffing Five-Star Rating December 2014	Number of Facilities	PointRight® Pro Long Stay™ Adjusted Hospitalization Rate CY 2014
1	104	15.0%
2	277	14.9%
3	571	14.1%
4	588	13.9%
5	376	12.3%

*Person Correlation Coefficient = -.1455 significant at p < .05

Figure 2. Scatter Plot of Pro Long Stay Adjusted Hospitalization Rates and Pro30 Adjusted Rehospitalization Rates

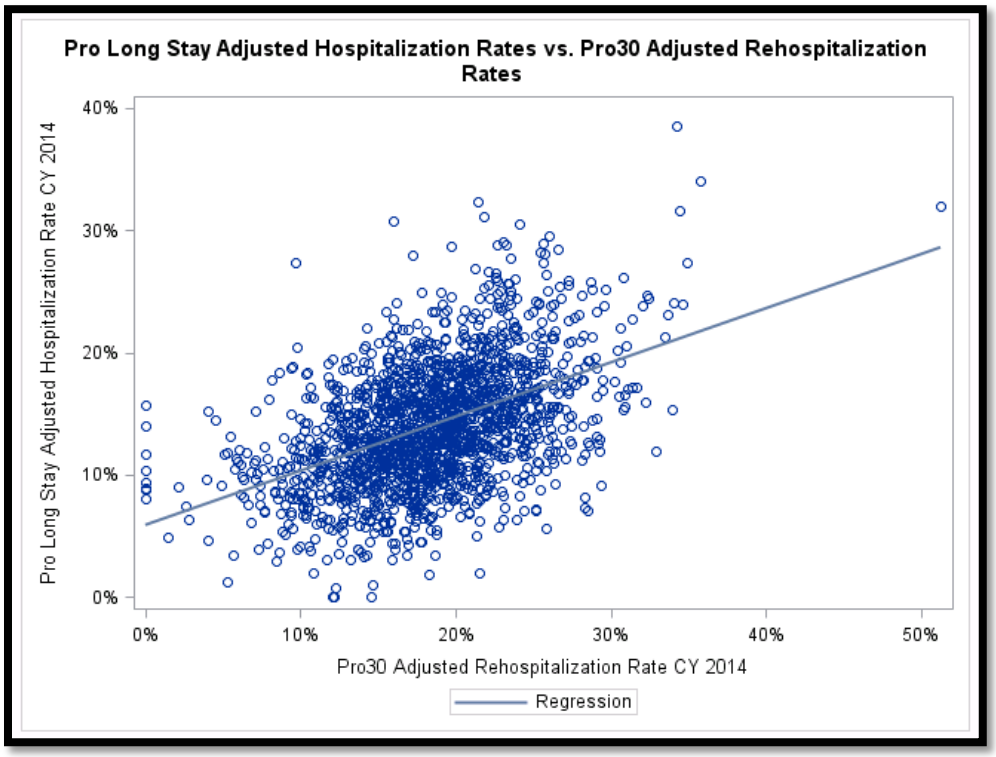


TABLE 14. Pearson Correlation Coefficients with CMS Long Stay Quality Measures

QM	Correlation Coefficient with Pro Long Stay Adjusted Hospitalization Rate	p-value
High-Risk Residents with Pressure Ulcers (Long Stay)	0.21	<.0001
Low-Risk Residents Who Lose Control of Their Bowels or Bladder (Long Stay)	-0.01	0.52
Residents Assessed and Appropriately Given the Pneumococcal Vaccine (Long Stay)	-0.05	0.0315
Residents Assessed and Appropriately Given the Seasonal Influenza Vaccine (Long Stay)	-0.03	0.1886
Residents Experiencing One or More Falls with Major Injury (Long Stay)	0.05	0.033
Residents Who Have Depressive Symptoms (Long Stay)	-0.06	0.0128
Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay)	0.06	0.0043
Residents Who Lose Too Much Weight (Long Stay)	0.12	<.0001
Long-Stay Residents Who Received An Antipsychotic Medication	0.04	0.0517
Residents who Self-Report Moderate to Severe Pain (Long Stay)	0.08	0.0003
Residents Who Were Physically Restrained (Long Stay)	0.08	0.0003
QM	**Correlation Coefficient with Pro Long Stay Adjusted Hospitalization Rate**	**p-value**
Residents Whose Need for Help with Activities of Daily Living Has Increased (Long Stay)	0.11	<.0001
Residents with a Urinary Tract Infection (Long Stay)	0.10	<.0001

*Highlighted cells represent correlation coefficients that were found statistically significant at p < .05

*Correlation coefficients in green font represent an improvement in the QM when Adjusted Long Stay Hospitalization rates decrease

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?*)

Critical Data Elements Agreement of Model Dependent Variables -

Table 8 presents a comparison of hospitalizations identified by claims and hospitalizations identified by the MDS, based on data from 241,857 long stay discharges for patients enrolled in Medicare Part A on discharge and existing in our extract of Medicare SNF patients in 2012, covering 15,091 SNFs, showed that 86% of hospitalizations of Medicare FFS patients identified by the MDS are confirmed by Medicare FFS claims; in the other direction, 98% (208,891 of 213,772) of acute inpatient claims found near an MDS discharge have an MDS discharge code of acute hospital. In other words, MDS discharge assessments appear to be overstating the rate of acute hospitalizations to a moderate degree. Independent

confirmation of the accuracy of the dependent variable for patients with other payers was not feasible as there is no central repository of hospitalization data for such residents. Overall, the relatively high level of agreement between MDS discharge coding and claims supports the validity of the measure.

Performance Measure Score- Correlation with SNF Industry Measures of Quality:
As hypothesized the PointRight Pro Long Stay Hospitalization Measure was correlated with other measures of quality. This supports using the Pro Long Stay Adjusted rates as a measure of a SNF's quality of care.

As Tables 9 through 13 above shows, higher star ratings are associated with lower adjusted long-stay hospitalization rates, and the relationship is strongest for the Quality and RN Staffing stars.

Pro Long Stay Adjusted Hospitalization rates had a statistically significant positive relationship with Pro30 Adjusted Rehospitalization rates at $p < .01$. The correlation coefficient was .47. In Figure 2 above find the two measures plotted against one another for CY 2014. We were expecting to see a strong correlation between the two measures, although it is still possible for a facility to perform well on of the measures and poorly on the other.

Table 14 shows the relationship between specific long-stay quality measures and the long-stay hospitalization measure. The correlation coefficients are statistically significant at $p < .05$. The negative correlation with depression may, given the modest p-value, be spurious, or if genuine may suggest lower hospitalization rates in facilities that do a better job of *recognizing* the depressive symptoms known to have a higher prevalence in SNFs in rigorous epidemiologic studies than they do in the published MDS-based CMS Quality Measures.

Validity of MDS Assessment
The measure developers did not specifically conduct analysis of the validity of the MDS items upon which the long-stay hospitalization measure is based. The overall reliability and validity of MDS 3.0 assessment was shown to be satisfactory prior to its adoption by CMS. RAND Corporation, the developer of MDS 3.0 as a contractor to CMS published the results of its reliability and validity testing in 2008 (Saliba and Buchanan, 2008). Details, and confirmatory studies, then appeared in peer-reviewed articles, some of which are referenced below.

Saliba , D., & Buchanan, J. (2008). Development & validation of a revised nursing facility assessment tool: MDS 3.0. RAND Health Corporation.

Saliba, D. & Buchanan, J. (2012). Making the investment count: Revision of the minimum data set for nursing facility's, MDS 3.0. J Am Med Dir Assoc. 13(7), 602-610.

Saliba, D., Buchanan, J., Eldelen, M.O., Streim, J., Ouslander, J., Berlowitz, D., & Chodosh, J. (2012). MDS 3.0: Brief interview for mental status. J Am Med Dir Assoc. 13(7), 611-617.

Saliba, D., DeFilippo, S., Edelen, M.O., Kroenke, K., Buchanan, J., & Streim, J. (2012) Testing the PHQ-9 interview and observational versions (PHQ-9 OV) for MDS 3.0. J Am Dir Assoc. 13(7), 618-625.

Saliba, D., Jones, M., Streim, J., Ouslander, J., Berlowitz, D., & Buchanan, J. (2012) Overview of significant changes in the minimum data set for nursing facilities version 3.0. J Am Dir Assoc. 13(7), 595-601.

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)

There are no exclusions from the denominator; all patients in the facility on the snapshot date who meet the long stay criterion on that date are included. However, the measure will not be reported for a SNF if the denominator population over the measure period's 4 snapshot dates is less than 30.

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*)

Table 15 below shows that the PointRight® Pro Long Stay™ Hospitalization Rates are steady across time with the exception of SNFs having small denominators. The 17 facilities with denominators less than 30 experienced changes in rates of greater than 50 percentage points. Ultimately excluding SNFs with denominators less than 30 resulted in excluding less than 1% of our sample.

TABLE 15. Change in Adjusted Rates from Quarter to Quarter by Denominator Size

Denominator Size	Number of Facilities	Variable	Mean	Standard Deviation	5th Percentile	95th Percentile
Denominator < 30	17	Adjusted Rate CY 2014	8.4%	10.3%	0.0%	36.1%
		Change in Adj Rates "July 2013 -June 2014" to "April 2013- Mar 2014"	4.8%	15.5%	-5.8%	57.4%
		Change in Adj Rates "Oct 2013 -Sept 2014" to "July 2013- June 2014"	-3.5%	7.9%	-24.8%	3.5%
		Change in Adj Rates "Jan 2014 -Dec 2014" to "Oct 2013- Sept 2014"	0.1%	7.4%	-18.9%	13.9%
30<= Denominator < 100	84	Adj Rate CY 2014	13.0%	7.2%	3.1%	24.7%
		Change in Adj Rates "July 2013 -June 2014" to "April 2013- Mar 2014"	0.5%	3.6%	-4.5%	6.9%
		Change in Adj Rates "Oct 2013 -Sept 2014" to "July 2013- June 2014"	0.1%	4.6%	-6.9%	7.8%
		Change in Adj Rates "Jan 2014 -Dec 2014" to "Oct 2013- Sept 2014"	0.4%	4.4%	-5.5%	8.0%
100<= Denominator < 300	1205	Adj Rate CY 2014	14.0%	5.1%	6.4%	22.9%
		Change in Adj Rates "July 2013 -June 2014" to "April 2013- Mar 2014"	0.3%	2.7%	-3.6%	4.3%
		Change in Adj Rates "Oct 2013 -Sept 2014" to "July 2013- June 2014"	0.2%	2.7%	-3.8%	4.5%
		Change in Adj Rates "Jan 2014 -Dec 2014" to "Oct 2013- Sept 2014"	0.3%	2.5%	-3.6%	4.5%
300<= Denominator < 450	514	Adj Rate CY 2014	14.5%	4.6%	7.4%	22.9%
		Change in Adj Rates "July 2013 -June 2014" to "April 2013- Mar 2014"	0.5%	2.2%	-2.8%	4.3%
		Change in Adj Rates "Oct 2013 -Sept 2014" to "July 2013- June 2014"	0.2%	2.0%	-2.9%	3.5%
		Change in Adj Rates "Jan 2014 -Dec 2014" to "Oct 2013- Sept 2014"	0.3%	1.9%	-2.7%	3.5%
450<= Denominator	229	Adj Rate CY 2014	13.8%	4.2%	6.6%	20.8%
		Change in Adj Rates "July 2013 -June 2014" to "April 2013- Mar 2014"	0.4%	2.1%	-2.8%	4.5%
		Change in Adj Rates "Oct 2013 -Sept 2014" to "July 2013- June 2014"	0.2%	1.5%	-2.2%	3.1%
		Change in Adj Rates "Jan 2014 -Dec 2014" to "Oct 2013- Sept 2014"	0.3%	1.5%	-2.2%	2.8%

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 results show that the average change of Adjusted Pro Long Stay Hospitalization Rates from quarter to quarter significantly drops once a facility has a denominator greater than or equal to 30. For this reason we felt rates are unstable for SNFs with denominators less than 30 and feel validated in excluding these SNFs. This exclusion only resulted in 17 (<1%) of facilities being excluded.

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 27 risk factors**
☐ **Stratification by ___risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (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 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 and not related to disparities*)

Our overall approach was to begin with reliable and rarely-missing patient-level SDS variables nominated by our clinical experts: Medicaid status (as a proxy for financial assets and income), black versus non-black, Hispanic/Latino versus non-Hispanic/Latino, and the interactions of Medicaid status and race. The significance of these variables in predicting hospitalization rates was tested in fixed-effects logistic regression models.

The risk adjustment model employed in the PointRight Pro Long Stay Hospitalization Rate utilizes four logistic regression models applied to four discrete subgroups of the denominator population to estimate risk of any hospitalization during a quarter at risk. This risk estimate is multiplied by a fixed factor to estimate the expected number of total hospitalization during the quarter.

Calculation of a patient's risk of any hospitalization (or equivalently, the risk of a first hospitalization) during a quarter at risk begins by assigning the patient to one of four subgroups of the denominator population based on the duration of the patient's current stay in the SNF as of the snapshot date. For each group the risk of one or more discharges from the SNF directly to an acute care hospital during the quarter is estimated by a logistic regression. The independent variables in each logistic regression model come from the patient's most recent MDS 3.0 assessment prior to the snapshot date that has the variable; most of the independent variables are common to the four models.

Our 4 logistic regression models will estimate a patient's risk of one or more hospitalizations. In order to estimate the patient's expected number of hospitalizations we multiply the risk estimate by a fixed factor. The factor was determined by modeling the relationships between the rate of first hospitalizations to the rate of all hospitalizations for a sample of 1,029 SNFs that had 100% complete reporting of their outcomes.

The decision to first segment the denominator population and then estimate logistic regression models was based on findings that the continuous length of stay at the beginning of a period at risk was the single most powerful predictor of hospitalization risk. (See Table 16). Estimating risk within strata defined by continuous length of stay gave greater predictive power than estimating a similar model without stratification, in part because risk factors operate somewhat differently within different LOS strata. The decision to estimate the risk of any hospitalization and then multiply it to get the expected rate of total hospitalizations was based on two considerations: (1) Models of binary dependent variables are much more widely known by clinical end-users of performance measures than models of dependent variables that can take multiple discrete values; this makes them more accessible and useful; (2) The relationship between SNFs' rates for first hospitalizations and their rates of all hospitalizations for their long-stay residents is extremely tight, with a linear equation accounting for 96% of the variance. Thus, modeling the risk of first hospitalizations is a rational approach to modeling of risk of all hospitalizations.

The selection of risk factors (independent variables) involved an iterative process. A panel of clinicians with extensive SNF experience recommended potential risk adjusters. These, and a full set of sociodemographic and contextual factors were tested for univariate relationships with hospitalizations. The variables with the strongest univariate correlations were then used to build multivariate models. The multivariate models (logistic regressions for each stratum of LOS) were reviewed by a larger panel of clinicians and potential users of the measure. Variables were rejected and replaced if their coefficients were opposite to their univariate correlation with the hospitalization, or if they were viewed as potentially under the control of the SNF – i.e., creating a risk of over-adjustment.

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

The initial stratification of the denominator population reflects the following analysis of the relationship of LOS on the snapshot date to the risk of a first hospitalization:

Table 16. Average Hospitalization Rates of Any (First) Hospitalization by Long Stay Groups

Long Stay Group	Continuous days form most recent admission to snapshot date	N Denominator Population July 1st 2013 to June 30th 2014 (% of total sample)	Observed Rate of First Long Stay Hospitalization (patient level) July 1st 2013 to June 30th 2014	N Denominator Population July 1st 2014 to June 30th 2015 (% of total sample)	Observed Rate of First Long Stay Hospitalization (patient level) July 1st 2014 to June 30th 2015
1	current LOS ≤ 100 days but cumulative days in SNF >100 days	67,156 (9.6%)	26.0%	56,242 (10.3%)	26.2%
2	100 days < LOS ≤ 181 days	98,992 (14.1%)	16.3%	74,236 (13.6%)	17.3%
3	181 days < LOS ≤ 364 days	144,789 (20.6%)	11.9%	109,794 (20.1%)	13.0%
4	LOS > 364 days	390,502 (55.7%)	7.3%	304,988 (55.9%)	7.8%

The models applicable to each of the subgroups of the denominator population are displayed in table 17. The table also shows the prevalence of the IV in population used to estimate the models, and indicates the model C-statistic. In the table cells are shaded red if the risk factor increases hospitalization risk and green if the risk factor decreases hospitalization risk.

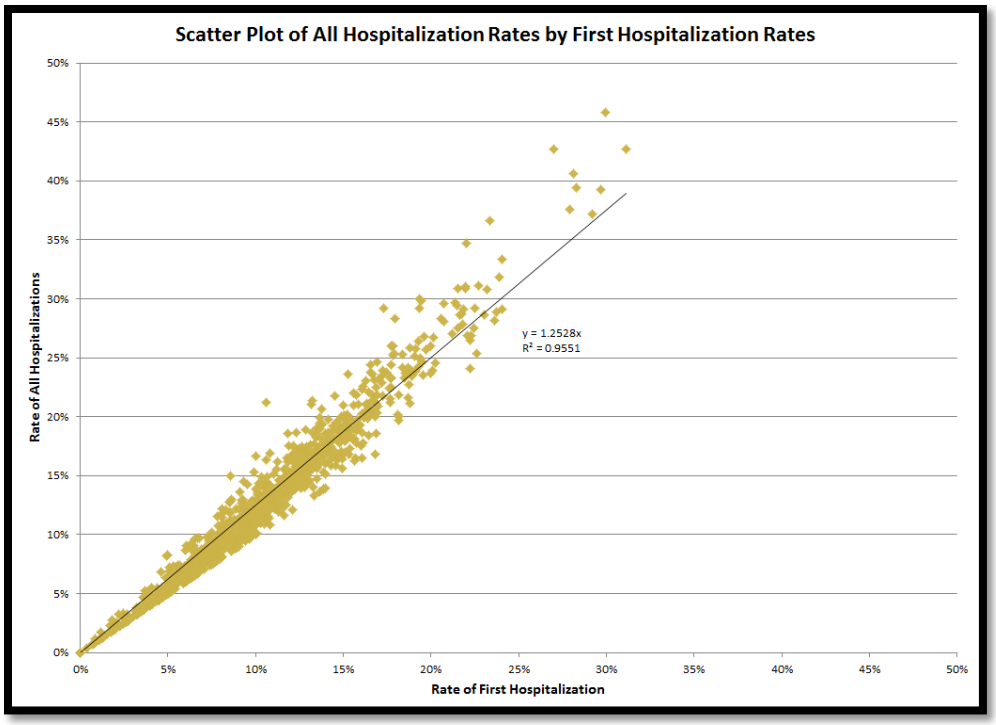
Table 17. PointRight® Pro Long Stay™ Hospitalization Measure Logistic Regression Models

Type of Independent Variable	Independent Variable	Long Stay Group 1 (current stay LOS ≤100 days but cumulative days in SNF >100) C-statistic = .64	Long Stay Group 1 Prevalence of Independent Variable	Long Stay Group 2 (100 days < LOS ≤181 days) C-statistic = .63	Long Stay Group 2 Prevalence of Independent Variable	Long Stay Group 3 (181 days < LOS ≤364 days) C-statistic = .62	Long Stay Group 3 Prevalence of Independent Variable	Long Stay Group 4 (LOS > 364 days) C-statistic = .63	Long Stay Group 4 Prevalence of Independent Variable
Constant	Intercept	-1.42	-	-1.97	-	-1.14	-	-1.01	-
Active Diagnoses	Anemia	0.12	37.3%	0.20	29.5%	0.20	28.1%	0.12	27.0%
Active Diagnoses	Asthma, COPD, or Chronic Lung Disease whether or not on oxygen	0.16	27.5%	0.19	27.7%	0.20	19.9%	0.17	16.3%
Active Diagnoses	Asthma, COPD, or Chronic Lung Disease on oxygen	X	X	X	X	0.28	6.3%	X	
Active Diagnoses	Diabetes on insulin	0.21	27.4%	0.30	21.7%	0.34	20.0%	0.30	16.5%
Active Diagnoses	Gastroesophageal Reflux Disease (GERD) or ulcer	X	X	X	X	X	X	0.12	23.0%
Active Diagnoses	Heart Failure	0.14	26.7%	0.20	20.3%	0.27	18.9%	0.18	16.4%
Active Diagnoses	Hypertension	X	X	X	X	X	X	0.22	69.4%
Active Diagnoses	Viral Hepatitis	0.26	0.9%	X	X	X	X	X	
Active Diagnoses	Neurogenic bladder	0.22	4.5%	X	X	0.43	2.6%	0.28	2.3%
Active Diagnoses	Renal failure, insufficiency, or ESRD	X	X	0.19	9.4%	0.17	7.7%	0.10	6.2%
Incontinence	Total bowel incontinence	X	X	0.20	27.9%	0.20	29.3%	0.13	35.3%
Demographics	Age 90 or over	-0.23	16.1%	-0.15	16.1%	-0.14	17.0%	-0.09	14.9%
Demographics	Male	0.14	37.7%	0.24	34.4%	0.20	31.8%	0.26	26.5%
Medications Received	Anticoagulant within 7 days prior to ARD	0.15	23.7%	X	X	X	X	0.08	17.4%
Medications Received	Antibiotics within 7 days prior to ARD	X	X	0.24	1.2%	X	X	X	
Stay History	Admitted from hospital (current stay)	0.38	96.4%	0.39	84.0%	0.27	82.5%	0.26	82.0%
Stay History	In this SNF 6 months before snapshot date (any stay)	-0.36	54.5%	X	X	X	X	X	
Stay History	In this SNF 12 months before snapshot date (any stay)	-0.33	31.7%	X	X	X	X	X	

Type of Independent Variable	Independent Variable	Long Stay Group 1 (current stay LOS ≤100 days but cumulative days in SNF >100) C-statistic = .64	Long Stay Group 1 Prevalence of Independent Variable	Long Stay Group 2 (100 days < LOS ≤181 days) C-statistic = .63	Long Stay Group 2 Prevalence of Independent Variable	Long Stay Group 3 (181 days < LOS ≤364 days) C-statistic = .62	Long Stay Group 3 Prevalence of Independent Variable	Long Stay Group 4 (LOS > 364 days) C-statistic = .63	Long Stay Group 4 Prevalence of Independent Variable
Stay History	Natural log of (Length of current stay minus 100) *prevalence is of LOS	X	X	-0.12	138.5	-0.29	265.9	-0.34	1160.8
Symptoms	Dyspnea on exertion	0.18	13.7%	X	X	X	X	0.19	5.0%
Skin	Surgical wound(s)	X		X	X	0.38	1.1%	0.38	0.5%
Hospice Status	Receiving hospice care	-1.51	4.6%	-1.23	5.3%	-1.24	5.3%	-1.05	4.7%
Recent Treatments	IV fluid or meds within 7 days before last MDS	0.18	5.3%	0.56	1.5%	0.36	0.9%	0.39	0.5%
Recent Treatments	Oxygen in 7 days before last MDS	X		0.38	13.9%	X	X	0.22	7.2%
Socioeconomic Status	Black resident on Medicaid	0.09	16.5%	-0.03	12.3%	X	X	X	
Socioeconomic Status	Black resident not on Medicaid	0.18	2.2%	0.07	3.9%	X	X	-0.12	2.1%
Socioeconomic Status	Non black resident not on Medicaid	-0.16	18.3%	0.01	32.2%	-0.08	25.3%	X	

The scatter plot below displays facility level linear regression model that resulted from the process described in section 2b4.3.

Figure 3. Scatter Plot of All Hospitalization Rates by First Hospitalization Rates



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)

Our overall approach was to begin with reliable and rarely-missing patient-level SDS variables nominated by our clinical experts: Medicaid status (as a proxy for financial assets and income), black versus non-black, Hispanic/Latino versus non-Hispanic/Latino, and the interactions of Medicaid status and race. The significance of these variables in predicting hospitalization rates was tested in fixed-effects logistic regression models. We reasoned that patient-level effects that were significant in models that included facility-specific constant terms probably reflected otherwise-unmeasured differences in baseline health status. Our final risk adjustment models were single-level logistic regression models in which the coefficients on the SDS variables were forced to be the same as in the fixed-effects model. Essentially this approach adjusts for the *within-facility* differences in long shay hospitalization rates associated with the SDS factor, but does not adjust for the *between-facility* differences in long stay hospitalization rates associated with the SDS factor. The within-facility effects are essentially those beyond those associated with facility quality differences. In all cases this made the effect of the SDS variables smaller than it would be in a single-level logistic regression that did not account for facility effects. We did not want to adjust away facility-level effects related to worse care at SNFs with large minority populations.

Finally, we note that the SDS factors we did include in risk adjustment have two interesting features that in our view argue for their inclusion in the model even though adding them did not meaningfully increase the models' c-statistics. The first is that while black race in general increases hospitalization risk, black long-stay SNF patients not on Medicaid actually have lower risk than non-black long-stay patients not on Medicaid. This may reflect that in the current cohort of elderly black patients having higher economic status is associated with overall better baseline health in ways not otherwise captured by our patient-level covariates. The second is that the effects of race and Medicaid status is different in patients with very long SNF stays than it is in those who have recently transitioned from post-acute care. Remaining in the SNF for a long continuous stay despite Medicaid status may reflect a high level of medical stability and/or a relatively low burden of care. Since facilities differ in their proportions of residents with long continuous stays the impact of the SDS variables is greater than would be evident from any one of our four LOS-specific risk adjustment models.

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*)

We compared our model coefficients to the mean coefficients from bootstrap analysis, expressed as actual values, standard deviation (S.D.) and percentage. Our sampling distribution consisted of 100 draws with replacement from our modeling data set.

All our covariates have less than 5% variation from the bootstrap mean, making the absolute value and/or the number of standard deviations clinically acceptable.

Table 18. Bootstrap Analysis Long Stay Group 1 current LOS <= 100 days but cumulative days in SNF >100 days

Variable Type	Independent Variable	Long Stay Group 1 Coefficient	Bootstrap Mean	S.D.	Difference	Difference in S.D.	Difference in %
Constant	Intercept	-1.432	-1.446	0.065	0.014	-0.009	-1.01%
Active Diagnoses	Anemia	0.123	0.121	0.018	0.001	0.001	1.02%
Active Diagnoses	Asthma, COPD, or Chronic Lung Disease whether or not on oxygen	0.161	0.163	0.021	-0.003	0.000	-1.61%
Active Diagnoses	Diabetes on insulin	0.207	0.209	0.021	-0.002	-0.001	-0.78%
Active Diagnoses	Heart Failure	0.145	0.146	0.018	0.000	0.003	-0.25%
Active Diagnoses	Viral Hepatitis	0.262	0.268	0.090	-0.006	-0.003	-2.41%
Active Diagnoses	Neurogenic bladder	0.222	0.229	0.044	-0.007	-0.003	-3.02%
Demographics	Age 90 or over	-0.229	-0.223	0.027	-0.006	0.001	2.58%
Demographics	Male	0.134	0.132	0.020	0.002	-0.001	1.28%
Medications Received	Anticoagulant within 7 days prior to ARD	0.149	0.148	0.020	0.001	0.000	0.83%
Stay History	Admitted from hospital (current stay)	0.391	0.404	0.066	-0.013	-0.012	-3.28%
Stay History	In this SNF 6 months before snapshot date (any stay)	-0.360	-0.360	0.022	0.000	0.000	-0.07%
Stay History	In this SNF 12 months before snapshot date (any stay)	-0.329	-0.326	0.024	-0.002	0.002	0.72%
Symptoms	Dyspnea on exertion	0.179	0.179	0.028	-0.001	-0.001	-0.31%
Hospice Status	Receiving hospice care	-1.514	-1.511	0.074	-0.003	-0.003	0.21%
Recent Treatments	IV fluid or meds within 7 days before last MDS	0.186	0.184	0.039	0.002	-0.001	0.92%

Table 19. Bootstrap Analysis Long Stay Group 2 100 days < LOS <= 181 days

Variable Type	Independent Variable	Long Stay Group 2 Coefficient	Bootstrap Mean	S.D.	Difference	Difference in S.D.	Difference in %
Constant	Intercept	-1.970	-1.980	0.036	0.010	0.004	-0.49%
Active Diagnoses	Anemia	0.203	0.202	0.020	0.002	-0.001	0.75%
Active Diagnoses	Asthma, COPD, or Chronic Lung Disease whether or not on oxygen	0.199	0.200	0.022	-0.001	0.000	-0.32%
Active Diagnoses	Diabetes on insulin	0.306	0.307	0.019	-0.001	0.001	-0.37%
Active Diagnoses	Heart Failure	0.200	0.203	0.021	-0.003	0.001	-1.54%

Active Diagnoses	Renal failure, insufficiency, or ESRD	0.191	0.192	0.029	-0.001	-0.001	-0.69%
Incontinence	Total bowel incontinence	0.200	0.204	0.019	-0.004	0.000	-1.91%
Demographics	Age 90 or over	-0.147	-0.148	0.025	0.001	0.001	-0.61%
Demographics	Male	0.247	0.247	0.017	0.000	0.001	-0.04%
Medications Received	Antibiotics within 7 days prior to ARD	0.226	0.235	0.068	-0.009	0.004	-3.88%
Stay History	Admitted from hospital (current stay)	0.388	0.387	0.028	0.001	0.000	0.31%
Stay History	Natural log of (Length of current stay minus 100) *prevalence is of LOS	-0.125	-0.123	0.009	-0.002	0.000	1.84%
Hospice Status	Receiving hospice care	-1.221	-1.226	0.054	0.005	0.005	-0.40%
Recent Treatments	IV fluid or meds within 7 days before last MDS	0.583	0.581	0.052	0.002	0.006	0.33%
Recent Treatments	Oxygen in 7 days before last MDS	0.388	0.387	0.022	0.001	0.003	0.28%

Table 20. Bootstrap Analysis Long Stay Group 3 181 days < LOS <= 364 days

Variable Type	Independent Variable	Long Stay Group 3 Coefficient	Bootstrap Mean	S.D.	Difference	Difference in S.D.	Difference in %
Constant	Intercept	-1.150	-1.163	0.120	0.013	0.006	-1.10%
Active Diagnoses	Anemia	0.202	0.200	0.016	0.002	0.002	0.80%
Active Diagnoses	Asthma, COPD, or Chronic Lung Disease whether or not on oxygen	0.195	0.196	0.023	-0.001	0.000	-0.36%
Active Diagnoses	Asthma, COPD, or Chronic Lung Disease on oxygen	0.275	0.282	0.036	-0.007	-0.001	-2.43%
Active Diagnoses	Diabetes on insulin	0.343	0.344	0.021	-0.001	-0.002	-0.34%
Active Diagnoses	Heart Failure	0.273	0.269	0.019	0.004	0.002	1.38%
Active Diagnoses	Neurogenic bladder	0.431	0.431	0.045	0.000	-0.002	0.09%
Active Diagnoses	Renal failure, insufficiency, or ESRD	0.174	0.173	0.026	0.001	0.003	0.86%
Incontinence	Total bowel incontinence	0.205	0.209	0.016	-0.004	0.002	-1.77%
Demographics	Age 90 or over	-0.136	-0.132	0.022	-0.004	0.002	2.58%
Demographics	Male	0.203	0.203	0.016	0.001	0.001	0.34%
Stay History	Admitted from hospital (current stay)	0.271	0.274	0.021	-0.003	0.004	-1.25%
Stay History	Natural log of (Length of current stay minus 100) *prevalence is of LOS	-0.284	-0.282	0.023	-0.002	0.001	0.67%
Skin	Surgical wound(s)	0.377	0.368	0.062	0.009	0.003	2.39%
Hospice Status	Receiving hospice care	-1.244	-1.248	0.057	0.004	0.001	-0.29%
Recent Treatments	IV fluid or meds within 7 days before last MDS	0.347	0.332	0.069	0.015	0.001	4.33%

Table 21. Bootstrap Analysis Long Stay Group 4 LOS > 364 days

Variable Type	Independent Variable	Long Stay Group 4 Coefficient	Bootstrap Mean	S.D.	Difference	Difference in S.D.	Difference in %
Constant	Intercept	-1.018	-1.018	0.061	-0.001	0.006	0.06%
Active Diagnoses	Anemia	0.118	0.120	0.013	-0.002	0.001	-1.79%
Active Diagnoses	Asthma, COPD, or Chronic Lung Disease whether or not on oxygen	0.167	0.169	0.017	-0.002	0.000	-1.20%
Active Diagnoses	Gastroesophageal Reflux Disease (GERD) or ulcer	0.117	0.117	0.014	0.001	0.000	0.43%
Active Diagnoses	Diabetes on insulin	0.296	0.295	0.015	0.001	0.000	0.20%
Active Diagnoses	Heart Failure	0.177	0.177	0.015	0.001	0.001	0.34%
Active Diagnoses	Hypertension	0.225	0.229	0.015	-0.004	0.001	-1.98%

Active Diagnoses	Neurogenic bladder	0.280	0.282	0.039	-0.002	-0.003	-0.56%
Active Diagnoses	Renal failure, insufficiency, or ESRD	0.104	0.106	0.023	-0.002	0.001	-1.90%
Incontinence	Total bowel incontinence	0.129	0.128	0.011	0.000	0.002	0.09%
Demographics	Age 90 or over	-0.093	-0.095	0.019	0.002	0.000	-2.53%
Demographics	Male	0.260	0.259	0.014	0.001	0.000	0.42%
Medications Received	Anticoagulant within 7 days prior to ARD	0.072	0.075	0.017	-0.003	0.000	-4.53%
Stay History	Admitted from hospital (current stay)	0.261	0.258	0.018	0.003	0.000	1.16%
Stay History	Natural log of (Length of current stay minus 100) *prevalence is of LOS	-0.338	-0.338	0.009	0.001	0.001	-0.16%
Symptoms	Dyspnea on exertion	0.194	0.192	0.026	0.002	0.000	0.89%
Skin	Surgical wound(s)	0.377	0.368	0.071	0.009	-0.003	2.31%
Hospice Status	Receiving hospice care	-1.032	-1.036	0.043	0.004	0.002	-0.37%
Recent Treatments	IV fluid or meds within 7 days before last MDS	0.405	0.406	0.061	-0.001	0.005	-0.22%
Recent Treatments	Oxygen in 7 days before last MDS	0.223	0.223	0.025	-0.001	-0.002	-0.42%

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*)**:**

Provided below are the c-statistics and r-squared values for the 5 models in the PointRight Pro Long Stay Hospitalization Measure.

Logistic Regression Model Long Stay Group 1 c-statistic = .64
Logistic Regression Model Long Stay Group 2, c-statistic = .63
Logistic Regression Model Long Stay Group 3, c-statistic = .62
Logistic Regression Model Long Stay Group 4, c-statistic = .63
Linear Regression Model Rate of all Hospitalizations, R-squared = .96

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

Table 22. Hosmer-Lemeshow Statistic Long Stay Group 1 (current LOS <= 100 days but cumulative days in SNF >100 days)

Partition for the Hosmer and Lemeshow Test			
Group	Total Patients at Risk	One or More Hospitalizations	
		Observed	Expected
1	6113	641	625
2	6655	1036	1082
3	6748	1309	1284
4	6712	1424	1450
5	6700	1606	1632
6	6682	1735	1801
7	6636	1984	1946
8	6641	2176	2135

9	6655	2438	2358
10	7447	3072	3107

Table 23. Homer-Lemeshow Statistic Long Stay Group 2 (100 days < LOS <= 181 days)

Partition for the Hosmer and Lemeshow Test			
Group	Total Patients at Risk	One or More Hospitalizations	
		Observed	Expected
1	9860	711	691
2	9853	864	1040
3	9845	1071	1174
4	9874	1313	1312
5	9860	1540	1433
6	9864	1651	1582
7	9861	1855	1741
8	9858	1965	1947
9	9857	2311	2243
10	9841	2839	2958

Table 24. Homer-Lemeshow Statistic Long Stay Group 3 (181 days < LOS <= 364 days)

Partition for the Hosmer and Lemeshow Test			
Group	Total Patients at Risk	One or More Hospitalizations	
		Observed	Expected
1	14404	721	745
2	14441	1064	1119
3	14410	1180	1255
4	14420	1322	1387
5	14410	1547	1514
6	14400	1733	1667
7	14410	1906	1837
8	14409	2109	2061
9	14413	2415	2387
10	14387	3167	3192

Table 25. Hosmer-Lemeshow Statistic Long Stay Group 4 (LOS > 364 days)

Partition for the Hosmer and Lemeshow Test			
Group	Total Patients at Risk	One or More Hospitalizations	
		Observed	Expected
1	38837	1150	1151
2	38834	1615	1679
3	38834	1973	1974
4	38833	2165	2229
5	38829	2444	2479
6	38843	2794	2750
7	38833	3128	3070

8	38837	3577	3471
9	38834	4054	4046
10	38825	5451	5504

Table 26. Linear Regression of All Hospitalizations by Decile of Expected Rates

Group	Observed Rate of Total Hospitalizations	Expected Rate of Total Hospitalizations
1	2.2%	2.5%
2	6.1%	6.6%
3	8.0%	8.7%
4	9.9%	10.4%
5	11.4%	11.9%
6	13.1%	13.4%
7	14.8%	15.0%
8	16.8%	16.9%
9	19.3%	19.4%
10	26.9%	25.5%

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

Figure 4. Calibration Plot Long Stay Group 1

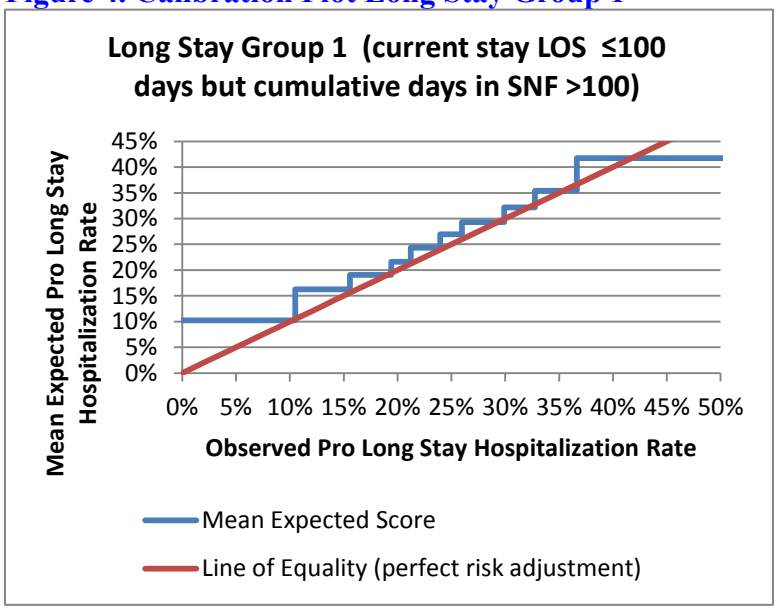


Figure 5. Calibration Plot Long Stay Group 2

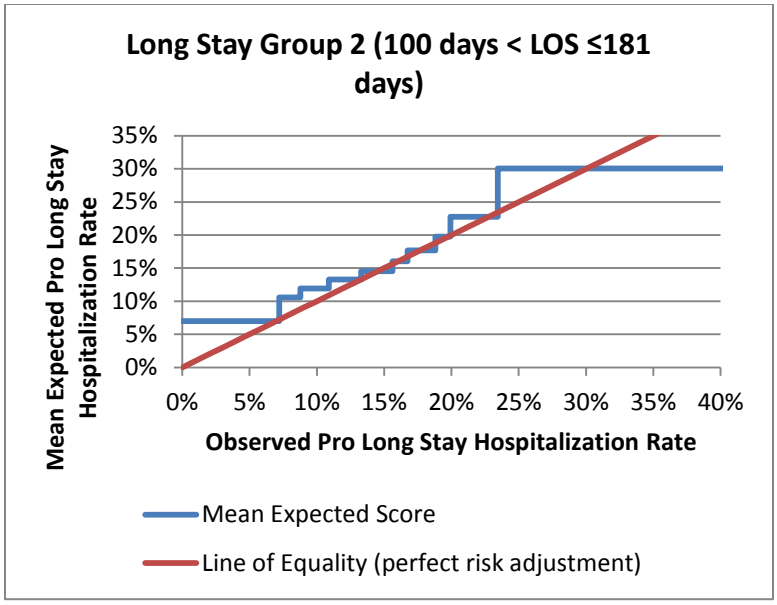


Figure 6. Calibration Plot Long Stay Group 3

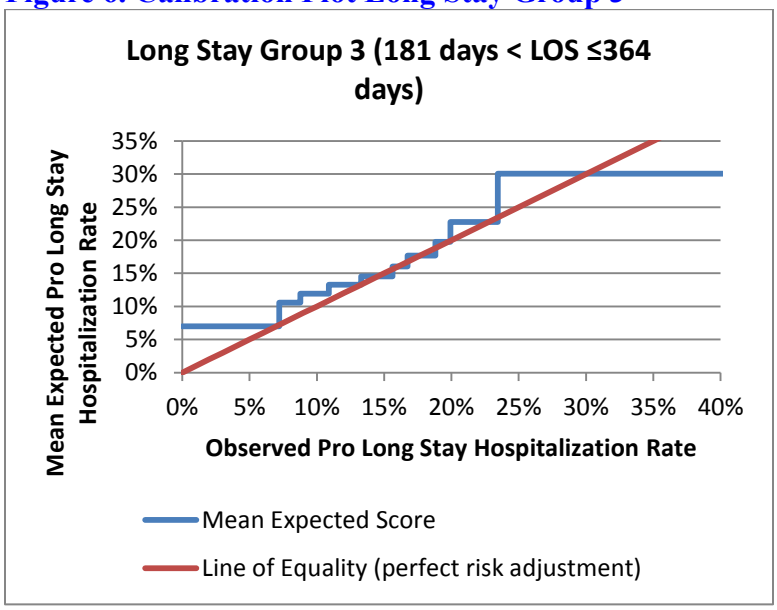


Figure 7. Calibration Plot Long Stay Group 4

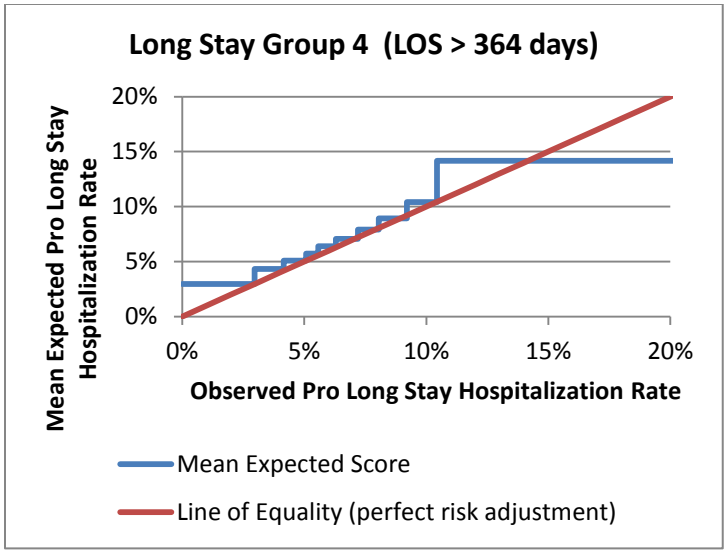
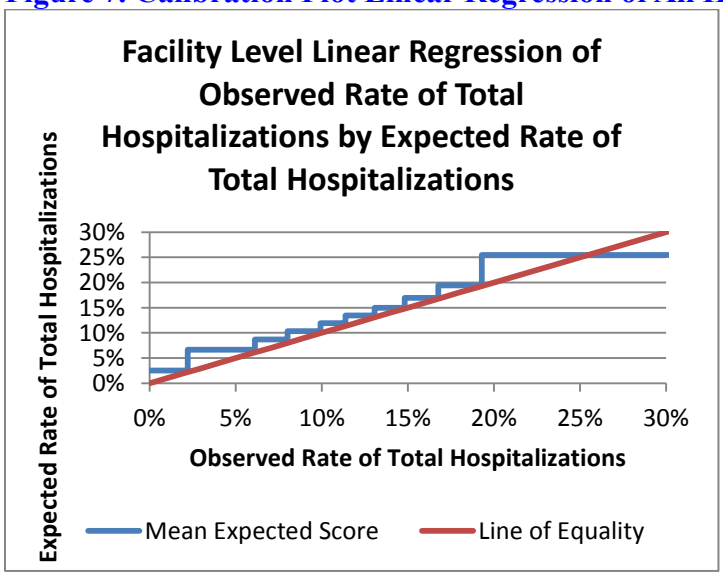


Figure 7. Calibration Plot Linear Regression of All Hospitalizations



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*)
N/A

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*)
N/A

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)*

The PointRight Pro Long Stay Hospitalization Measure will be reported to SNF providers quarterly. To determine what amount of change in risk adjusted rates will be considered meaningful from one quarter to the next we observed the distribution of changes amongst our testing sample. We started with risk adjusted rates covering the 12 month measurement period of October 1st, 2013 to September 30th, 2014 and observed the quarterly changes up until the 12 month measurement period of January 1st, 2014 to December 31st, 2014. We bucketed our sample into deciles of change in adjusted rates and calculated the average change for each bucket. We preformed the same analysis on subsets of our sample, where we divided the sample into 3 groups based on denominator size.

Large – Denominator > 400 patient quarters
Medium – 400 >= Denominator > 200 patient quarters
Small – 200 >= Denominator > 30 patient quarters

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*)

Table 27. Quarterly Changes in Risk Adjusted Rates (All Facilities N = 1,970)

	Average Quarterly Differences in Adjusted PointRight Pro Long Stay Hospitalization Rates		
Decile of Adjusted PointRight Pro Long Stay Hospitalization Rate Quarterly Changes	July 2013 to June 2014 minus April 2013 to Mar 2014	Oct 2013 to Sept 2014 minus July 2013 to June 2014	Jan 2014 to Dec 2014 minus Oct 2013 to Sept 2014
1	4.9%	4.1%	4.5%
2	2.4%	2.2%	2.5%
3	1.5%	1.4%	1.6%
4	0.9%	0.8%	1.0%
5	0.5%	0.3%	0.4%
6	0.0%	-0.2%	-0.1%
7	-0.6%	-0.7%	-0.6%
8	-1.2%	-1.3%	-1.2%
9	-2.0%	-2.1%	-2.0%
10	-3.7%	-3.9%	-3.7%

Table 28. Quarterly Changes in Risk Adjusted Rates (Large Facilities N=318)

	Average Quarterly Differences in Adjusted PointRight Pro Long Stay Hospitalization Rates		
Decile of Adjusted PointRight Pro Long Stay Hospitalization Rate Quarterly Changes	July 2013 to June 2014 minus April 2013 to Mar 2014	Oct 2013 to Sept 2014 minus July 2013 to June 2014	Jan 2014 to Dec 2014 minus Oct 2013 to Sept 2014
1	4.9%	3.3%	3.0%
2	2.4%	1.9%	1.9%
3	1.6%	1.2%	1.3%
4	1.0%	0.7%	0.9%
5	0.5%	0.4%	0.5%
6	0.1%	-0.1%	0.1%
7	-0.3%	-0.4%	-0.3%
8	-0.8%	-0.8%	-0.7%
9	-1.5%	-1.3%	-1.2%
10	-3.0%	-2.3%	-2.4%

Table 29. Quarterly Changes in Risk Adjusted Rates (Medium Facilities N=1,081)

	Average Quarterly Differences in Adjusted PointRight Pro Long Stay Hospitalization Rates		
Decile of Adjusted PointRight Pro Long Stay Hospitalization Rate Quarterly Changes	July 2013 to June 2014 minus April 2013 to Mar 2014	Oct 2013 to Sept 2014 minus July 2013 to June 2014	Jan 2014 to Dec 2014 minus Oct 2013 to Sept 2014
1	4.4%	3.7%	4.1%
2	2.2%	1.9%	2.4%
3	1.4%	1.2%	1.5%
4	0.9%	0.7%	0.9%
5	0.4%	0.2%	0.3%
6	0.0%	-0.3%	-0.1%
7	-0.5%	-0.8%	-0.6%
8	-1.2%	-1.3%	-1.2%
9	-1.9%	-2.0%	-1.9%
10	-3.4%	-3.6%	-3.4%

Table 30. Quarterly Changes in Risk Adjusted Rates (Medium Facilities N=571)

	Average Absolute Quarterly Differences in Adjusted PointRight Pro Long Stay Hospitalization Rates		
Decile of Adjusted PointRight Pro Long Stay Hospitalization Rate Quarterly Changes	July 2013 to June 2014 minus April 2013 to Mar 2014	Oct 2013 to Sept 2014 minus July 2013 to June 2014	Jan 2014 to Dec 2014 minus Oct 2013 to Sept 2014

1	5.8%	4.9%	5.5%
2	2.8%	2.9%	3.1%
3	1.8%	2.0%	2.0%
4	1.1%	1.2%	1.3%
5	0.5%	0.4%	0.5%
6	-0.2%	-0.2%	-0.2%
7	-0.9%	-0.9%	-0.9%
8	-1.6%	-1.7%	-1.7%
9	-2.5%	-2.5%	-2.6%
10	-4.4%	-5.0%	-4.7%

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 distribution of changes in adjusted rates, for our full sample, was similar across all 4 quarters, where for each quarter the average change for deciles 2 through 8 was less than +/- 3%. Deciles 1 and 10 had average changes greater than +/- 3.5%. The distribution of differences was larger for facilities with smaller denominators and this indicated that recommendations of clinically meaningful difference should be dependent upon facility size.

We made the following recommendations as we attempted to identify changes in adjusted rates that would move a facility several deciles in our sample's distribution.

Large Facilities - 4%
Medium Facilities - 3%
Small Facilities - 2%

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 criterion is directed 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).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of 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 demonstrating comparability of 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*)

On occasion facilities fail to submit MDS assessments adhering to the MDS submission schedule specified by regulation. This can result in a patient being included in a quarterly denominator population but not having a known outcome during the quarter following the snapshot date. The known outcome rate for the measure period is the sum of the counts of patients over the four quarterly denominator populations that have known outcomes, divided by the measure period denominator.

Knowing the outcome of a patient in the quarter at risk is vital the measure's accuracy. We've reviewed the known outcome rates across our sample to ensure that missing data is not a major factor. For this analysis we used our full sample of 2,361 PointRight clients, before excluding facilities for having known outcome rates less than 90%(N=2,096). This distribution is provided below in table 28.

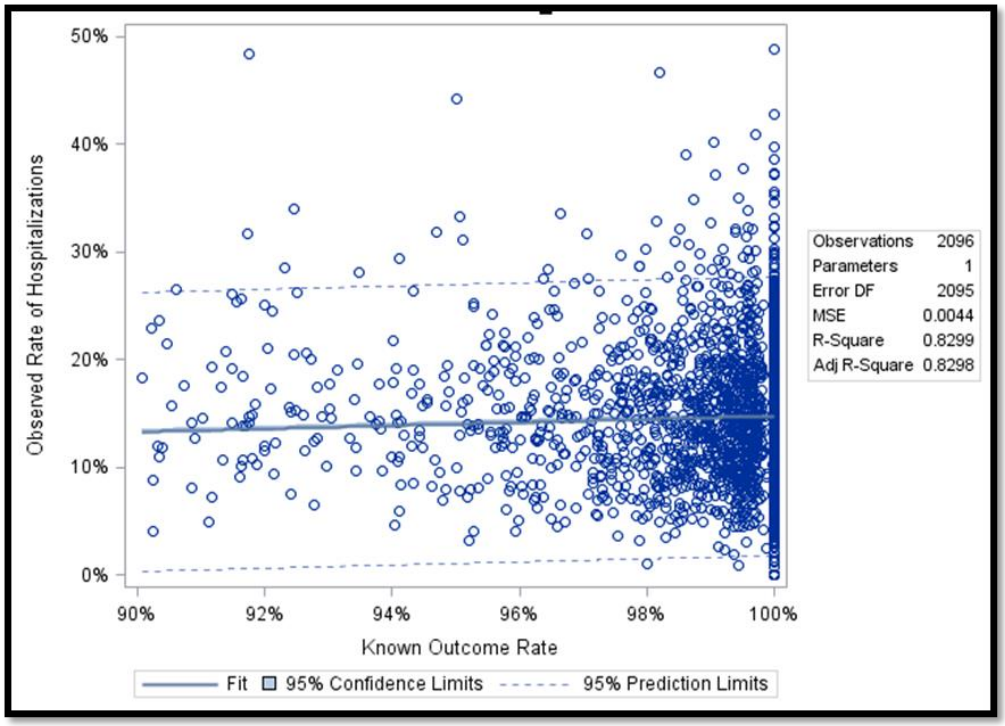
In addition to reviewing the missing data distribution across our sample, we also examined the relationship between the observed rate of hospitalizations and the known outcomes rate. For facilities with known outcomes rates between 100% and 90%, patient quarters at risk with unknown outcomes will contribute a .8 to the numerator. We impute .8 because nationally 80% of long stay patients with known outcomes are discharged to the hospital. If we are appropriately imputing the rate of hospitalization we would expect to see little to no correlation between known outcome rates and hospitalization rates. A scatter plot of the two rates is found in figure 8.

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*)

Table 31. Distribution of Known Outcome Rates before Exclusion

Quantile	
Quantile	Estimate
100% Max	100%
99%	100%
95%	100%
90%	100%
75% Q3	100%
50% Median	99%
25% Q1	98%
10%	89%
5%	82%
1%	73%
0% Min	70%

Figure 8. Scatter Plot of Known Outcome Rates and Observed Hospitalization Rates



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*)

The median known outcome rate in our full sample of PointRight facilities was 99%, leading us to conclude missing data was not an issue for the majority of facilities. We selected a known outcome rate of 90% to the be the minimum threshold. This threshold excluded 11% of our sample and was a good balance between the availability and utility of the measure.

In examining the relationship between known outcome rates and observed hospitalization rates we do see a very slight positive correlation. The Pearson correlation coefficient was .07 with a p-value of .0021 and the Spearman correlation coefficient was .03 with a p-value of .2340. These results lead us to believe that there is no significant bias for facilities with known outcomes rates greater than 90%.

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)

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 clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)

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.**

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.

Several of the decisions made in measure development were based on our objective of creating a measure that would be reliable and valid, readily computable based on data available to providers, and comprehensible and credible to providers so that they will adopt it in their quality improvement efforts. These decisions are described in several sections above. The following specific decisions are emphasized:

1)A one year rolling measure period was selected to ensure adequate denominators (>30) for virtually all SNFs that have long-stay patients at all.

2)A “snapshot” approach was adopted rather than a complex survival model to make the measure more comprehensible to users.

3)Simple logistic regressions were selected for risk adjustment because more complex models did not work better at the individual patient level, and we explicitly did not want to adjust for facility effects via a multilevel mode.

4)Variability in hospitalization explained by race and Medicaid status was divided into variance due to facility effects and variance

due to individual effects, with adjustment of outcomes only for the latter. The aim was to get facilities to take responsibility for the part of sociodemographic disparities more likely to be under their control than due to otherwise unmeasured differences in baseline health status.\

5)Observation of the rates of unknown outcomes in PointRight client population (see table A.7) motivated us to exclude facilities with more than 10% unknown outcomes, and to impute hospitalization for facilities with less than 100% but 90% or more known outcomes. Doing so implies that rates will be available for 91% of all facilities. Insisting a 95% rate of known outcomes would imply that rates would not be reported for 14% of all SNFs – a problem that would limit its value for quality improvement. With 10% unknown outcomes the maximum potential for overestimating the measure is 8% and the maximum potential for underestimating the measure is 2%. Clinically the problem with underestimation is erroneously identifying a poor-performing facility as a good-performing facility, and then either referring more patients there or otherwise supporting the facility's status quo. Given the distribution of the measure a 2% improvement will not bring a facility in the worst quartile of performance to better than median performance. Thus, the 10% threshold appears to be an acceptable compromise between availability and utility of the measure.

6)Observation of the very tight correlation the rate of first hospitalizations with the rate of all hospitalizations led us to base our risk adjustment on modeling of first hospitalizations and subsequent multiplication of the results by a conversion factor. This approach is much more comprehensible to many end users than use of models with non-binary dependent variables.

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)*.
Computation of the measure requires a license to use software for large-scale data management and calculation of risk estimates using logistic regression models. These are capabilities of all typical analytics software packages used by healthcare organizations (e.g., SAS, SPSS, Stata, and R). Healthcare organizations would thus not incur additional expense to implement the measure. Utilization of the measure specifications does not require a fee. However, there is a requirement that display, disclosure or publication of the measure include the measure's trademark (viz., PointRight Pro Long Stay Hospitalization Measure) and that it indicate that the measure specifications are copyrighted by PointRight.

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)
Public Reporting Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Quality Improvement (Internal to the specific organization)	

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

The measure is not currently in use.

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?*)

The American Health Care Association (AHCA) is currently implementing the PointRight® Long Stay Hospitalization measure on national MDS data, which it will then publish on its website for free public use, and also in its member data profiling and tracking tool, LTC Trend Tracker®. Once published, the measure developer and measure steward would like to see the measure adopted for regulatory and payment purposes, rather than a measure based on Fee-for-Service Medicare claims; with the increasing penetration of managed care for Medicare, Medicaid and dual eligible programs, and the significant proportion of private pay and commercial LTC insurance financing of long-term SNF care, a measure based on Medicare FFS claims alone could mischaracterize the performance of many SNFs.

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.*)

PointRight intends to provide the PointRight Pro Long Stay Hospitalization Measure to its customers beginning in the second half of 2016; AHCA intends to make the measure available to its members (and to other selected stakeholders) on its website in the second half of 2016. If the measure is endorsed by the NQF, AHCA and PointRight will advocate for its adoption by CMS as a publicly reported quality measure that contributes to CMS's evaluation of SNFs' clinical performance.

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**

Appendix 1b.2 table A.2 presents PointRight Pro Long Stay Hospitalization rates for years ending 2013q2, 2014q2 and 2015q2. PointRight compared the distributions of the PointRight Pro Long Stay Hospitalization Measure over three consecutive one-year measure periods, for a sample of 1,535 facilities that consistently submitted data to PointRight over all three periods and had known outcome rates of 90% or greater. The data show that long stay hospitalization rates have increased from 13.2% in the year ending 2013q2 to 14.0% in the year ending 2014q2, and 14.3% in the year ending 2015q2.

The geographic distribution of facilities used in the sample can be found in appendix 1b.2 on table A.3.

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.**

It is unclear why long stay hospitalization rates have increased, when short stay rehospitalization rates have consistently decreased over the same period. The short stay companion to the PointRight® Long Stay hospitalization measure shows a decrease from 17.5% in the year ending 2013q1 to 17.4% in the year ending 2014q1, to 17.3% in the year ending 2015q1 (according to AHCA analysis of PointRight® Pro 30™ rates calculated on national MDS data). However, this pattern speaks strongly to the need for a long stay hospitalization measure that nursing homes can incorporate into their quality assurance/performance improvement (QAPI) programs.

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.**
No unintended consequences have been identified or are anticipated to occur as a result of this measure.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

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?

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Currently there are no NQF-endorsed measures of hospitalizations for long stay nursing home patients.

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 competing measures at this time.

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:** PointRight_Pro_Long_StayTM_Appendix_Final.docx

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** American Health Care Association
Co.2 **Point of Contact:** Urvi, Patel, upatel@ahca.org, 202-842-4444-2858
Co.3 **Measure Developer if different from Measure Steward:** PointRight
Co.4 **Point of Contact:** Thomas, Martin, Thomas.Martin@pointright.com, 781-457-5900-5944

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The following is a list of members who served on the post acute care workgroup. This workgroup reviewed the measure specifications and provided PointRight advice on how to construct the measure for most of the different steps; numerator, denominator, risk adjustment and exclusions.

Barry Lazarus - HCR ManorCare
Holly Harmon - American Health Care Association
James Muller - American Health Care Association
Barbara Yody - Genesis
Tami Johnson - Kindred
Joanne Wisely - Genesis
Vincent Mor - Brown University
Bill Goulding - Aegis Therapies
Douglas Burr - Health Care Navigator

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2016
Ad.3 **Month and Year of most recent revision:** 11, 2015
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?** 01, 2017

Ad.6 **Copyright statement:** Copyright Notice:

Ad.7 **Disclaimers:** "The products and services provided by PointRight, including without limitation, feedback on data integrity or quality (clinical or otherwise), are not intended to give, and shall not be construed as, specific recommendations for the diagnosis or treatment of any medical condition or placement of the patient in any particular care environment. The products and services provided by PointRight are intended for the purpose of helping to promote a more accurate assessment, indicating where there may be errors or omissions requiring correction, and prompting more complete and accurate documentation of assessments performed by the Customer and its employees and contractors. The services provided by PointRight do not include any direct assessment of any resident or patient– either on-site or via electronic communication – nor the rendering of any opinion regarding the clinical diagnosis or treatment of any resident or patient. All patient care and activities resulting from decisions of the medical and social services community, are the sole responsibility of these groups for such care.

All medical practice management, patient care and placement decisions made in which the Services may be utilized, and the consequences thereof, will be exclusively the responsibility of the Customer, as well as physicians, other clinical practitioners with privileges at the Customers licensed facility(ies) and social services workers related to such patient care, transition and placement."

Ad.8 **Additional Information/Comments:** N/A

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 #: 2858
De.2. **Measure Title:** Discharge to Community
Co.1.1. **Measure Steward:** American Health Care Association
De.3. **Brief Description of Measure:** The Discharge to Community measure determines the percentage of all new admissions from a hospital who are discharged back to the community alive and remain out of any skilled nursing center for the next 30 days. The measure, referring to a rolling year of MDS entries, is calculated each quarter. The measure includes all new admissions to a SNF regardless of payor source.
1b.1. **Developer Rationale:** Improving national discharge to community rates directly aligns with the three aims of the National Quality Strategy led by the Agency for Healthcare Research and Quality (AHRQ) on behalf of the U.S. Department of Health and Human Services (HHS). Those three aims being: Better Care, Healthy People/Health Communities, and Affordable Care (AHRQ, 2011).

Better Care
Two central tenets to better care is making care more patient-centered and safe. The ability to measure and improve discharge to community rates furthers both.

With regards to patient-centered care, studies show the majority of nursing home residents prefer community discharge over remaining in post-acute and long-term care (Arling et al., 2010; Eckhert et al., 2004; McAuley & Blieszner, 1985). In spite of this, an estimated 10%-20% of nursing home residents capable of successfully residing in the community with appropriate rehabilitative services and support in place do not get discharged (Buttar et al., 2001; Chapin et al., 1998; Chapin et al., 2009; Mollica, 2006; Mor et al., 2007).

Remaining unnecessarily in institutionalized care is not benign, which is why improving discharge to community rates would lead to safer care.

Extended SNF stays increase a patient's risk and exposure to health care-related infections and serious illnesses, such as Methicillin-resistant Staphylococcus aureus (MRSA) and clostridium difficile (C. difficile). Approximately 2 million infections occur in nursing homes each year (Strausbaugh & Joseph, 2000). Nearly 10-30% of nursing home residents are colonized with C. difficile at any given time (Makris & Gelone, 2007). C. difficile is a particular challenge for nursing homes because they have limited or no areas for isolating infected patients. Also, C. difficile persists in the environment as spores that contaminate inanimate surfaces, such as bed rails, furniture, and toilets (Montoya & Mody, 2011). Therefore, reducing the number of days patients reside in SNFs by discharging them to the community has the potential of reducing unnecessary exposure to C. difficile and other infectious organisms.

Extended stays also have the potential to exacerbate the psychosocial toll of residing in a SNF. This psychosocial toll on wellbeing stems from patients potentially feeling socially isolated from friends and family and being required to sacrifice personal privacy and autonomy, while in SNF care (Calkins, 2009).

Healthy People/Healthy Communities
The utilization of SNFs and discharge to community rates is not uniform across the nation or between communities. Studies show facility and market factors can affect SNF to community discharge rates (Arling et al., 2011; Holup et al., 2015).

Non-uniform rates are also reflective of inconsistent community practices and engagement in the SNF discharge to community process. A reliable and well-reasoned discharge to community rate has the potential to help identify and spread effective and innovative care transition practices, as well as bring the issue to a head for those providers and communities lagging significantly behind.

Additionally, in order to have healthier communities, health inequalities and disparities have to be addressed. Measuring and improving discharge to community rates is another avenue towards this aim. A case in point highlighting this need is Leland et al.'s (2015) study, which found blacks were less likely than similar whites to be successfully discharged to the community from post-acute care for hip fractures. Overall, the study found only 57% of patients were successfully discharged to the community between 1999 and 2007.

Affordable Care

With SNF days being increasingly assigned to the most intensive rehabilitation case mix groups, which cost the most, reducing SNF patient days is more important now than it has ever been for Medicare's fiscal sustainability (MedPac, 2015).

Measuring and improving discharge to community rates is not only beneficial to Medicare. It could also lead to greater economies in service delivery and lower Medicaid spending growth by shifting resources from nursing homes to home and community-based services (Kaye et al., 2010; Kaye et al., 2009).

Better discharge to community rates can also help reduce patient health expenditures. Under Medicare Part A, patients are responsible for co-payments of $157.50 per day after their first 20 days of SNF care (CMS, 2014). Each unnecessary day in SNF care prevented means more money for patients to spend on other health care needs and lower chances of accruing medical debt.

Agency for Healthcare Research and Quality (AHRQ). (2011). "About the National Quality Strategy." Retrieved from http://www.ahrq.gov/workingforquality/about.htm Accessed 19 Jan 2016.

Arling G., Kane R. L., Cooke V., & Lewis T. (2010). "Targeting Residents for Transitions from Nursing Home to Community." Health Services Research. 45(2010): 691-711.

Arling G., Abrahamson,K. A. , Cooke V., et al. "Facility and Market Factors Affecting Transitions From Nursing Home to Community." Medical Care. 49 (2011): 790-796.

Buttar A., Blaum C., & Fries B. "Clinical Characteristics and Six-Month Outcomes of Nursing Home Residents with Low Activities of Daily Living Dependency." The Journals of Gerontology, Series A. 56 (2001): M292-M297.

Calkins M. P. "Evidence-Based Long Term Care Design." NeuroRehabilitation. 25(2009): 145–154.

Centers for Medicare & Medicaid Services (CMS). (2014). "Update to Medicare Deductible, Coinsurance, and Premium Rates for 2015." Medicare Learning Network Matters. MM8982. Retrieved from https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNMattersArticles/downloads/MM8982.pdf

Chapin R., Wilkinson D.S., Rachlin R., et al. "Going Home: Community Reentry of Light Care Nursing Facility Residents Age 65 and Over." Journal of Health Care Finance. 25 (1998): 35-48.

Chapin R., Baca B., Macmillan K., et al. "Residential Outcomes for Nursing Facility Applicants Who Have Been Diverted: Where are they 5 Years Later?" The Gerontologist. 49 (2009): 46-56.

Eckhert J. K., Morgan L. A., & Swamy N. "Preferences for Receipt of Care Among Community-Dwelling Adults." Journal of Aging & Social Policy. 16 (2004): 49–65.

Holup A. A., Gassoumis Z. D., Wilber K. H., and Hyer K. "Community Discharge of Nursing Home Residents: The Role of Facility Characteristics." Health Services Research. (2015).

Kaye H.S., LaPlante M.P., & Harrington C. "Do Noninstitutional Long-Term Care Services Reduce Medicaid Spending?" Health Affairs. 28 (2009): 262–272.

Kaye H.S., Harrington C., & LaPlante M.P. "Long-Term Care: Who Gets It, Who Provides It, Who Pays, and How Much?" Health Affairs. 29 (2010): 11–21.

Leland N. E., Gozalo P., Christrian T. J., et al. "An Examination of the First 30 Days After Patients are Discharged to the Community

From Hip Fracture Postacute Care." Medical Care. 53 (2015): 879-887.

Makris A, & Gelone S. "Clostridium difficile in the Long-Term Care Setting." Journal of American Directors Association. 8 (2007): 290–299.

McAuley W. J. & Blieszner R. "Selection of Long-Term Care Arrangements by Older Community Residents." The Gerontologist. 25 (1985): 188–193.

Medicare Payment Advisory Commission (MedPac) & Urban Institute. (2015) "The Need to Reform Medicare's Payments to Skilled Nursing Facilities is as Strong as Ever." Retrieved from http://www.medpac.gov/documents/reports/january-2015-medpac-and-urban-institute-report-the-need-to-reform-medicare-s-payments-to-skilled-nursing-facilities-is-as-strong-as-ever.pdf?sfvrsn=0

Mollica R. (2006). Transitions From Nursing Homes to Community Settings. In Berkman B, D'Ambruoso S (Eds.), Handbook of Social Work in Health and Aging (pp. 661-665). New York, NY: Oxford University Press.

Montoya A. & Mody,L. "Common Infections in Nursing Homes: A Review of Current Issues and Challenges." Aging Health. 7 (2011): 889–899.

Mor V., Zinn J., Gozalo P., et al. "Prospects for Transferring Nursing Home Residents to the Community." Health Affairs. 26 (2007): 1762-1771.

Strausbaugh L & Joseph C. "The Burden of Infection in Long-Term Care." Infection Control and Hospital Epidemiology. 21 (2000): 674–679.

S.4. **Numerator Statement:** The outcome measured is the number of new admissions from an acute care hospital discharge to community from a skilled nursing center. More specifically, the numerator is the number of stays discharged back to the community (i.e. private home, apartment, board/care, assisted living, or group home as indicated on the MDS discharge assessment form) from a skilled nursing center within 100 days of admission and remain out of any skilled nursing center for at least 30 days.
S.7. **Denominator Statement:** The denominator is the total number of all admissions from an acute hospital (MDS item A1800 "entered from"=03 (indicating an "acute care hospital") to a center over the previous 12 months, who did not have a prior stay in a nursing center for the prior 100 days (calculated by subtracting 100 from the admission date (MDS item A1900 "admission date"). Please note, the denominator only includes admissions from acute hospitals (MDS item A1800 "entered from"=03 (indicating an "acute care hospital") regardless of payor status.
S.10. **Denominator Exclusions:** The denominator has three exclusions (see below).

First, stays for patients less than 55 years of age are excluded from the measure.

Second, stays for which we do not where the patient entered from, or for which we do not observe the patient's discharge, are excluded from being counted in the denominator.

Third, stays with no available risk adjustment data (clinical and demographic characteristics listed in Section S.14) on any MDS assessment within 18 days of SNF admission are excluded from the measure.

Note, while not denominator exclusions, we also suppress the data for facilities that have fewer than 30 stays in the denominator, or for whom the percent of stays with a known outcome is less than 90%. The suppression of risk adjusted to community rates for facilities with fewer than 30 stays in the denominator is to improve the reliability of the measure, as detailed in the testing section (2b3). The suppression of rates for facilities for whom fewer than 90% of stays had a known outcome is done to improve the reliability of the measure and avoid perverse incentives about submitting MDS assessments for patients not discharged to the community.

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Electronic Clinical Data
S.26. **Level of Analysis:** Facility

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

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?** Not Applicable

New Measure -- Preliminary Analysis

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.

The developer provides the following evidence for this outcome measure:

- This measure determines the percentage of all new admissions from a hospital who are discharged back to the community alive and remain out of any skilled nursing center for the next 30 days. It includes all new admissions to a SNF regardless of payor source.
- The developer clearly states the rationale for the measure that improving national discharge to community rates directly aligns with NQS 3 aims of Better Care, Healthy People/Health Communities, and Affordable Care.
- The developer lists several studies from peer-reviewed journals that provide examples of clinical actions (identifying warning symptoms, medication reconciliation, follow-ups on labs and appointments, etc.) especially continuous communication between the patient/his family, staff at acute care hospitals and SNF staff lead to a patient- and family-centered improvement of quality of care.

Question for the Committee:

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

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 developer provides the following information:

- Studies show the majority of nursing home residents prefer community discharge over remaining in post-acute and long-term care but an estimated 10%-20% of nursing home residents capable of successfully residing in the community with appropriate rehabilitative services and support in place do not get discharged and remain unnecessarily in institutionalized care.
- Extended SNF stays increase a patient's risk and exposure to health care-related infections and serious illnesses, such as Methicillin-resistant Staphylococcus aureus (MRSA) and clostridium difficile (C. difficile). Approximately 2 million infections occur in nursing homes each year (Strausbaugh & Joseph, 2000). Nearly 10-30% of nursing home residents are colonized with C. difficile at any given time (Makris & Gelone, 2007).
- Extended stays also have the potential to exacerbate the psychosocial toll of residing in a SNF that stems from patients potentially feeling socially isolated from friends and family and being required to sacrifice personal privacy and autonomy, while in SNF care.
- The utilization of SNFs and discharge to community rates is not uniform across the nation or between communities. Non-uniform rates are reflective of inconsistent community practices and engagement in the SNF discharge to community process.
- Reducing SNF patient days will lead to Medicare and Medicaid spending, by shifting resources from nursing homes to home and community-based services and better discharge to community rates can also help reduce

total patient health expenditures.

Disparities

- A 2015 study (Leland et al.) found blacks were less likely than similar whites to be successfully discharged to the community from post-acute care for hip fractures. Overall, the study found only 57% of patients were successfully discharged to the community between 1999 and 2007.
- The developer used Sex and Age as risk factors and found the mean risk adjusted rates of discharge to community for the present measure were each essentially equal to the stay-weighted mean, i.e., 63.3%. Men were 63.4%, and women were 63.4%. For those 55-64, the mean rate was 63.47%; for those 65-74, 63.0%; for those 75-84, 63.48%; and for those 85 or greater, 63.48%. For race, the mean risk adjusted rates for Whites was 63.2%, Blacks was 64.0%, for Hispanics was 65.5%, for Hawaiians and Pacific Islanders was 67.2%, for American Indians was 63.7%, and for Asians was 64.1%.

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**

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

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

- This measure uses electronic clinical data collected via the Minimum Data Set (MDS).
- This measure calculates the rate of all new admissions from a hospital who are discharged back to the community alive and remain out of any skilled nursing center for the next 30 days.
- The Numerator is the number of new admissions from an acute care hospital discharge to community from a skilled nursing center. More specifically, the numerator is the number of stays discharged back to the community (i.e. private home, apartment, board/care, assisted living, or group home as indicated on the MDS discharge assessment form) from a skilled nursing center within 100 days of admission and remain out of any skilled nursing center for at least 30 days.
- The denominator is the total number of all admissions from an acute hospital (MDS item A1800 "entered from"=03 (indicating an "acute care hospital") to a center over the previous 12 months, who did not have a prior stay in a nursing center for the prior 100 days (calculated by subtracting 100 from the admission date (MDS item A1900 "admission date").
- The numerator and denominator populations are defined using MDS items; a list of applicable items is included in the submission.
- The measure is based on a rolling 12 month window of MDS entries, which is updated quarterly.
- The measure is risk-adjusted using a statistical risk model (see details below).

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

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

- The developers used two methods to test for reliability.
- Method 1: Replacement Bootstrapping
  - The developer compared outcomes before and after resampling with replacement bootstrapping.
  - The developer conducted a random resampling of the population with replacement to simulate a facility or two facilities of similar size independently drawing patients from the same underlying patient population and compared outcomes before and after resampling.
- Method 2: Performance Comparison between Quarters
  - The second method used to test reliability was a comparison of facility rates from one quarter to the next to assess the reliability of the measure across time.
  - The developer compared outcomes in Q3-2014 to Q4-2014 to analyze the actual stability in a facility's rate between consecutive measurement quarters.

Results of reliability testing

- Method 1: Replacement Bootstrapping
  - The developer found that if a SNF's patients were completely redrawn from the same underlying population (e.g. the same SNF a year in the future) or if two SNFs who each drew patients from the same underlying population were compared, 68% of the time they will remain ranked within ten percentiles of where they were before redrawing patients. In 96% of cases, they would shift less than thirty percentiles after random resampling.
  - The developer also notes that a SNF's rate would shift within five points 76% of the time and within ten points 95% of the time if it completely redrew its population.
- Method 2: Performance Comparison between Quarters
  - The developers found that between Q3-2014 and Q4-2014, 74% of facilities remained ranked within ten percentiles; 98% remained ranked within thirty percentiles.
  - Facility rates also stayed relatively stable from one quarter to the next with 85% changing less than five points and 98% changing less than ten points.

Guidance from the Reliability Algorithm [Algorithm guidance]

Question 1: Submitted specifications are precise, unambiguous, and complete.
Question 2: Empirical reliability testing was conducted using a adjusted rate bootstrapping.
Question 3: Empirical validity testing of patient-level data was conducted.
Question 4: Reliability testing was conducted with computed performance measure scores for each measure
Question 5: The replacement bootstrapping method was appropriate for assessing the proportion of variability due to real differences among measured entities.
Question 6: The bootstrapping procedure shows that if a SNF's patients were completely redrawn from the same underlying population (e.g. the same SNF a year in the future) or if two SNFs who each drew patients from the same underlying population were compared, 68% of the time they will remain ranked within ten percentiles of where they were before redrawing patients.

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?*

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

2b. Validity

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**

- This measure calculates the rate of all new admissions from a hospital who are discharged back to the community alive and remain out of any skilled nursing center for the next 30 days.
- As a rationale for measuring this health outcome, the developers that there are a number of interventions such as staffing levels, improved communication and appropriate treatment that SNFs can undertake to improve rates of discharge to the community.

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.

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 tested the validity of the measure two ways. First, the coding of discharges was validated against matched Part A claims data. Secondly, the developer performed construct validity testing by correlating risk adjusted discharge to community rates with certain other measures hypothesized to be driven by the same factors driving discharge to community rates.
- Method 1: Empirical MDS validation of the coding of discharges to the community
  - To validate the accuracy of the MDS coding of discharges to the community, the developer matched Part A and Medicare enrollment data with the MDS discharges for short stays (< 100 days) that admitted from the acute hospital.
  - The developer hypothesized that if there is a high proportion of the MDS discharges to the community that does not have a Part A claim, this is strong supporting evidence of the quality of the MDS discharge coding.
- Method 2: Construct validity correlating discharge to community rates with other quality metrics
  - The developer conducted construct validity testing comparing the discharge to community measure to other measures of SNF quality.
  - The developer hypothesized that facilities with higher discharge to community would correlate with Five Star ratings (particularly the quality measure (QM) component of Five Star), CMS nursing home compare short stay quality measures, and the facility's 30 day rehospitalization rate.
  - The developers used the survey, quality, and staffing components from the Five Star rating. Specifically, the developers grouped facilities by their Five Star rating and calculated the mean discharge to community measures for each grouping and calculated the Pearson's correlation coefficient. The developers hypothesized that facilities with higher Five Star ratings would have higher discharge to community measure scores.

Validity testing results:

- Method 1: Empirical MDS validation of the coding of discharges to the community
  - Of the 993,916 MDS discharges to community in the analytical dataset, the developer confirmed that 95% of MDS discharges do not have a Part Medicare claim in the inpatient or SNF settings on the MDS discharge date or the four days following.
  - The developer interprets this as indicative that the coding of discharges to the community is valid and reliable.
- Method 2: Construct validity correlating discharge to community rates with other quality metrics
  - The developers found a negative and statistically significant relationship between the discharge to community rate and the short stay rehospitalization rate (Pearson's correlation =-0.092, p<.0001).
    - The developer noted this negative coorelation was expected because higher scores of discharge to community measure are indicative of higher quality, whereas lower scores of the short stay rehospitalization rate are indicative higher quality.
  - The developer also found statistically significant correlations between the discharge to community rate and the CMS Nursing Home Compare Short Stay quality measures.

CMS NHC Short Stay Quality Measures	Discharge to Community Score	P value
Short stay pressure ulcers	-0.11249	<.0001
Short stay pain	0.01952	0.0374
Short stay antipsychotics	-0.23376	<.0001
Short stay influenza	0.16500	<.0001
Short stay pneumococcal	0.19441	<.0001

  - Additionally the developer found a positive correlation between the Five Star rating and the discharge to the community measure.
  - The developer interpreted these findings as supporting the construct validity of the discharge to community measure.

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:

- Patients in the following categories are excluded from the measure:
  - less than 55 years of age
  - stays for which it is unknown where the patient entered from, or for which the patient's discharge is not observed
  - stays with no available risk adjustment data (clinical and demographic characteristics listed in Section S.14) on any MDS assessment within 18 days of SNF admission
- The developer notes that all exclusions were determined an expert of long term care clinicians a priori to the development and testing of the measure.
- To determine the impact of exclusions, the developer calculated the discharge to community measure including residents of all ages, and compared that to the discharge to community measure where those younger than 55 are excluded, all other things being equal. Therefore, any differences in the two percentages are attributable to this age exclusion.
  - The measure is incalculable with the cases in the second and third exclusions so the developer was unable to test their effects on the measure.

- Finally, the developer tested the reliability suppression rules (fewer than 30 annual stays and tracking rates below 90%) using the same resampling approach we used in the main reliability testing (details in section 2a2).
- Excluding residents younger than 55 years old lead to excluding 5.5% of the total number of stays.
- The developer noted that testing shows that the sample loses reliability as the sample size decreases below 30 stays in the year.

Questions for the Committee:

- *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:

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

- A panel of clinicians with extensive SNF experience recommended the initial pool of potential risk adjusters. The developers then conducted a series of logistic regression analyses to estimate the effect of different patient characteristics ranging from medical diagnoses and clinical conditions to functional and demographic characteristics on the likelihood of successful discharge to community.
- In developing the risk adjusted model using logistic regression analysis, the developer used a backward elimination selection process with a significance level cutoff of $p<.10$ to determine which of the dozens of possible diagnoses, conditions, and characteristics available on the MDS contributed statistically to the likelihood of community discharge.
- This process resulted in 60 risk adjustment variables, which were encoded in 116 variables in the final risk model (including interaction terms, multilevel factor variables, etc.).

Conceptual analysis of the need for SDS adjustment

- Available and analyses SDS variables included age, gender, marital status, and race.
- In the CY 2014 MDS data set the majority (i.e. 61.8%) of the residents were female. The age of this group ranged from 55.0 to 116.6 years old, with a mean of 79.0 years old, and the standard deviation of 10.3 years.
- Marital status was missing for 3.5 % of this population. The two most prevalent marital status categories were widowed (i.e. 39.0%), followed by married (i.e. 34.7%).
- The majority (i.e. 80%) of the stays were white, 10% were black, and 4% were Hispanic

Empirical analysis of SDS factors:

- To develop the risk adjustment model, the developers state that they considered a selection of SDS risk adjustors including age, sex, language, and marital status.
- SDS variables were analyzed in the same way as all other variables. The developer did not do any separate analyses on these variables.
- Ultimately the developers included age, sex, and marital status.

Risk model diagnostics

- To assess the overall performance of their risk-adjustment model, the developers used several standard regression diagnostic techniques.
- The developers examined the examined the receiver operator characteristics (ROC) curve of the sensitivity and specificity of the model's prediction accuracy and found it to have an area under the ROC curve of 0.8147.
- The resulting model demonstrated a "pseudo" R-squared of 0.235 for the stay level. This is to say that about 23.5% of the variation in the outcome was completely explained by the regression model.
- The C-statistic was 0.820.
    - A c-statistic of 0.22 means that for 82% of all possible pairs of patients—one who died and one who lived—the model correctly assigned a higher probability to those who died. Generally, a c-statistic of at least 0.70 is considered acceptable.

- The facility level R-square is 55.9%. This means that 55.9% of the variance in rates was explained by the risk adjustment model.
- The developer performed a Hosmer-Lemeshow test for the goodness of fit of the logistic regression models. The test assesses whether or not the observed event rates match expected event rates in subgroups of the model population.
- The developers conclude that the risk model is an effective and well specified device for controlling for unwanted variance in discharge to community rates.

Results of the Hosmer and Lemeshow Test

Partition for the Hosmer and Lemeshow Test					
Group	**Total**	**outcome = 1**		**outcome = 0**	
		Observed	**Expected**	**Observed**	**Expected**
1	196533	29181	20135.08	167352	176397.9
2	196532	56209	56288.75	140323	140243.3
3	196532	81819	89116.23	114713	107415.8
4	196533	106943	115252.4	89590	81280.55
5	196533	129338	134681.6	67195	61851.38
6	196533	146128	148586.7	50405	47946.32
7	196536	159421	158795.9	37115	37740.12
8	196529	170083	166745.2	26446	29783.82
9	196533	178516	173576.9	18017	22956.09
10	196527	186699	181158.0	9828	15369.02

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	**DF**	**Pr > ChiSq**
11734.4060	8	<.0001

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.*
-

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

- To determine statistically significant and meaningful performance differences, the developer used the results from the reliability analyses.
- To compare between different facilities, the developer measured the percent of facilities whose discharge to communities change by various degrees when the facility's stays are resampled from the same underlying population of stays.

- After resampling a facility's stays (from the same underlying distribution), 75% of facilities stay within 5 percentage points of the original risk adjusted discharge to community rate, 90% stay within 8 percentage points of the original risk adjusted rate, and 95% stay within 10 percentage points.
- The developer states that this means they are 75% confident that a difference in rates of 5 percentage points is statistically significant, 90% confident that a difference in risk adjusted rates of 8 percentage points is statistically significant, and 95% confident that a difference in risk adjusted rates of 10 percentage points is statistically significant.

- To compare a single facility over time, the developer measured the percent of facilities whole risk adjusted rates change by various degrees from one quarter to the next by looking at the tails of this distribution – which measure directly the statistically significant differences in risk adjusted discharge to community rates at various levels of certainty.
  - 85% of facilities stay within 5 percentage points of the original risk adjusted discharge to community rate, 90% stay within 6 percentage points of the original risk adjusted rate, and 98% stay within 10 percentage points.
  - The developer states that they are 85% confident that a difference in rates of 5 percentage points is truly the case, 90% confident that a difference in risk adjusted rates of 6 percentage points is truly the case, and 98% confident that a difference in risk adjusted rates of 10 percentage points is truly the case.
- The developers note that their analysis concludes that an 8 percent difference in risk adjusted discharge to community rates between two different facilities is sufficient to confidently conclude one is better than the other; and a 6 percentage point difference in risk adjusted rates for a given facility between two quarters is sufficient to confidently conclude the facility's performance has improved/gotten worse.

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

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

Criterion 3. Feasibility

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.

The developer states:

- ALL measure data elements are in defined fields in electronic claims (e.g., clinical registry, nursing home MDS, home health OASIS) and routinely collected by and used by healthcare personnel during the provision of care.
- The measure does not present collection burden because it relies solely on data items from the MDS 3.0 that all facilities are already required to submit.
- Because the measure is collected and calculated on a quarterly basis but spanning a 12 month period, fewer facilities are expected to have missing rates due to small sample size. This prevents large fluctuations from one measurement period to the next.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*

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

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

Criterion 4: Usability and Use

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**
OR
Planned use in an accountability program? ☐ **Yes** ☐ **No**

Accountability program details

- Public Reporting: AHCA/NCAL's Research and Data Website, Measure Downloads
- Quality Improvement with Benchmarking (external benchmarking to multiple organizations): AHCA/NCAL's Research and Data Website, Measure Downloads; AHCA/NCAL LTC Trend TrackerSM
- Quality Improvement (Internal to the specific organization): AHCA/NCAL LTC Trend TrackerSM
- The developer provides links to the websites listed above.

Improvement results

- The developer reports that the national mean risk adjusted discharge to community score has increased from 57.9% in the 4th quarter of 2012, to 59.3% in the 4th quarter of 2013, to 60.0 in the 4th quarter of 2014. This is an increase of 3.6% since the end of 2014.

Potential harms; N/A

Feedback : N/A

***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**

Criterion 5: Related and Competing Measures

Related or competing measures

- This measure has no related or competing measures identified.

Harmonization

N/A

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** N/A

Measure Title: Discharge to Community

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: N/A

Date of Submission: 1/29/2016

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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: Discharge to Community

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

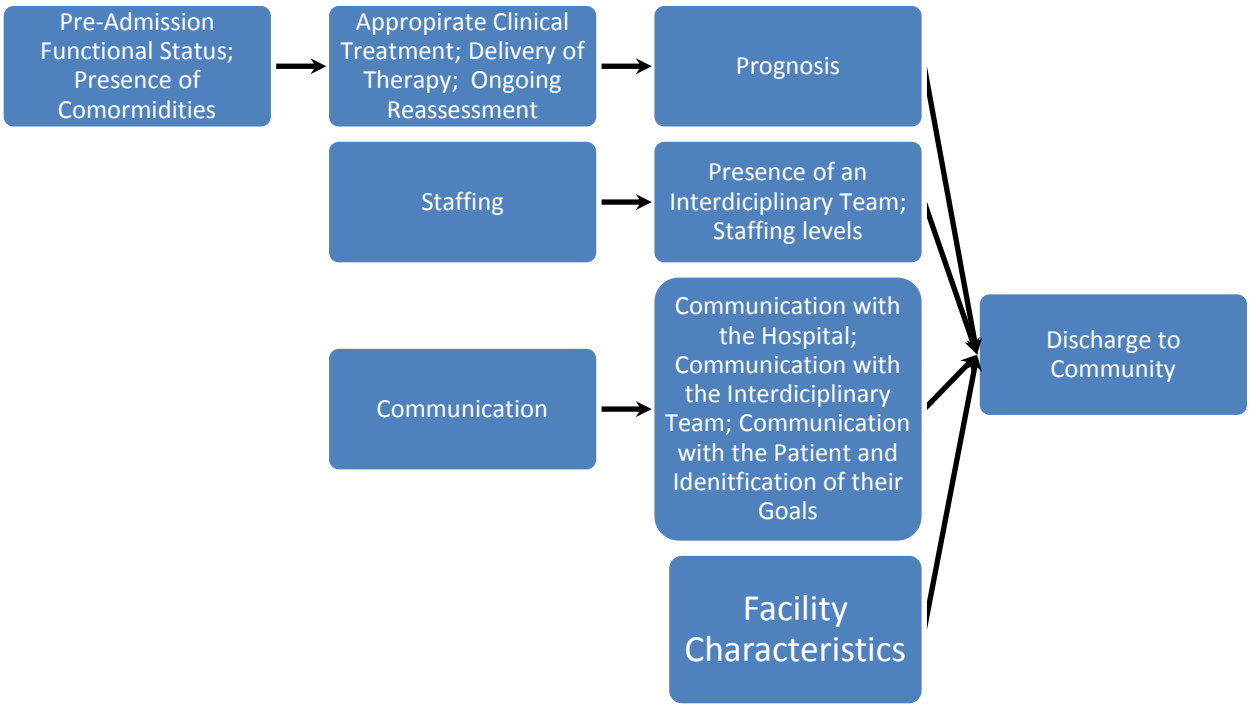
☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.



1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Approximately 20% of the Medicare beneficiaries are admitted to the hospital each year with almost 40% discharged to one of four post-acute care (PAC) settings for additional nursing or therapy services. In 2008, almost half (47%) of this group entered into a skilled nursing facility (Gage et al., 2011). Given their clinical complexity, these patients may be vulnerable and costly to the system. This presents important implications for

management of resources, as well as meeting the "triple aim" effort to improve quality of care, improve health, and reduce costs to the health care system (Ottenbacher et al., 2014; Chandra A, Dalton MA, Holmes J, 2013; Ouslander, J, Maslow K, 2012).Therefore, timely and safely discharge to the community ensures that these patients get the most optimal care while reducing costs to the system.

There is a cascade of events which leads to timely and safely discharge to community from a SNF. First, upon receiving acute care, patients must have an appropriate discharge to the SNF. Discharges from an acute care hospital a SNF, have numerous barriers that need to be overcome. In the most complex cases, in order for a discharge to occur, the staff and patients must overcome legal barriers, financing barriers, behavior management, availability of medical management at the next level of care, and finally inter-institution relations and transfers (Mackenzie et al., 2012). Mackenzie et al., 2012 identified these barriers among complex patients at the Denver Health an urban, public, safety-net health care system, however they are generally indicative of the need for coordination of care between the acute care hospital and skilled nursing facilities (ibid).

Moreover, these barriers speak to the need for continuous communication between three different tiers: the patient and family with the staff at the acute care hospitals, between the acute care hospital staff with the SNF staff, and ultimately once admitted to the skilled nursing facility, between the patients and the SNF staff as well as internally within SNF staff. This results in a patient- and family-centered improvement of quality of care (Ouslander, J, Maslow K, 2012). Elements of an effective care transition include; communication between clinicians on care plan, summary of patient history, patient preferences; preparing the patient and their caregiver on what to expect in the next setting and how to identify warning symptoms; medication reconciliation; and follow up on outstanding labs and appointments (Coleman, 2013).

Communicating with patients is very important. A 2010, study showed that the probability of discharge fell after the first 90 days in a SNF. As such, starting the conversations on community discharge before 90 days is important (Arling, et al., 2010). The same study also found that patients who have a support person and prefer to return to the community have a higher rate of discharge to community. As such, it is necessary to communicate with the patient and understand their preferences and needs (ibid).

Information about a patient's pre-admission level of physical and cognitive function and number/type of chronic comorbidities serve as an indicator for patient acuity factors, risk adjustors and expected changes in functional limitation. Expected functional improvements from a rehabilitation program, which ultimately influence the probability of discharge to the community, have been shown to be correlated to the patient's preadmission level of physical and cognitive function (Buntin et al., 2004; Cornette et al., 2005; Deutsch et al., 2005; Gage et al., 2005; Givens et al., 2008; Heruti et al., 2002; Kane et al., 2000; Kramer et al., 1997; Lieberman et al., 2006; Munin et al., 2005; Murtaugh et al., 2007; Stineman et al., 2003; Walsh et al., 2006). Therefore, SNF staff must use pre-admission information about chronic conditions and prognosis to tailor the level of care and the amount of resources allotted to meet the patient's rehabilitation goals.

The pre-existing number and type of chronic comorbidities will also impact the expected prognosis and subsequently the discharge to community. The Department of Health and Human Services (HHS) in 2010 recognized that as the number of chronic conditions in an individual increases, the risk of poor functional status and unnecessary hospitalizations also increases. In response, HHS has developed a strategic framework to address the prevalence of and the resource implications for addressing multiple chronic conditions (U.S. Department of Health, 2010). This is important because Wodchis et al. (2005) found different effects of different levels of rehabilitation therapy on time to discharge based on short term prognosis type (i.e. positive or

negative) (Wodchis et al., 2005). Specifically, they discovered that very high rehabilitation therapy intensity had little effect on time to discharge home in cases when the short-term prognosis was positive. On the other hand, residents with less optimistic prognosis benefited from additional rehabilitation therapy (ibid). Jette and colleagues found that in general higher intensity of therapy was associated with shorter lengths of stay (Jette et al., 2005). A recent study has found evidence that in particular for patients with hip fractures, having an additional hour of therapy every week is associated with a 3.1 percent increase in the likelihood for discharge to community (Jung et al., 2016).

Following admission into a skilled nursing facility, licensed medical professionals and therapists provide on-site appropriate clinical treatments and therapy. The therapist selects appropriate interventions based on comprehensive physical examination and collaboration with the interdisciplinary healthcare team (Health Care Association of New Jersey, 2012; National Institute for Health and Clinical Excellence, 2011). Literature suggests that the type of interventions selected should be specific to the patient's condition, the patient's functional impairments, and the patient's goals (Juhl et al., 2014; Silva et al., 2014; Iellmo et al., 2013; Lee et al., 2012). Post-admission, ongoing reassessment and the modification of the specific treatment approaches relative to the patient's response is critical to affect a change in mobility (Lohse et al., 2014; Veerbeek et al., 2014). All of these processes combined ensure that the patient is provided safe, timely, effective and efficient care so that the resident may be discharged to the community.

Staffing levels are also shown to have a positive influence on discharge to community. Specifically, for stays less than 60 days, increasing the number of nurses corresponded an increase in the probability of patients leaving the nursing home recovered or in a stabilized condition (Decker, 2006). Additionally, increasing nursing hours per resident day are show to have a positive correlation with discharge to community (Arling et. al, 2011).

Lastly, studies have also found that facility characteristics have an influence on discharge to community rates. Most notably, facility size, occupancy rates and proportion of Medicaid residents have a correlation with discharge to community. Holup and colleagues found that facilities in California and Florida had a great odds ratio of discharging patients home if they had a larger number of total beds and higher occupancy rates. The authors theorize that this occurs because these facilities are more likely to have a larger amount of resources, better systems to develop care plans, and more capacity for discharge planning. At the same time, the reason that higher Medicaid facilities had a lower odds ratio for discharge to community is likely due to a lack of overall resources (Holup et al., 2015). Residents who have Medicare as a primary payer source, have also been found to have higher rates of community discharge (Arling et. al., 2011). Generally, SNFs that have a higher number of admissions from post-acute care and have a higher number of patients who prefer to be discharged home are more likely to have higher discharge to community rates (Arling, et al., 2010).

Arling, G., Abrahamson, K.A., Cooke, V., Kane, R.L., & Lewis, T. "Facility and market factors affecting transitions from nursing home to community." *Medical Care*. 49.9(2011): 790-796.

Arling, G., Kane, R.L., Cooke, V., & Lewis, T. "Targeting residents for transitions from nursing home to community." *Health Services Research*. 45.3(2010): 691-711.

Buntin, M. B., & Zaslavsky, A. M. "Too much ado about two-part models and transformation? Comparing methods of modeling Medicare expenditures." *Journal of Health Economics* 23.3 (2004):525–542.

Chandra A., Dalton M.A., Holmes J. "Large increases in spending on postacute care in Medicare point to the potential for cost savings in these settings." *Health Affairs* (Millwood). 32.5 (2013):864-872.

Cornette, P., Swine, C., Malhomme, B., Gillet, J. B., Meert, P., & D'Hoore, W. "Early evaluation of the risk of functional decline following hospitalization of older patients: Development of a predictive tool." *European Journal of Public Health.* 16.2 (2005):203-208.

Decker, F.H. "Nursing staff and outcomes of nursing home stays." *Medical Care*. 44.9(2006): 812-821.

Deutsch, A., Granger, C. V., Fiedler, R. C., DeJong, G., Kane, R. L., & Ottenbacher, K. J. "Outcomes and reimbursement of inpatient rehabilitation facilities and subacute rehabilitation programs for Medicare beneficiaries with hip fracture." *Medical Care.* 43.9 (2005):892-901.

Gage, B., Smith, L., Morley, M., Drozd, E. M., Abbate, J., Coots, L.,… & Mallinson, T. "Post Acute Care Payment Reform Demonstration Report to Congress Supplement – Interim Report." 2011. RTI International. CMS contract No. HHSM-500-2005-00029I.

Gage, B., Smith, L., Morley, M., Drozd, E. M., Abbate, J., Coots, L.,… & Mallinson, T. "Post Acute Care Payment Reform Demonstration Report to Congress Supplement – Interim Report." 2011. RTI International. CMS contract No. HHSM-500-2005-00029I.

Givens, J. L., Sanft, T. B., & Marcantonio, E. R. "Functional recovery after hip fracture: The combined effects of depressive symptoms, cognitive impairment, and delirium." *Journal of the American Geriatric Society.* 56.6 (2008):1075-1079.

Health Care Association of New Jersey (HCANJ). "Fall management guideline." National Guideline Clearinghouse. 2012. 34 p.

Heruti, R. J., Lusky, A., Dankner, R., Ring, H., Dolgopiat, M., Barell, V.,… & Adunsky, A. 'Rehabilitation outcome of elderly patients after a first stroke: Effect of cognitive status at admission on the functional outcome." *Archives of Physical Medicine and Rehabilitation.* 83.6 (2002):742-749.

Holup, A.A., Gassoumis, Z.D., Wilber, K.H., & Hyer, K. "Community discharge of nursing home residents: The role of facility characteristics." *Health Serv Res.* 1-21.

Iellamo, F., Manzi, V., Caminiti, G., Vitale, C., Castagna, C., Massaro, M.,…& Volterrani, M. "Matched dose interval and continuous exercise training induce similar cardiorespiratory and metabolic adaptations in patients with heart failure." *International Journal of Cardiology*. 167.6 (2013): 2561-2565.

Jette, D.U., Warren, R.L., & Wirtalla, C. "The relation between therapy intensity and outcomes of rehabilitation in skilled nursing facilities." Arch Phys Med Rehabil. 86(2005):373-379.

Juhl, C., Christensen, R., Roos, E. M., Zhang, W., & Lund, H. "Impact of exercise type and dose on pain and disability in knee osteoarthritis: A systematic review and meta-regression analysis of randomized controlled trials." Arthritis and Rheumatology. 66.3 (2014):622-636.

Jung, H., Trivedi, A. N., Grabowski, D.C., & Mor, V. "Does more therapy in skilled nursing facilities lead to better outcomes in patients with hip fracture?" American Physical Therapy Association. 96(2016):81-89.

Kane, R. L., Chen, Q., Finch, M., Blewett, L., Burns, R., & Moskowitz, M. "The optimal outcomes of post-hospital care under Medicare." *Health Services Research.* 35.3 (2000): 615-661.

Kramer, A. M., Steiner, J. F., Schlenker, R. E., Eilertsen, T. B., Hrincevich, C. A., Tropea, D. A.,… & Eckhoff, D. G. "Outcomes and costs after hip fracture and stroke. A comparison of rehabilitation settings." *Journal of the American Medical Association.* 277.5 (1997): 396-404.

Lee, W. J., Cheng, Y. Y, Liu, C. Y., Peng, L. N., Liu, L. K., & Chen, L. K. "Dose-dependent effect of rehabilitation in functional recovery of older patients in the post-acute care unit." Archives of Gerontology and Geriatrics. 54.3(2012): e290-293.

Lieberman, D., Friger, M., & Lieberman, D. "Inpatient rehabilitation outcome after hip fracture surgery in elderly patients: A prospective cohort study of 946 patients." *Archives of Physical Medicine and Rehabilitation.* 87.2 (2006): 167-171.

MacKenzie T., Kukolja T., House R., Loehr A., Hirsh L., Boyle K, Sabel A., and Mehler P. " "A Discharge Panel At Denver Health, Focused On Complex Patients, May Have Influenced Decline In Length-Of-Stay. " *Health Affairs.* 31.8 (2012):1786-1795. doi:10.1377/hlthaff.2012.0515.

Munin, M. C., Seligman, K., Dew, M. A., Quear, T., Skidmore, E. R., Gruen, G.,… & Lenze, E. J. "Effect of rehabilitation site on functional recovery after hip fracture." *Archives of Physical Medicine and Rehabilitation.* 86(2005):367-372.

Murtaugh, C., Peng, T., Aykan, H., & Maduro, G. "Risk adjustment and public reporting on home health care." *Health Care Financing Review.* 28.3 (2007): 77-94.

Ottenbacher K., Karmarkar A., Graham J., Kuo YF., Deutsch A., Reistetter T., Al Snih S., Granger C. "Thirty-Day Hospital Readmission Following Discharge From Postacute Rehabilitation in Fee-for-Service Medicare Patients." *JAMA.* 311.6 (2014): 604-614. doi:10.1001/jama.2014.8.

Ouslander J., Maslow K. "Geriatrics and the Triple Aim: Defining Preventable Hospitalizations in the Long-Term Care Population." *JAGS.* 60.12 (2012):2313-2318.

Silva, N. L., Oliveira, R. B., Fleck, S. J., Leon, A. C, & Farinatti P. "Influence of strength training variables on strength gains in adults over 55 years-old: A meta-analysis of dose-response relationships." Journal of Science and Medicine in Sport. 17.3(2014):337-344.

Stineman, M. G., Ross, R. N., Granger, C. V., & Maislin, G. "Predicting the achievement of 6 grades of physical independence from data routinely collected at admission to rehabilitation." *Archives of Physical Medicine and Rehabilitation.* 84.11 (2003): 1647-1656.

U.S. Department of Health and Human Services. *Multiple Chronic Conditions—A Strategic Framework: Optimum Health and Quality of Life for Individuals with Multiple Chronic Conditions.* Washington, DC. December 2010.

Veerbeek, J. M., van Wegen, E., van Peppen, R., van der Wees, P. J., Hendriks, E., Rietbert, M., & Kwakkel, G. "What is the evidence for physical therapy poststroke? A systematic review and meta-analysis." PLOS One. 9.2(2014): 33 p.

Walsh, M. B., & Herbold, J. "Outcome after rehabilitation for total joint replacement at IRF and SNF: A case-controlled comparison." *American Journal of Physical Medicine and Rehabilitation.* 85.1 (2006): 1-5.

Wodchis W.,Teare G., Naglie G., Bronskill S., Gill S., Hillmer M., Anderson G., Rochon P., Fries B. "Skilled Nursing Facility Rehabilitation and Discharge to Home." *Arch Phys Med Rehabil*. 86 (2005): 442-448.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of 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
Discharge_to_Community_Evidence_Final.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)*
Improving national discharge to community rates directly aligns with the three aims of the National Quality Strategy led by the Agency for Healthcare Research and Quality (AHRQ) on behalf of the U.S. Department of Health and Human Services (HHS). Those three aims being: Better Care, Healthy People/Health Communities, and Affordable Care (AHRQ, 2011).

Better Care
Two central tenets to better care is making care more patient-centered and safe. The ability to measure and improve discharge to community rates furthers both.

With regards to patient-centered care, studies show the majority of nursing home residents prefer community discharge over remaining in post-acute and long-term care (Arling et al., 2010; Eckhert et al., 2004; McAuley & Blieszner, 1985). In spite of this, an estimated 10%-20% of nursing home residents capable of successfully residing in the community with appropriate rehabilitative services and support in place do not get discharged (Buttar et al., 2001; Chapin et al., 1998; Chapin et al., 2009; Mollica, 2006; Mor et al., 2007).

Remaining unnecessarily in institutionalized care is not benign, which is why improving discharge to community rates would lead to safer care.

Extended SNF stays increase a patient's risk and exposure to health care-related infections and serious illnesses, such as Methicillin-resistant Staphylococcus aureus (MRSA) and clostridium difficile (C. difficile). Approximately 2 million infections occur in nursing homes each year (Strausbaugh & Joseph, 2000). Nearly 10-30% of nursing home residents are colonized with C. difficile at any given time (Makris & Gelone, 2007). C. difficile is a particular challenge for nursing homes because they have limited or no areas for isolating infected patients. Also, C. difficile persists in the environment as spores that contaminate inanimate surfaces, such as bed rails, furniture, and toilets (Montoya & Mody, 2011). Therefore, reducing the number of days patients reside in SNFs by discharging them to the community has the potential of reducing unnecessary exposure to C. difficile and other infectious organisms.

Extended stays also have the potential to exacerbate the psychosocial toll of residing in a SNF. This psychosocial toll on wellbeing stems from patients potentially feeling socially isolated from friends and family and being required to sacrifice personal privacy and autonomy, while in SNF care (Calkins, 2009).

Healthy People/Healthy Communities
The utilization of SNFs and discharge to community rates is not uniform across the nation or between communities. Studies show facility and market factors can affect SNF to community discharge rates (Arling et al., 2011; Holup et al., 2015).

Non-uniform rates are also reflective of inconsistent community practices and engagement in the SNF discharge to community process. A reliable and well-reasoned discharge to community rate has the potential to help identify and spread effective and innovative care transition practices, as well as bring the issue to a head for those providers and communities lagging significantly behind.

Additionally, in order to have healthier communities, health inequalities and disparities have to be addressed. Measuring and improving discharge to community rates is another avenue towards this aim. A case in point highlighting this need is Leland et al.'s (2015) study, which found blacks were less likely than similar whites to be successfully discharged to the community from post-acute care for hip fractures. Overall, the study found only 57% of patients were successfully discharged to the community between 1999 and 2007.

Affordable Care

With SNF days being increasingly assigned to the most intensive rehabilitation case mix groups, which cost the most, reducing SNF patient days is more important now than it has ever been for Medicare's fiscal sustainability (MedPac, 2015).

Measuring and improving discharge to community rates is not only beneficial to Medicare. It could also lead to greater economies in service delivery and lower Medicaid spending growth by shifting resources from nursing homes to home and community-based services (Kaye et al., 2010; Kaye et al., 2009).

Better discharge to community rates can also help reduce patient health expenditures. Under Medicare Part A, patients are responsible for co-payments of $157.50 per day after their first 20 days of SNF care (CMS, 2014). Each unnecessary day in SNF care prevented means more money for patients to spend on other health care needs and lower chances of accruing medical debt.

Agency for Healthcare Research and Quality (AHRQ). (2011). "About the National Quality Strategy." Retrieved from http://www.ahrq.gov/workingforquality/about.htm Accessed 19 Jan 2016.

Arling G., Kane R. L., Cooke V., & Lewis T. (2010). "Targeting Residents for Transitions from Nursing Home to Community." Health Services Research. 45(2010): 691-711.

Arling G., Abrahamson,K. A. , Cooke V., et al. "Facility and Market Factors Affecting Transitions From Nursing Home to Community." Medical Care. 49 (2011): 790-796.

Buttar A., Blaum C., & Fries B. "Clinical Characteristics and Six-Month Outcomes of Nursing Home Residents with Low Activities of Daily Living Dependency." The Journals of Gerontology, Series A. 56 (2001): M292-M297.

Calkins M. P. "Evidence-Based Long Term Care Design." NeuroRehabilitation. 25(2009): 145–154.

Centers for Medicare & Medicaid Services (CMS). (2014). "Update to Medicare Deductible, Coinsurance, and Premium Rates for 2015." Medicare Learning Network Matters. MM8982. Retrieved from https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNMattersArticles/downloads/MM8982.pdf

Chapin R., Wilkinson D.S., Rachlin R., et al. "Going Home: Community Reentry of Light Care Nursing Facility Residents Age 65 and Over." Journal of Health Care Finance. 25 (1998): 35-48.

Chapin R., Baca B., Macmillan K., et al. "Residential Outcomes for Nursing Facility Applicants Who Have Been Diverted: Where are they 5 Years Later?" The Gerontologist. 49 (2009): 46-56.

Eckhert J. K., Morgan L. A., & Swamy N. "Preferences for Receipt of Care Among Community-Dwelling Adults." Journal of Aging & Social Policy. 16 (2004): 49–65.

Holup A. A., Gassoumis Z. D., Wilber K. H., and Hyer K. "Community Discharge of Nursing Home Residents: The Role of Facility Characteristics." Health Services Research. (2015).

Kaye H.S., LaPlante M.P., & Harrington C. "Do Noninstitutional Long-Term Care Services Reduce Medicaid Spending?" Health Affairs. 28 (2009): 262–272.

Kaye H.S., Harrington C., & LaPlante M.P. "Long-Term Care: Who Gets It, Who Provides It, Who Pays, and How Much?" Health Affairs. 29 (2010): 11–21.

Leland N. E., Gozalo P., Christrian T. J., et al. "An Examination of the First 30 Days After Patients are Discharged to the Community From Hip Fracture Postacute Care." Medical Care. 53 (2015): 879-887.

Makris A, & Gelone S. “Clostridium difficile in the Long-Term Care Setting.” Journal of American Directors Association. 8 (2007): 290–299.

McAuley W. J. & Blieszner R. “Selection of Long-Term Care Arrangements by Older Community Residents.” The Gerontologist. 25 (1985): 188–193.

Medicare Payment Advisory Commission (MedPac) & Urban Institute. (2015) “The Need to Reform Medicare’s Payments to Skilled Nursing Facilities is as Strong as Ever.” Retrieved from http://www.medpac.gov/documents/reports/january-2015-medpac-and-urban-institute-report-the-need-to-reform-medicare-s-payments-to-skilled-nursing-facilities-is-as-strong-as-ever.pdf?sfvrsn=0

Mollica R. (2006). Transitions From Nursing Homes to Community Settings. In Berkman B, D’Ambruoso S (Eds.), Handbook of Social Work in Health and Aging (pp. 661-665). New York, NY: Oxford University Press.

Montoya A. & Mody,L. “Common Infections in Nursing Homes: A Review of Current Issues and Challenges.” Aging Health. 7 (2011): 889–899.

Mor V., Zinn J., Gozalo P., et al. “Prospects for Transferring Nursing Home Residents to the Community.” Health Affairs. 26 (2007): 1762-1771.

Strausbaugh L & Joseph C. “The Burden of Infection in Long-Term Care.” Infection Control and Hospital Epidemiology. 21 (2000): 674–679.

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.*

Calendar Year 2014 data was utilized to derive the performance score, a description of this data sample is presented in the measure testing form as the same sample was utilized. Table 1b.2.1 presents the discharge to community cutoff scores for each decile, as well as the descriptive statics for 1,913,510 stays among the 11,939 facilities. Note this applies the facility suppressions of denominators smaller than 30 stays and facilities for whom fewer than 90% of stays had a known outcome. The mean discharge to community rate was 59.60%, with a SD of 10.70%; and the median discharge to community rate for this population was 61.40%. The minimum risk adjusted discharge to community rate after applying the exclusions is 0.00%, the 10th percentile is 45.8%, the 20th is 51.9%, 30th is 55.8%, 40th is 58.9%, 50th (median) is 61.4%, 60th is 63.6%, 70th is 65.8%, 80th is 68.1%, 90th is 71.2% and maximum is 100%.

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.

Not applicable.

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.*

To provide information relevant to our measure, we produced the following statistics utilizing the same data sample used for measure testing (see measure testing form for a description of the sample). Note the distributions differ from the performance scores in Section 1b.2 because these are weighted by number of stays.

Because sex and age were adjustors in this NQF application, the mean risk adjusted rates of discharge to community for the present measure were each essentially equal to the stay-weighted mean, i.e., 63.3%. Men were 63.4%, and women were 63.4%. For those 55-64, the mean rate was 63.47%; for those 65-74, 63.0%; for those 75-84, 63.48%; and for those 85 or greater, 63.48%. For race, the mean risk adjusted rates for Whites was 63.2%, Blacks was 64.0%, for Hispanics was 65.5%, for Hawaiians and Pacific Islanders was 67.2%, for American Indians was 63.7%, and for Asians was 64.1%.

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.

Not Applicable

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, High resource use, Patient/societal consequences of poor quality

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.**

Skilled nursing facilities (SNFs) serve a pivotal role in providing post-acute care to patients requiring rehabilitation therapy or daily skilled nursing services. The Medicare SNF Part A benefit is designed principally for the purpose of either providing therapy to individuals who were hospitalized or completing their nursing care in a less intensive setting. Skilled nursing facilities' ultimate goal is to provide optimal care to improve patients' functional abilities so that they may safely and timely return to their prior living situation in the community (Wodchis et al., 2005).

With an estimated 10%-20% of nursing home residents capable of successfully residing in the community with appropriate rehabilitative services and support in place but remaining as inpatients, SNFs are falling short of their goal (Buttar et al., 2001; Chapin et al., 1998; Chapin et al., 2009; Mollica, 2006; Mor et al., 2007).

The consequence of failing to appropriately discharge nursing home residents is vast and sizable. Medicare data on SNF utilization, charges, and payments show no substantial improvements from 2008-2012. During this time frame, Medicare covered every year over 2.5 million SNF admissions, which translated to over 68 million patient days per year. The per day covered SNF charges increased by 24% from $505 in 2008 to $627 in 2012. Total Medicare payments to SNFs increased by 15% from $301 billion in 2008 to $345 billion in 2015 (CMS, 2013).

Improving safe discharge to community rates from SNFs has the potential to lead to meaningful reductions in unnecessary SNF patient days, charges, and payments. Currently, there is no standard assessment tool to measure discharge to community among SNFs. Without a proper means to measure discharge to community, SNFs are challenged in knowing if changes they enact to improve care transitions are in fact making an impact. Knowing whether or not a change in practice is an improvement is a fundamental aspect of health care quality improvement (Langley, 2009).

The passage of the Improving Post-Acute Care Transformation Act (IMPACT Act) in October 2014 illustrates this very need for better measurement of care transitions from SNFs and other post-acute providers. More specifically, the IMPACT Act calls for the development of measures pertaining to resource use, hospitalization, and discharge to the community (CMS, 2014).

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

Buttar A., Blaum C., & Fries B. "Clinical Characteristics and Six-Month Outcomes of Nursing Home Residents with Low Activities of Daily Living Dependency." The Journals of Gerontology, Series A. 56 (2001): M292-M297.

Centers for Medicare & Medicaid Services (CMS). (2013). "Medicare & Medicaid Statistical Supplement, 2013 Edition." Office of Enterprise Data and Analytics. Retrieved from https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MedicareMedicaidStatSupp/2013.html Accessed on 18 Jan 2016.

Centers for Medicare & Medicaid Services (CMS). (2014). "IMPACT Act of 2014 Data Standardization & Cross Setting Measure" Retrieved from https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/IMPACT-Act-of-2014/IMPACT-Act-of-2014-Data-Standardization-and-Cross-Setting-MeasuresMeasures.html Accessed on 15 Jan 2016.

Chapin R., Wilkinson D.S., Rachlin R., et al. "Going Home: Community Reentry of Light Care Nursing Facility Residents Age 65 and Over." Journal of Health Care Finance. 25 (1998): 35-48.

Chapin R., Baca B., Macmillan K., et al. "Residential Outcomes for Nursing Facility Applicants Who Have Been Diverted: Where are

they 5 Years Later?" The Gerontologist. 49 (2009): 46-56.

Langley G.L., Moen R., Nolan K.M., et al. (2009). The Improvement Guide: A Practical Approach to Enhancing Organizational Performance (2nd edition). San Francisco: Jossey-Bass Publishers.

Mollica R. (2006). Transitions From Nursing Homes to Community Settings. In Berkman B, D'Ambruoso S (Eds.), Handbook of Social Work in Health and Aging (pp. 661-665). New York, NY: Oxford University Press.

Mor V., Zinn J., Gozalo P., et al. "Prospects for Transferring Nursing Home Residents to the Community." Health Affairs. 26 (2007): 1762-1771.

Wodchis W., Teare G., Naglie G., et al. "Skilled Nursing Facility Rehabilitation and Discharge to Home After Stroke." Arch Phys Med Rehabil. 86 (2005): 442-448.

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.)*
Not Applicable

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):*

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Care Coordination : Readmissions, Safety : Readmissions

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.)*
Not Applicable

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)*
No data dictionary **Attachment:**

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Not Applicable

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 measured is the number of new admissions from an acute care hospital discharge to community from a skilled nursing center. More specifically, the numerator is the number of stays discharged back to the community (i.e. private home, apartment, board/care, assisted living, or group home as indicated on the MDS discharge assessment form) from a skilled nursing center within 100 days of admission and remain out of any skilled nursing center for at least 30 days.

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.)*
The measure is based on a rolling 12 month window of MDS entries, which is updated quarterly.

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.
Data for the numerator comes from MDS 3.0 discharge assessments.

The numerator is the number of new admissions from an acute care hospital discharged back to the community (as indicated by MDS item A2100=01 'discharge into the community') alive from a skilled nursing center within 100 days of admission and remain out of any skilled nursing center for at least 30 days. All new admissions (regardless of payor status at time of admission to the facility or time of discharge back to the community) are counted as long as they are discharged back to the community within 100 days and do not have a subsequent stay in any nursing center within 30 days.

The "within 100 days from admission" time frame is measured by subtracting date of admission (MDS item A1900 "admission date") from date of discharge (MDS item A2000 "discharge date"). Subsequent stays in any nursing center within 30 days of discharge are determined by subtracting admission date (MDS item A1900 "admission date") from target date (MDS itemTRGT_DT) and ensuring that this isn't greater than 130 days (i.e. 100 days (of admission for this entry) + 30 days (after discharge) <=130).

Stays that discharge to death are not counted as a discharge in the numerator.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The denominator is the total number of all admissions from an acute hospital (MDS item A1800 "entered from"=03 (indicating an "acute care hospital") to a center over the previous 12 months, who did not have a prior stay in a nursing center for the prior 100 days (calculated by subtracting 100 from the admission date (MDS item A1900 "admission date").
Please note, the denominator only includes admissions from acute hospitals (MDS item A1800 "entered from"=03 (indicating an "acute care hospital") regardless of payor status.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk : Dual eligible beneficiaries, Populations at Risk : Individuals with multiple chronic conditions, Senior Care

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 denominator is the number of all stays (regardless of payor status) admitted from an acute care hospital (as indicated by MDS item A1800 "entered from"= 03 "acute care hospital") to a center over the prior 12 months, who did not have a prior stay in a nursing center for the prior 100 days (as indicated by MDS item A1600 "most recent admission/entry or reentry to this facility: entry date," and item A1800 "entered from").

For example, if the "entry date" (MDS item A1600) is within 100 days from the current admission and the "entered from" (MDS item A1800) is 02 "another nursing home" then these patients are excluded from denominator.

Note that our stay grouping algorithm allows interruptions in the stay, so long as the patient returns to the same facility within 100 days of the original admission. Once a new stay has started, if the patient discharges from the SNF and then returns to the same facility within 100 days of the original admission date, then that subsequent time in the SNF is considered to be part of that original stay. Then, when the patient discharges and does not return to the facility (within 100 days of the original admission date), the

discharge status code (community discharge, acute hospital, etc.) is the final outcome. For example, if Bill first entered the SNF on February 14th and then was hospitalized on March 10th, returned to the same SNF on March 15th, and then discharged to the community on April 1st, and never came back to the SNF, then Bill would count once in the denominator and once in the numerator. The original and subsequent stay start dates are identified using the entry date, MDS item A1600.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
The denominator has three exclusions (see below).

First, stays for patients less than 55 years of age are excluded from the measure.

Second, stays for which we do not where the patient entered from, or for which we do not observe the patient's discharge, are excluded from being counted in the denominator.

Third, stays with no available risk adjustment data (clinical and demographic characteristics listed in Section S.14) on any MDS assessment within 18 days of SNF admission are excluded from the measure.

Note, while not denominator exclusions, we also suppress the data for facilities that have fewer than 30 stays in the denominator, or for whom the percent of stays with a known outcome is less than 90%. The suppression of risk adjusted to community rates for facilities with fewer than 30 stays in the denominator is to improve the reliability of the measure, as detailed in the testing section (2b3). The suppression of rates for facilities for whom fewer than 90% of stays had a known outcome is done to improve the reliability of the measure and avoid perverse incentives about submitting MDS assessments for patients not discharged to the community.

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)*
First, individuals less than 55 years of age (as indicated by subtracting birth date, MDS item A0900, from admission date, MDS item A1900) are excluded from the measure.

Second, exclusions are made for admissions for which there is missing data over the previous 12 months for MDS item A1800 "Entered From" or MDS item A2100 "Discharge Status".

Third, if individuals have no available risk adjustment data on any MDS assessment within 18 days of SNF admission, they are excluded from the measure.

As noted above, in addition to the denominator exclusions, we also suppress data for facilities that have fewer than 30 stays in the denominator or for whom the percent of stays with a known outcome is less than 90%. Facilities with fewer than 30 stays in the denominator, are identified by counting the stays remaining after applying the exclusions in this section to the denominator. Facilities for whom fewer than 90% of stays have known outcomes, are measured by looking at all entries for the facility and seeing how many of those entries also have a discharge assessment.

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)*
Not Applicable

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)*
Risk adjustment for the measure was completed by means of logistic regression using independent variables drawn from the admission to SNF and discharge from SNF MDS 3.0 assessments. When information was not available on the admission MDS assessment, information from the next available MDS of any type (except discharge MDS assessment) was used, as long as the MDS was completed within 18 days of admission to the center; if no such complete assessment exists on entry or within 18 days, the stay

is excluded from the denominator per the denominator exclusions.

The following lists the variables used in the logistic regression risk adjustment model. There are 60 different MDS items, which are encoded across 116 variables in the final risk model (e.g., age and age-squared; interaction terms; etc.). The respective MDS 3.0 codes used to determine whether or not each variable contributes to the calculation are provided in Section S.15 below.

Demographic:
-Age
-Gender
-Marital Status

Functional Status:
-Vision
-Makes Self-understood
-Ability to Understand

Functional Status (cognitive, mobility and self care):
-Any Sign/symptom of Delirium
-Major Depression
-Behavioral Code (i.e. Hallucination, Delusion, Physical Behavior, Verbal Behavior, Other Behavior)
-Any Rejection of Care
-Medicare RUG IV Hierarchical Group
-Activities (i.e Bed Mobility, Transfer, Walk in Corridor, Locomotion, Eating, and Personal Hygiene)
-ADL summary (Combination of Bed Mobility, Transfer, Locomotion, Dressing, Eating, Toilet Use, Hygiene)
-ADL*Cognitive Impairment: Interaction Term
-Bathing
-Balance (i.e. Moving from Seated to Standing, Walking, Turning Around and Facing the Opposite Direction, and Moving On and Off Toilet)
-Urinary Incontinence
-Bowel Incontinence

Prognosis:
-Any acute Hospitalization within 30 days of Admission
-Special Treatment/Programs: Hospice Post-Admission
- Life Expectancy of less than 6 months

Clinical Conditions:
-Shortness of Breath when Exertion
-Shortness of Breath when Sitting
Shortness of Breath when Lying Flat
-Any Swallowing Disorder
-Weight Loss
-Pressure Ulcer
-Wound Infection
-Hemiplegia
-Paraplegia

Clinical Treatments:
-Oxygen Post-admit
-Tracheostomy Post-admit
-Ventilator Post-admit
-Dialysis Post-admit
-Max Number Injections
-Antipsychotic Use

Clinical Diagnosis:
-Anemia

-Heart Failure
-Hypertension
-Pneumonia
-Septicemia
-Urinary Tract Infection (UTI)
-Viral Hepatitis
-Diabetes Mellitus
-Hyperkalemia
-Hyperlipidemia
-Hip Fracture
-Other Fracture
-Alzheimer's Disease
-Stroke
-Dementia
-Huntington's
-Malnutrition
-Anxiety Disorder
-Depression
-Manic Depression
-Psychotic
-Schizophrenia
-Asthma, COPD, Chronic Lung Disease

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.
Provided in response box S.15a

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
The expected rate for a single SNF is calculated using the formula below. The calculation must be performed at least 45 days after the end of the target rolling 12-month period. This is to allow a full 30 days to elapse to capture re-admissions to any SNF that may occur on the last day of the target period plus another 14 days to allow facilities to submit data to CMS (and then rounding up to 45). In addition to the 45 days, to ensure maximum data availability, we recommend waiting up to 3 weeks beyond the 45 day window to ensure maximum data availability for MDS assessments not submitted during the 14 day period.

FORMULA
Note, the detailed mathematical definitions of each of these variables is presented in the appendix at S.15. For each of these variables, if the value is missing, first attempt to impute using the facility mean for the variable, and if the variable was entirely missing for the facility, impute using the population mean.

We then calculate the expected discharge to community rate for the stay as follows.

Log Odds= -1.41199
-0.27016*Pressure ulcer
-0.02102*ADL*cog_impair1
+0.010727 *ADL*cog_impair2
+0.058283 *ADL*cog_impair3
+0.055554 *Age
-0.00043*Age squared
-0.16875*Antipsychotic use
-1.98184*Entry from an acute care hospital (up to 30 days before SNF entry)
-0.27486*Behavior
-0.1844 *Patient rejected care during their current stay
-0.21135*Bathing: Supervision-Oversight help only (i.e.
-0.25798*Bathing: Physical help limited to transfer only
-0.32613*Bathing: Physical help in part of bathing activity

-0.40808*Bathing: Total dependence or activity did not occur or activity occurred only once or twice

+0.171826 *Bed mobility: Supervision
+0.336257 *Bed mobility: Limited assistance
+0.505795 *Bed mobility: Extensive assistance
+0.535251 *Bed mobility: Total dependence , Activity occurred only once or twice, Activity did not occur

-0.17849*Bowel incontinence: frequently incontinent or always incontinent
-0.07739*Bowel incontinence: Not rated
+0.801006 *Cognitive impairment: Score at least 13
-0.13541*Cognitive impairment: Score between 8 and 12
-1.24006*Cognitive impairment: Score less than 8
-0.0661 *Eating - Self-Performance: Supervision
-0.07192*Eating - Self-Performance: Limited assistance
-0.17375*Eating - Self-Performance: Extensive assistance
-0.31162*Eating - Self-Performance: Total dependence, activity occurred only once or twice, or activity did not occur

-0.0783 *Moving from seated to standing, Not steady, but able to stabilize without human assistance

-0.07909*Moving from seated to standing, Not steady, only able to stabilize with human assistance

-0.25663*Moving from seated to standing, Activity did not occur
+0.032238 *Walking: Not steady, but able to stabilize without human assistance
-0.02074*Walking: Not steady, only able to stabilize with human assistance
-0.21705*Walking: Activity did not occur
+0.036453 *Turning around and facing the opposite direction: Not steady, but able to stabilize without human assistance

+0.002928 *Turning around and facing the opposite direction: Not steady, only able to stabilize with human assistance

-0.04765*Turning around and facing the opposite direction: Activity did not occur
-0.00847*Moving on and off toilet: Not steady, but able to stabilize without human assistance

+0.031553 *Moving on and off toilet: Not steady, only able to stabilize with human assistance

-0.1183 *Moving on and off toilet: Activity did not occur
-0.06206*Personal Hygiene: Supervision
-0.27528*Personal Hygiene: Limited assistance
-0.44864*Personal Hygiene: Extensive assistance
-0.47705*Personal Hygiene: Total dependence, activity occurred only once or twice, or activity did not occur

-0.24638*Any Sign/Symptoms of Delirium
+0.094727 *Locomotion on unit: Supervision
+0.174147 *Locomotion on unit: Limited assistance
+0.216346 *Locomotion on unit: Extensive assistance
+0.139412 *Locomotion on unit: Total dependence, Activity occurred only once or twice, activity did not occur

-0.27615*Major Depression
-0.15642*Male
+0.335516 *Married
+0.016322 *Maximum number of injections of N0300 and N0350a
+0.015199 *Anemia
-0.12524*Heart Failure (CHF)
+0.054845 *Hypertension
+0.065189 *Pneumonia
+0.165911 *Septicemia
-0.01395*Urinary Tract Infection (UTI)
-0.13166*Viral Hepatitis

+0.042279 *Wound Infection
-0.09013*Diabetes Mellitus
-0.13801*Hyperkalemia
+0.125916 *Hyperlipidemia
+0.362973 *Hip Fracture
+0.196092 *Other Fracture
-0.23826*Alzheimer's Disease
-0.0651 *Stroke (CVA or TIA or Stroke)
-0.22697*Dementia
-0.16149*Hemiplegia
+0.138843 *Paraplegia
-0.70189*Huntington's
-0.1581 *Malnutrition
-0.06198*Anxiety Disorder
-0.04714*Depression
-0.07149*Manic Depression
-0.16682*Psychotic
-0.48623*Schizophrenia
-0.01384*Asthma, COPD, Chronic Lung Disease
-0.84092*Prognosis: (Life expectancy of < 6 months)
-0.08602*Shortness of Breath With Exertion
-0.11828*Shortness of Breath When Sitting
-0.0923 *Shortness of Breath When Lying Flat
-0.10945*Oxygen Post-admit
-0.32907*Tracheostomy Post-admit
-0.71096*Ventilator Post-admit
-0.25879*Dialysis Post-admit
-1.55938*Special Treatments/Programs: Hospice Post-admit
+0.16806 *Medicare RUG IV Hierarchical Group2: moderate/high nursing, no therapy
+0.664998 *Medicare RUG IV Hierarchical Group3: very low nursing and therapy
+0.632329 *Medicare RUG IV Hierarchical Group4: lower nursing, therapy, but have both

+0.972862 *Medicare RUG IV Hierarchical Group5: moderate nursing, moderate/high therapy

+0.859821 *Medicare RUG IV Hierarchical Group6: moderate nursing, moderate/high therapy

+0.705103 *Medicare RUG IV Hierarchical Group7: moderate nursing, moderate/high therapy
+1.14548 *Medicare RUG IV Hierarchical Group8: moderate nursing, moderate/high therapy

+0.987026 *Medicare RUG IV Hierarchical Group9: moderate nursing, high therapy
+0.801595 *Medicare RUG IV Hierarchical Group10: moderate nursing, high therapy
+0.548214 *Medicare RUG IV Hierarchical Group11: high nursing, low therapy
+0.813968 *Medicare RUG IV Hierarchical Group12: high nursing, high therapy
-0.0882 *Makes Self Understood: Usually understands
-0.1021 *Makes Self Understood: Sometimes understood or rarely/never understood
-0.13229*Swallowing Disorder
+0.200977 *Transfer - Self-Performance: Supervision
+0.293408 *Transfer - Self-Performance: Limited supervision
+0.30866 *Transfer - Self-Performance: Extensive supervision
+0.053075 *Transfer - Self-Performance: Total dependence, Activity occurred only once or twice, activity did not occur

-0.11822*Ability to understand : Usually understands
-0.10047*Ability to understand (B0800) is 2 or 3: Sometimes understands or rarely/never understands

-0.23221*Urinary incontinence: frequently incontinent or always incontinent
-0.29577*Urinary incontinence: Not rated
-0.14014*Resident had impaired vision

-0.16773*Resident had moderately impaired, highly impaired, or severely impaired vision

+0.160179 *Walk in Corridor: Activity occurred with supervision
+0.105846 *Walk in Corridor: Activity occurred with limited assistance
+0.002214 *Walk in Corridor: Activity occurred with extensive assistance
-0.10974*Walk in Corridor: Activity occurred only once or twice, or activity did not occur (activity or any part of the ADL was not performed by resident, or staff at all over the entire 7-day period), or activity occurred with total dependence.

-0.05961*Weight loss

Finally, calculate the stay's expected discharge back to community rate as 1/ (1+exp (-LogOdds)).

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 = Higher 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.)*
The formula for the risk-adjusted discharge to community rate is:

((Observed discharge to community alive within 100 days of admission and remaining out of any SNF for at least 30 days rate)/ (Expected discharge to community alive within 100 days of admission and remaining out of any SNF for at least 30 days rate)) * (National discharge to community alive within 100 days of admission and remaining out of any SNF for at least 30 days rate).

Note: The national rate and the expected rate need to be calculated for the same time period so that their ratio across the nation will center around 1.0, i.e., the risk adjustment does not systematically bias up or down the rates.We recommend the national rate and expected rates be recalibrated at least annually.

1. Build the denominator population, applying exclusions:
-Establish the 12 month rolling time period and collect all the assessments for an admissions from an acute care hospital (for patients who did not have a prior stay in a nursing center for the prior 100 days) that fall within the time period.
-Identify all MDS assessments through the stay, up to discharge. If no discharge is observed, the stay does not have a known outcome and is excluded from the denominator population. Note that if the patient is discharged (e.g., a hospitalization after which the patient returns to the SNF), but then returns to the same SNF within 100 days of the original admission, then the stay is continued to be ongoing, and we continue to search for the final discharge.
-If the stay had missing data on the "admitted from" MDS item (to identify admissions from the acute hospital) or on the "discharged to" item (to identify discharges to the community).
-Identify whether the patient was seen in a SNF in the 30 days after discharge from the current stay, which indicates the patient's outcome was not a successful community discharge for the purpose of this measure. This is accomplished by looking for any MDS for that individual in any SNF during the 30 day widow following SNF discharge to the community.
-Identify any MDS assessments for the patient in the 100 days prior to the stay's admission. If any are found, exclude the stay from the denominator.
-If the patient was under 55 years of age on admission to the stay, exclude the stay from the denominator population.

2. Observed Rate Calculation:
-The formula for a facility's observed discharge to community rate is:
(The number of stays discharged back to the community (i.e. private home, apartment, board/care, assisted living, or group home as indicated on the MDS 3.0 discharge assessment form) from a skilled nursing center within 100 days of admission and remain out of any skilled nursing center for at least 30 days)/ (all admissions from an acute hospital to a center over the prior 12 months that do not meet the exclusions)
-The numerator is the number of stays in the denominator that are discharged back to the community from a SNF within 100 days of

admission and remain out of any skilled nursing center for at least 30 days upon discharge, during a rolling 12 month period.
-For example, if a center discharged 130 stays (that were admitted from an acute care hospital and that did not have a prior stay in a nursing center for the prior 100 days), but 30 of them were readmitted to a skilled nursing center within 30 days following discharge, the numerator would be 100 (i.e. 130-30=100).
-Divide the numerator by the denominator to obtain the observed rate for the skilled nursing center.

3. Expected Rate Calculation
-See S.15
-For each SNF, calculate the facility-level mean of the stay-level expected rates of discharging back to the community, from the calculation in S.15; this is the overall expected rate of discharging back to the community for the SNF based on its denominator population.

4. National Average
-The national average is calculated as the sum of all residents in the nation who were discharged to the community (and remained out of a SNF for at least 30 days) divided by the sum of all admissions to SNF (regardless of payor status) from acute care hospitals during a calendar year and did not have a prior stay in the nursing home.

5. Divide the observed rate by the expected rate and multiply by the national rate to obtain the adjusted discharge to community rate for the center.

6. Suppress the risk adjusted discharge to community rates for SNFs with fewer than 30 stays in the denominator, or with a "known outcome rate" of less than 90%. The known outcome rate for the facility is the proportion of stays in the denominator (excepting the known outcome exclusion) for which the outcome is unknown.

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)*
No diagram provided

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.
Not Applicable

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.
Not Applicable

S.22. Missing data (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Missing risk adjusters were imputed with the facility average where available, if this was not available the population average was used. If all risk adjusters were missing the stay was dropped from the denominator. Similarly, if no discharge assessment was found for the stay then the stay was dropped from the denominator.

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

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.
Minimum Data Set (MDS) 3.0

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)*
Available in attached appendix at A.1

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)*
Post Acute/Long Term Care Facility : Nursing Home/Skilled Nursing 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.)*
Not Applicable

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
Discharge_to_Community_Testing_Final.docx

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

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Discharge to Community
Date of Submission: 1/29/2016
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☒ Outcome (*including PRO-PM*)
☐ 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: Nursing Facility MDS 3.0	☒ other: Nursing Facility MDS 3.0; Nursing Home Compare Five Star Ratings and Quality Measures; PointRight Pro 30 Rehospitalization Rates

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*).

To develop the specifications (i.e. exclusion criteria, risk adjustment variables) for the discharge to community measure, we used admission assessments from the MDS 3.0.

To test the validity of the measure we used two samples. First, to test the validity of MDS discharge coding, we matched calendar year 2012 MDS assessments with Part A Medicare inpatient and SNF claims and enrollment data. Second, to validate the relationship between the discharge to community measure and other measures we used discharge to community rates calculated on calendar year 2014 MDS 3.0 data. The other measures included Five Star Ratings from the December 2014 release of Nursing Home Compare, Nursing Home Quality Metrics for the 3 quarter period ending December 2014, and PointRight® Pro 30™ rehospitalization rates for the year ending December 2014.

1.3. What are the dates of the data used in testing? 01/01/2014-12/31/2014; 01/01/2012-12/31/2012

To perform measure testing, we used assessments with entry dates: 01/01/14-12/31/14, and for the validation matching with claims and enrollment data, we used MDS discharge assessments dated 1/1/2012-12/31/2012.

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

Measure Specified to Measure Performance of:	**Measure Tested at Level of:**

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

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Table 1.5.a presents the facility-level summary of measured entities in the two analytical samples described in Sections 1.2 and 1.3. The Five Star ratings, Nursing Home Compare Quality Measures, and PointRight Pro-30 Short Stay Reshopitalization used in the construct validity analysis included all SNFs in the country, i.e., essentially matching the samples in Tables 1.5.a.

Table 1.5.a. Facility characteristics of the two analytical samples

	Main CY2014 MDS stays		2012 MDS discharges matched with Medicare FFS data	
	N	%	N	%
Total facilities	***15,464***	***100%***	***15,488***	***100%***
Ownership control type:				
For profit	10,684	69%	9,328	60%
Non-profit	3,659	24%	3,314	21%
Government	898	6%	695	4%
Unknown	223	1%	2,151	14%
Certified beds:				
1-49	1,868	12%	1,621	10%
50-99	5,615	36%	4,936	32%
100-149	5,151	33%	4,630	30%
150-199	1,678	11%	1,534	10%
200+	929	6%	862	6%
Unknown	223	1%	1,905	12%

1.6. How many and which patients 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*)

Table 1.6.a presents the breakdown of stays in the two analytical datasets used in the testing, as described in Section 1.5 and Table 1.5.a.

Table 1.6.a, part 1. Demographic and diagnostic characteristics of the stays in the two testing samples

	Main CY2014 MDS stays		2012 MDS discharges matched with Medicare FFS data	
Total stays	***1,965,321***	***100%***	***1,879,934***	***100%***
Age				
<55	-	0%	58,636	3%
55-64	215,389	11%	116,663	6%
65-74	453,872	23%	386,335	21%
75-84	652,985	33%	639,265	34%
84-94	574,292	29%	600,973	32%
95+	68,783	3%	78,062	4%
Sex				
Male	749,963	38%	728,455	39%
Female	1,215,358	62%	1,151,335	61%
Marital status				
Never married	211,547	11%	209,553	11%
Married	680,105	35%	616,001	33%
Widowed	767,616	39%	795,347	42%
Separated	20,778	1%	21,144	1%
Divorced	217,477	11%	202,525	11%
Unknown	67,798	3%	35,364	2%
Race				
White	1,564,736	80%	1,532,074	81%
Black	193,057	10%	200,287	11%
Asian	32,180	2%	23,528	1%
Hispanic	80,697	4%	71,803	4%
American indian	6,666	0%	6,285	0%
Hawaiian	2,700	0%	2,047	0%
Unknown	85,285	4%	43,910	2%

Table 1.6.a, part 2. Demographic and diagnostic characteristics of the stays in the two testing samples

	Main CY2014 MDS stays		2012 MDS discharges matched with Medicare FFS data	
Total stays	***1,965,321***	***100%***	***1,879,934***	***100%***
Active diagnoses coded on MDS assessment (note %s exclude missing values from the denominator)				
I0200 Active Diagnoses: Anemia	539,305	27%	213,229	32%
I0600 Active Diagnoses: Heart Failure (CHF)	397,472	20%	164,239	25%
I0700 Active Diagnoses: Hypertension	1,489,215	76%	495,806	75%
I2000 Active Diagnoses: Pneumonia	181,261	9%	69,850	10%
I2100 Active Diagnoses: Septicemia	38,902	2%	13,465	2%
I2300 Active Diagnoses: Urinary Tract Infection (UTI)	275,326	14%	256,981	15%
I2400 Active Diagnoses: Viral Hepatitis	13,014	1%	1,783	1%
I2500 Wound Infection	34,880	2%	16,336	2%
I2900 Active Diagnoses: Diabetes Mellitus (DM)	643,652	33%	624,091	36%
I3200 Active Diagnoses: Hyperkalemia	17,587	1%	7,810	1%
I3300 Active Diagnoses: Hyperlipidemia	846,727	43%	252,227	38%
I3900 Active Diagnoses: Hip Fracture	155,513	8%	40,079	6%
I4000 Active Diagnoses: Other Fracture	219,091	11%	49,977	8%
I4200 Active Diagnoses: Alzheimers Disease	70,180	4%	9,681	4%
I4500 Active Diagnoses: Stroke (CVA or TIA or Stroke)	217,417	11%	76,982	12%
I4800 Active Diagnoses: Dementia	320,097	16%	121,867	18%
I4900 Active Diagnoses: Hemiplegia	74,860	4%	27,484	4%
I5000 Active Diagnoses: Paraplegia	5,112	0%	3,192	0%
I5250 Active Diagnoses: Huntingtons	488	0%	648	0%
I5600 Active Diagnoses: Malnutrition	66,502	3%	53,175	3%
I5700 Active Diagnoses: Anxiety Disorder	346,873	18%	336,794	19%
I5800 Active Diagnoses: Depression	544,866	28%	208,834	31%
I5900 Active Diagnoses: Manic Depression	35,174	2%	45,571	3%
I5950 Active Diagnoses: Psychotic	43,232	2%	64,073	4%
I6000 Active Diagnoses: Schizophrenia	23,889	1%	33,921	2%
I6200 Active Diagnoses: Asthma, COPD, Chronic Lung Disease	453,792	23%	180,762	27%

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.

We consistently used the two main analytical samples described in Sections 1.5 and 1.6 throughout the testing section. There were three exceptions to this. First, we tested reliability by comparing the final risk adjusted discharge to community measure for the year ending 2014q4 against the final risk adjusted discharge to community measure for the year ending 2014q3. Second, we tested construct validity by correlating the risk adjusted discharge to community rates against other, publicly available quality measures. Last, we analyzed the effect of the age<55 exclusion by adding those patients back into the analytical sample.

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).

The patient-level sociodemographic variables available and utilized in the analyses were: age, gender, marital status, and race.

Specifically, in the main calendar year 2014 MDS stays dataset, the majority (i.e. 61.8%) of the residents were female. The age of this group ranged from 55.0 to 116.6 years old, with a mean of 79.0 years old, and the standard deviation of 10.3 years. Marital status was missing for 3.5 % of this population. The two most prevalent marital status categories were widowed (i.e. 39.0%), followed by married (i.e. 34.7%). The majority (i.e. 80%) of the stays were white, 10% were black, and 4% were Hispanic (Table1.6.a).

In the 2012 MDS discharges matched with Medicare FFS data, the majority (i.e. 61.0%) of the stays were female residents (Table1.6.a). The age of this group ranged from 0 to 113 years old, with a mean of 79.1 years and a standard deviation of 11.1 years. The majority (i.e. 81%) of the stays were white residents, 11% were black residents, and 4% were Hispanic (Table1.6.a). A sizeable majority (i.e. 42%) of the stays were widowed residents, with married as the next most prevalent (i.e. 33%) marital status.

The exclusion testing included those younger than 55 years old, resulting in an age range from 0 to 116.6 years old, and a SD of 12.7 years, and a mean of 77.2 years old. The majority (i.e. 61.0%) of the patients were female. Here, the two most prevalent marital status categories were widowed (i.e. 37.0%) followed by married (i.e. 34.1%), with 3.6% of the population missing information on marital status.

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*)

Two methods were used to test for reliability. The first method compared outcomes before and after resampling with replacement bootstrapping. This essentially measures reliability between different facilities admitting patients from the same underlying distribution. The second was a comparison of facility rates from one quarter to the next. This measures the reliability of the measure across time.

Method #1: Replacement Bootstrapping

For this method, we conducted a random resampling of the population with replacement. This simulates a facility or two facilities of similar size independently drawing patients from the same underlying patient population. We compared facility-level outcomes after resampling to outcomes before sampling. To correctly measure the stability of the samples, we applied both the denominator size and known outcome rate suppression criteria after resampling.

Method #2: Performance Comparison between Quarters

For this method, we compared facility-level outcomes in Q3-2014 to Q4-2014. This analyzes the actual stability in a facility's rate between consecutive quarterly releases of the discharge to community measure. All reliability tests were at the facility level, i.e., the level of measurement of the performance measure.

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*)

Results from Method #1 (Replacement Bootstrapping)

Table 2a2.3.a: Changes in Risk Adjusted Absolute Rates after Re-Sampling

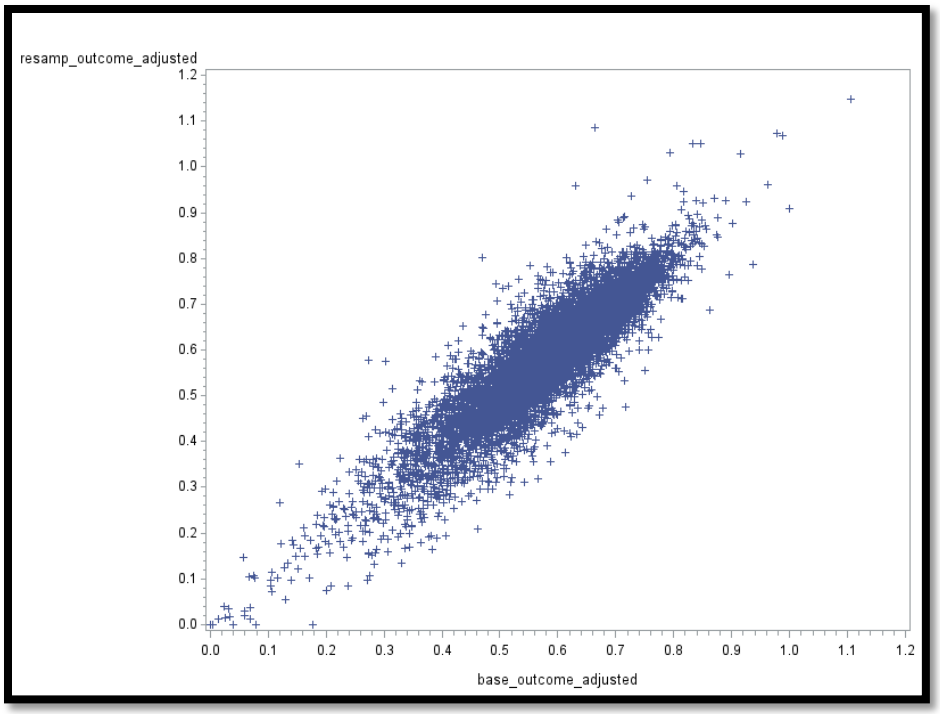
Absolute Rate Change Ranges	# of Facilities	Percent of Facilities	Cumulative Percent
0-5	8,980	76.8%	76.8%

>5-8	1,554	13.3%	90.1%
>8-10	502	4.3%	94.4%
>10-20	616	5.3%	99.7%
>20-30	36	0.3%	100.0%
>30	1	0.0%	100.0%

Table 2a2.3.b: Changes in Percentile Rankings after Re-Sampling

Percentile Change Ranges	# of Facilities	Percent of Facilities	Cumulative Percent
0-5 Percentiles	5,313	45.4%	45.4%
>6-10	2,642	22.6%	68.0%
>11-15	1,537	13.1%	81.1%
>16-20	940	8.0%	89.2%
>21-25	554	4.7%	93.9%
>26-30	293	2.5%	96.4%
>30	418	3.6%	100%

Figure 2a2.3.1: Facility Scatter Plot: Rate Before Re-Sampling vs Rate after Re-Sampling



Results from Method #2 (Quarter Performance Comparison)

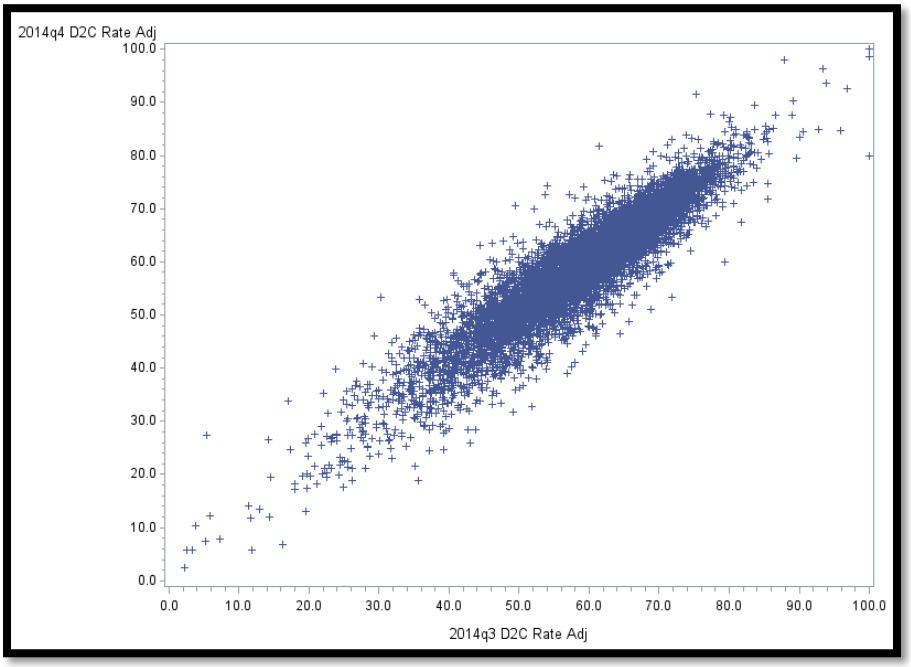
Table 2a2.3.c: Changes in Absolute Rates from Q3-2014 to Q4-2014

Absolute Rate Change Ranges	# of Facilities	Percent of Facilities	Cumulative Percent
0-5	9,421	84.8%	84.8%
>5-6	550	4.9%	90.0%
>5-10	906	8.2%	97.9%
>10-20	226	2.0%	100%
>20-30	6	0.1%	100%
>30	0	0	100%

Table 2a2.3.d: Changes in Percentile Rankings from Q3-2014 to Q4-2014

Percentile Change Ranges	# of Facilities	Percent of Facilities	Cumulative Percent
0-5 Percentiles	5,618	50.6%	50.6%
>6-10	2,550	23.0%	73.5%
>11-15	1,376	12.4%	85.9%
>16-20	718	6.5%	92.4%
>21-25	415	3.7%	96.1%
>26-30	204	1.8%	97.9%
>30	228	2.1%	100%

Figure 2a2.3.2: Facility Scatter Plot: from Q3-2014 Rate vs Q4-2014 Rate



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?*)
Based on the results from replacement bootstrapping, if a SNF's patients were completely redrawn from the same underlying population (e.g., the same SNF a year in the future), or if we compare two SNFs who each draw patients from the same underlying population, then 68% of the time they will remain ranked within ten percentiles of where they were before redrawing patients. In 96% of cases, they would shift less than thirty percentiles after random resampling. Similarly, a SNF's rate does not shift very much if it completely redrew its population. Their rate would shift within five points 76% of the time and within ten points 95% of the time.

The results of comparing performance between consecutive quarters further supports the idea that the measure is highly reliable. Between Q3-2014 and Q4-2014, 74% of facilities remained ranked within ten percentiles; 98% remained ranked within thirty percentiles. Facility rates also stayed relatively stable from one quarter to the next with 85% changing less than five points and 98% changing less than ten points. This validates the notion of gaining stability across time due to the same patients appearing in consecutive rolling years of risk adjusted discharge to community rates.

Given that the most important comparison between facilities is the top tertile (high performers), middle tertile (medium performers), and bottom tertile (low performers), the level of stability demonstrated in percentile rank and outcome rate is very acceptable and supports the measure's overall reliability.

_____

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)*
We performed two kinds of empirical validity testing. First, a validation of the coding of discharges to the community using matched Part A claims data. Second, we performed construct validity testing by correlating risk adjusted discharge to community rates with certain other measures hypothesized to be driven by the same factors driving discharge to community rates.

First, empirical MDS validation of the coding of discharges to the community.

To validate the accuracy of the MDS coding of discharges to the community, we matched Part A and Medicare enrollment data with the MDS discharges for short stays (< 100 days) that admitted from the acute hospital. This parallels our broad definition of the discharge to community measure. We searched for death records and for Part A inpatient hospital and SNF claims that indicated the person had not been discharged to the community following what we observe in the SNF. The Part A claim must have had a "from date" (first date the claim refers to) on the MDS discharge date, or within 4 days after discharge. While it is possible that non-Medicare Part A activity may have occurred, if there is a high proportion of the MDS discharges to the community that does not have a Part A claim, this is strong supporting evidence of the quality of the MDS discharge coding. We used the most recent claims data available, the CY2012 Research Identifiable Files from CMS.

Second, construct validity correlating discharge to community rates with other quality metrics.

We conducted construct validity testing comparing the discharge to community measure to other measures of SNF quality. We hypothesized that facilities with higher discharge to community would correlate with Five Star ratings (particularly the quality measure (QM) component of Five Star), CMS nursing home compare short stay quality measures, and the facility's 30 day rehospitalization rate. We did not test a relationship with other individual long stay CMS nursing home quality measures on nursing home compare, since this measure applies to short stay patients, i.e., those discharged to the community within 100 days of admission to a SNF and not to individuals requiring long term nursing home care.

We used the survey, quality, and staffing components from the Five Star rating. Specifically, we grouped facilities by their Five Star rating and calculated the mean discharge to community measures for each grouping and calculated the Pearson's correlation coefficient. We hypothesized that facilities with higher Five Star ratings would have higher discharge to community measure scores.

We also hypothesized that facilities with a lower score on their individual CMS short stay quality measures would have a higher discharge to community score (i.e., a negative correlation) since lower scores on the short stay quality measures indicates better quality and higher score for discharge to community indicates better quality. Specifically, the short stay antipsychotic rates would have the strongest negative correlation with the discharge to community rates, because severely ill patients are less likely to improve quickly and be discharged back to the community. Furthermore, we hypothesized pressure ulcers would have a negative correlation with the discharge to community measure. On the other hand, we hypothesized that being administered the influenza and pneumococcal vaccine is positively correlated with the discharge to community measure, because gaining immunity against these infections, prevents getting the flu and pneumonia, which would otherwise delay the discharge to community. We hypothesized that short stay pain would not necessarily be correlated with discharge to community, because such pain is likely to be resolved quickly, thus not impacting the discharge to community.

We also hypothesized that there would be a negative correlation between risk adjusted short stay rehospitalization rates (NQF measure 2375) and discharge to community score.

All validity tests were at the facility level, i.e., the level of measurement of the performance measure.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)
First, empirical MDS validation of the coding of discharges to the community.

Table 2b2.3.a presents the results of the analysis of matched CY2012 short stay MDS discharges (that admitted from the acute hospital) with CY2012 Part A Medicare claims and enrollment data. Of the 993,916 MDS discharges to community in our analytical dataset, we confirmed that 95% of MDS discharges do not have a Part Medicare claim in the inpatient or SNF settings on the MDS discharge date or the four days following.

Table 2b2.3.a: CY2012 Medicare Part A Claims Results

			Medicare Part A claims in the 4 days post-discharge…						
	MDS Discharges		**Alive but no IP/SNF claim**	**SNF/Swing Bed**	**STACH/ CAH**	**Psych Hospital/ Unit**	**IRF/Rehab Unit**	**LTCH**	**Died**
	N	Col %	Row %	Row %	Row %	Row %	Row %	Row %	Row %
Short stay MDS discharges	***1,879,934***	***100%***	***56%***	***5%***	***30%***	***1%***	***0%***	***0%***	***7%***
Community	993,916	53%	95%	1%	4%	0%	0%	0%	0%
SNF/NF	63,398	3%	32%	64%	3%	0%	0%	0%	0%
Acute hospital	661,054	35%	11%	6%	79%	1%	0%	0%	2%
Psychiatric hospital	9,064	0%	15%	1%	13%	69%	0%	1%	0%
IRF	4,422	0%	15%	14%	5%	0%	62%	3%	0%
LTCH	983	0%	36%	12%	8%	1%	2%	41%	0%
Other	21,709	1%	65%	4%	8%	1%	0%	0%	22%
Died	125,388	7%	2%	0%	0%	0%	0%	0%	98%

Second, construct validity correlating discharge to community rates with other quality metrics.

With respect to the relationship between the discharge to community rate and the short stay rehospitalization rate, as suspected we found a negative and statistically significant relationship (Pearson's correlation =-0.092, p<.0001). The negative correlation was expected, because higher scores of discharge to community measure are indicative of higher quality, whereas lower scores of the short stay rehospitalization rate are indicative higher quality.

With respect to the relationship between the discharge to community rate and CMS Nursing Home Compare Short Stay quality measures, we also found statistically significant correlations (See Table 2b2.3.b). As hypothesized, the strongest negative correlation is between the short stay antipsychotic measure and the discharge to community measure. This is due to the fact that patients who are administered short stay antipsychotics may be more likely to need other therapy and treatment, which would prolong their SNF stay, and thus impact their discharge to community. Pressure ulcers rates had a negative correlation, which reflects the resident's low mobility. Severely ill patients most likely have lower mobility and the sicker the patient, the more treatment they will need during their SNF stay, which may decrease their likelihood of being discharged within 100 days of admission to a SNF.

As expected the administration of influenza and pneumococcal vaccine are positively correlated with the discharge to community score. We did observe a positive and statistically significant correlation between short stay pain and discharge to community. We hypothesize this positive relationship is because those with short stay pain may get the appropriate treatment or therapy which will lead to their pain being resolved, and ultimately to discharge to community.

Table 2b2.3.b: Correlation between CMS Nursing Home Compare Short Stay QMs and Discharge to Community Measure

CMS NHC Short Stay Quality Measures	Discharge to Community Score	P value
Short stay pressure ulcers	-0.11249	<.0001
Short stay pain	0.01952	0.0374
Short stay antipsychotics	-0.23376	<.0001
Short stay influenza	0.16500	<.0001
Short stay pneumococcal	0.19441	<.0001

With respect to the relationship between the Five Star rating (i.e. the overall rating, the survey, staffing, and quality components) and the discharge to community, as expected, we found a positive correlation (See Table 2b2.3.c). More specifically, as the overall Five Star Rating increased, the mean discharge to community score also increased. The same was true for the Survey, Nurse Staffing, as well as Quality Measure components of Five Star. This was expected because a higher Five Star Rating is indicative of higher quality. The same is true for the discharge to community score.

Table 2b2.3.c: Discharge to Community Measure to Five Star Rating

Overall Combined Five Star Rating	Mean Discharge to Community Score
1	56.4468214
2	58.2263937
3	59.2285128
4	60.8159138
5	62.1883575
Pearson's Correlation 0.17876;	

p<.0001	
Five Star Survey Rating	**Mean Discharge to Community Score**
1	57.7589151
2	59.5979018
3	60.2973784
4	61.2321566
5	61.9988146
Pearson's Correlation 0.12900 ; p <.0001	
Five Star Staffing Rating	**Mean Discharge to Community Score**
1	56.3791045
2	57.6646541
3	59.3785527
4	60.9328267
5	63.9830142
Pearson's Correlation 0.19921; p <.0001	
Five Star Quality Measure Rating	**Mean Discharge to Community Score**
1	58.6421053
2	58.5401818
3	59.1675748
4	59.8616587
5	60.5988327
Pearson's Correlation 0.06048; P <.0001	

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?*)

First, empirical MDS validation of the coding of discharges to the community.

We found that 95% of CY2015 short stay MDS discharges that admitted from the acute hospital had no Medicare Part A inpatient or SNF claim on the MDS discharge date or the four days following. While some portion of this will be discharges that followed institutional activity billed under a non-Medicare payer, the high

match rate indicates the coding of discharging to the community is valid and reliable for use in the discharge to community measure.

Second, construct validity correlating discharge to community rates with other quality metrics.

As hypothesized, we found that the discharge to community measure was positively correlated with being administered the influenza and pneumococcal vaccines, as well as Five Star ratings; and, again as expected, we found a negative correlation to short stay rehospitalization rate, short stay antipsychotic rate and pressure ulcer rate. Overall, the discharge to community measure followed other high quality measures' performance. Additionally, the magnitudes of these correlation coefficients were typical of those found in similar NQF-endorsed validity testing sections such as the PointRight® Pro 30™ short stay rehospitalization rate (NQF #2375), CARE: Improvement in Mobility (NQF #2612), and CARE: Improvement in Self-Care (NQF #2613). This supports construct validity of the discharge to community 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*)

We applied three patient-level exclusions to the denominator. First, patients under 55 were excluded from the measure. It should be noted that, this exclusion was selected by an expert panel of long term care clinicians a priori to the development and testing of the measure. In order to assess whether this exclusion induced any bias on the discharge to community measure, we decided to include the age group in the exclusions analysis. This lead to including 110,822 additional residents and 113,875 additional stays. Thus, we calculated the discharge to community measure including residents of all ages, and compared that to the discharge to community measure where those younger than 55 are excluded, all other things being equal. Therefore, any differences in the two percentages are attributable to this age exclusion.

Second, we excluded stays for which we do not where the patient entered from, or for which we do not observe the patient's discharge, are excluded from being counted in the denominator. These cases make the measure incalculable, and so we cannot test the effect of it on the measure.

Third, we excluded stays with no available risk adjustment data, or for whom we did not observe the stay's outcome (i.e., discharge MDS assessment was not found). These cases make the measure incalculable, and so we cannot test the effect of it on the measure.

Last, we tested the reliability suppression rules (fewer than 30 annual stays and tracking rates below 90%) using the same resampling approach we used in the main reliability testing (section 2a2). That is, we resampled the stays with replacement, and then compared the percentile rankings of the facilities before and after the resampling. We analyzed this measure of stability (changes in percentile rankings when the stays are redrawn from the same underlying distribution) by number of stays in the denominator, and by tracking rate. We then chose thresholds for the tracking rate and denominator size that preserved stability and had face validity. Note that these are not denominator exclusions, but rather are facility-level rules to not publish data for facilities with small denominator size, or with low rates of known outcomes.

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*)

Age. Including residents younger than 55 years old lead to having a total of 2,009,361 residents in 15,521 facilities, representative of 2,079,196 stays. Whereas, excluding those younger than 55 years old resulted in having a total of 1,898,539 residents in 15,464 facilities, representative of 1,965,321 stays. Therefore, the

exclusions represent 5.5% of the overall residents (i.e., 1,898,539 /2,009,361), or 5.5% of the total number of stays (i.e., 113,876/2,079,196).

We examined the impact of the age exclusion on the overall denominator size in all SNFs in the country by applying the exclusions to all admission from a hospital in SNFs stratified by number of admissions. Table 2b3.2.a shows that <2% or an additional 226 SNFs will have a sample size of <30 admissions from an acute care hospital after the exclusion is applied.

Table 2b3.2.a: Impact of Exclusions on Number of SNFs Based on Admission Size

Admissions (2014)	Count of facilities before applying exclusion	Count of facilities after applying exclusion	Difference	
	N	N	N	%
ALL	***15,521***	***15,464***	**-57**	**-0.37%**
00-29	3,299	3,525	226	1.46%
30-39	1,008	1024	16	0.10%
40-49	825	894	69	0.45%
50-59	789	776	-13	-0.08%
60-69	667	702	35	0.23%
70+	8,933	8,543	-390	-2.52%

We calculated the mean observed, predicted, and adjusted discharge to community before and after applying the age exclusion. Table 2b3.2.b shows that the mean adjusted discharge to community rate for the exclusion analysis was 59.5%, whereas the mean adjusted discharge to community excluding those younger than 55 years old was 59.62%.

Table 2b3.2.b: Comparison of Discharge to Community Rate Before and After Applying Exclusion

	Before Applying Exclusion (i.e. Ages>=0 y/o)	After Applying Exclusion (i.e. Ages>=55 y/o)
Mean observed discharge to community (%)	52.11	51.87
Mean predicted discharge to community (%)	57.93	57.86
National mean discharge to community (%)	63.26	63.31
Mean adjusted discharge to community (%)	**59.53**	**59.62**

Denominator size. Table 2b3.2.c presents the stability analysis by denominator size. The green highlighted region is the denominator sizes we included in the final measure, and the red highlighted region is the denominator sizes we excluded. The rightmost column presents the percent of facilities whose percentile

ranking changed by 20 percentile points or greater when we resampled the stays. Once we have fewer than 30 stays, the percent of facilities shifting rank by a large amount increased beyond about 20%, which we deemed too high; below that it was in the high teens, up to about 20%.

Table 2b3.2.c: Denominator Size

	Facilities		Stays		*Mean Percentile Change is within…*			
	N	%	N	%	<10%	10-19%	<20%	>=20%
Overall	15,070	100%	1,932,613	100%	66%	22%	88%	12%
>=100 stays	6,619	44%	1,573,707	81%	75%	21%	96%	4%
75-99 stays	1,442	10%	124,811	6%	61%	27%	88%	12%
50-74 stays	1,793	12%	110,497	6%	56%	26%	83%	17%
45-49 stays	460	3%	21,567	1%	58%	26%	83%	17%
40-44 stays	417	3%	17,520	1%	54%	27%	82%	18%
35-39 stays	500	3%	18,498	1%	56%	27%	83%	17%
30-34 stays	499	3%	15,949	1%	55%	23%	78%	22%
25-29 stays	556	4%	15,040	1%	57%	22%	79%	21%
20-24 stays	589	4%	12,980	1%	56%	17%	74%	26%
15-19 stays	581	4%	9,887	1%	56%	18%	74%	26%
10-14 stays	594	4%	7,176	0%	59%	14%	73%	27%
5-9 stays	550	4%	3,817	0%	63%	12%	75%	25%
<5 stays	470	3%	1,164	0%	86%	4%	90%	10%

^*Facility and stay counts different from those described in the sample discussion due to additional data preparation done for these analyses.*

Tracking rates. Table 2b3.2.d presents the stability analysis by denominator size. The green shaded region is the denominator sizes we included in the final measure, and the red highlighted region is the denominator sizes we excluded. The rightmost column presents the percent of facilities whose percentile ranking changed by 20 percentile points or greater when we resampled the stays. For facilities with a tracking rate below 90%, the proportion of facilities with large percentile ranking changes is not significantly higher or lower, overall, than facilities with high tracking rates.

Table 2b3.2.d: Tracking Rates

	Facilities		Stays		*Mean Percentile Change is within…*			
	N	%	N	%	<10%	10-19%	<20%	>=20%
Overall	11,939	100%	1,913,510	100%	67%	23%	91%	9%
100%	1,142	10%	136,462	7%	63%	25%	88%	12%
95% - <100%	10,104	85%	1,695,452	89%	67%	23%	91%	9%
90% - <95%	484	4%	50,635	3%	69%	20%	90%	10%
85% - <90%	110	1%	14,928	1%	70%	20%	90%	10%
80% - <85%	45	0%	6,333	0%	78%	18%	96%	4%
75% - <80%	18	0%	2,437	0%	83%	6%	89%	11%
70% - <75%	16	0%	2,594	0%	88%	6%	94%	6%
0% - <70%	20	0%	4,669	0%	100%	0%	100%	0%

^*Facility and stay counts different from those described in the sample discussion due to additional data preparation done for these analyses.*

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*)

Age restriction. This exclusion represent a small percentage of the overall stays in skilled nursing facilities and on average exclude approximately less than 6% of the residents. Moreover, the exclusion only result in less than 2% of SNFs having a sample size too small (i.e. less than 30 in the denominator) for the measure to be reported.

While the proportion of stays that are excluded is relatively low, the exclusion was determined a priori to the development and testing of the measure by a panel of clinicians with extensive SNF experience. Thus, first and foremost, the exclusion is needed due to its clinical relevance and for face validity. Moreover, because the exclusion solely excludes a small percentage of the total residents, it does not present additional burden of data analysis for the facilities aiming to use our measure. Therefore, we kept the exclusion in the final measure.

Denominator size restriction. The results show that the sample quickly loses reliability as the sample size decreases below 30 stays in the year. Therefore, we kept this criterion of the minimum sample size of 30 stays for the final measure.

Tracking rate restriction. While the stability of the measure was not obviously different for facilities with low tracking rates, for face validity and, very importantly, to avoid the possibility of perverse MDS coding incentives, we chose a tracking rate of 90% as the minimum for the measure. Therefore, we kept this restriction for the final 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** Click here to enter number of factors **risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

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

Not Applicable

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*)

A panel of clinicians with extensive SNF experience recommended the initial pool of potential risk adjusters. In order to develop a risk adjusted community discharge measure, we conducted a series of logistic regression analyses to estimate the effect of different patient characteristics ranging from medical diagnoses and clinical conditions to functional and demographic characteristics on the likelihood of successful discharge to community.

In developing the risk adjustment models, we performed a number of diagnostic tests to determine the completeness of the various types of admission assessments from which these covariates are derived. Assessments were excluded if they had incomplete data (in general admission and 5 day Medicare assessments had little missing data but other types of assessments were more likely to have missing information).

Additionally, if the only source of covariates was their availability on the discharge record, these cases were also dropped since they were completed on the same day as discharge to the community or death and were therefore not independent of the outcome.

In developing the risk adjusted model using logistic regression analysis, we used a backward elimination selection process with a significance level cutoff of p<.10 to determine which of the dozens of possible diagnoses, conditions, and characteristics available on the MDS contributed statistically to the likelihood of community discharge. This process resulted in 60 risk adjustment variables, which were encoded in 116 variables in the final risk model (including interaction terms, multilevel factor variables, etc.).

2b4.4a. What were the statistical results of the analyses used to select risk factors?
First, presence of pain, pre-admission screening and resident review, functional rehabilitation potential, being obese, and being underweight were not considered due to too many missing values. Table 2b4.4a shows the variables were eliminated from the original covariate pool by using the backward elimination selection process with a significance level cutoff of p<.10. That is to say any covariate from this pool which had a p value >= to 0.1 was removed from the model.

Table 2b4.4a: Covariates removed

Variables	P Value
Hyponatremia	0.9835
Quadriplegia	0.7313
Age 65	0.7184
Balancing (Surface-to-surface transfer)	0.656
Dressing	0.6531
Tuberculosis	0.6537
Traumatic Brain Injury (TBI)	0.5675
Multi-drug Resistant Drug Organism (MDRO)	0.4948
Post-traumatic Stress Disorder (PTSD)	0.4904
Respiratory Failure	0.451
Seizure	0.4454
Cerebral Palsy	0.4181
Obstructive Uropathy	0.4087
Variables	**P Value**
Parkinson's	0.3953
Neurogenic Bladder	0.3745
Multiple Sclerosis	0.3398
Dementia*Age	0.3019
Toilet Use	0.278
Hypotension	0.2503
Hearing	0.2466
Aphasia	0.1452
Age85 and older	0.1378

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)
To develop the risk adjustment model, we considered a selection of sociodemographic risk adjustors include age, sex, language, and marital status, analyzing them in the same way as all of our other risk adjustors. That is, we did not do any separate analyses because they were sociodemographic risk adjustors. In the final model, following the selection process outlined above, we included age and age-squared, sex, and marital status. We had originally excluded race from the measure due to uncertainty about the appropriateness of including it, with respect to SNF admission incentives; we have separately analyzed the effect of adding race to this measure, and found that it has almost no effect on the final risk adjusted discharge to community scores.

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*)
Since logistic regression models seek to optimize the "fit" between the array of independent variables and the probability of the outcome state (i.e. successful community discharge), there is the risk that the model may be "over fit" or may not include certain types of variables that are relatively rare but strongly related to the outcome. To address these issues, we applied a number of standard regression diagnostic techniques. We examined the receiver operator characteristics (ROC) curve of the sensitivity and specificity of the model's prediction accuracy and found it to have an area under the ROC curve of 0.8147. This degree of fit is quite strong. Based on an examination of the standardized residuals and the predicted values of individual admissions in the model, we dropped 77 observations whose standardized residuals had a value of 6 or greater (the 99th percentile was 2.34). Other diagnostics did not reveal any substantial bias in terms of the pattern of residual predictions, suggesting that there were no obvious variables missing from the model.

To measure the effect of adding race to the risk model, we ran the entire calculation first without race indicators (i.e., current specifications), and then with race indicators. We calculated the risk adjusted discharge to community rates applying the denominator exclusions and facility suppression criteria, and then compared each facility's percentile ranking between the two versions of the measure. If the differences were small, then the effect of including or excluding race on the measure is small. The results of this are in the Appendix, section 2b4.5; the correlation between the calculated rates with and without race was 0.9996 and p<0.0001.

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*)**:**
The resulting model demonstrated a "pseudo" R-squared of 0.235 for the stay level. This is to say that about 23.5% of the variation in the outcome was completely explained by the regression model. Furthermore, the c-statistic for this model was 0.820, suggesting a great fit.

The facility level R-square is 55.9%. This means that the percent of facility level variance in rates explained by the risk adjustment model is 55.9%.

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):
Table 2b4.7.a presents the results of the Hosmer-Lemeshow test, which accompanies our calibration plot in Figure 2b4.8. The statistically significant (i.e. P<.0001) finding is the result of having a large sample size, which in our case was 1,965,321 observations. That is, we are highly confident that there is not an absolutely perfect prediction (which would be a perfect line on Figure 2b4.8, rather than a step); this test evaluates whether or not there is a perfect unbiasedness up and down the distribution, but, critically, does not test the *magnitude* of that bias, which is shown to be negligible in Figure 2b4.8 below.

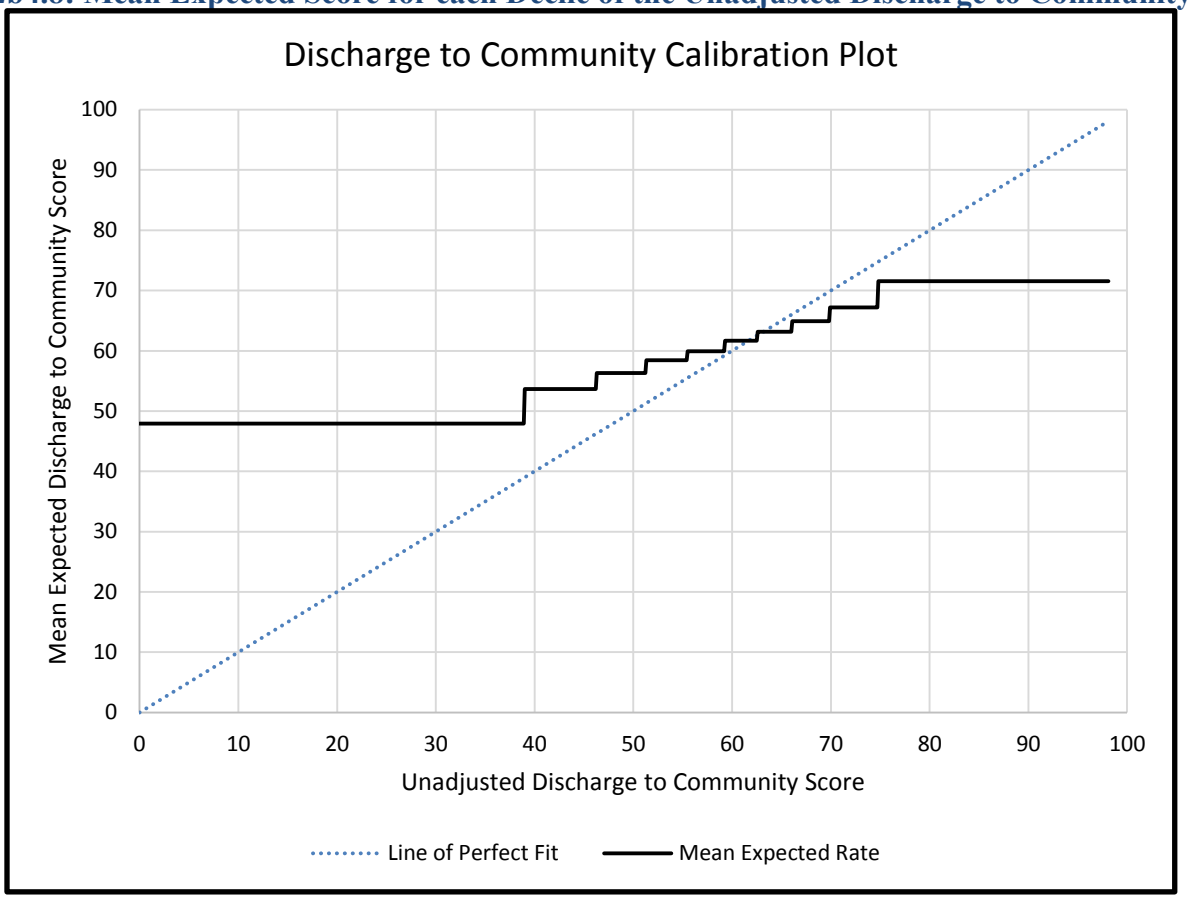
Table 2b4.7.a: Results of the Hosmer and Lemeshow Test

Partition for the Hosmer and Lemeshow Test					
Group	Total	outcome = 1		outcome = 0	
		Observed	Expected	Observed	Expected
1	196533	29181	20135.08	167352	176397.9
2	196532	56209	56288.75	140323	140243.3
3	196532	81819	89116.23	114713	107415.8
4	196533	106943	115252.4	89590	81280.55
5	196533	129338	134681.6	67195	61851.38
6	196533	146128	148586.7	50405	47946.32
7	196536	159421	158795.9	37115	37740.12
8	196529	170083	166745.2	26446	29783.82
9	196533	178516	173576.9	18017	22956.09
10	196527	186699	181158.0	9828	15369.02

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
11734.4060	8	<.0001

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
Figure 2b4.8 presents the mean score predicted by the risk model for each decile of unadjusted discharge to community scores. As shown, the risk model predicts quite well across almost the entire range of actual scores, only deviating by more than about 10 percentage points at the first and tenth deciles. Note we did not apply the denominator size and tracking rate suppression rules to this, but note the calibration plot measures systematic bias across the distribution of rates rather than instability of the sample, and so would not be affected by applying or not applying these suppression criteria.

Figure 2b4.8: Mean Expected Score for each Decile of the Unadjusted Discharge to Community Scores



2b4.9. Results of Risk Stratification Analysis:
Not Applicable

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*)
As aforementioned, the Hosmer and Lemeshow Test of goodness of fit test is detecting the upper and lower bounds of discharge to community scores. This is a result of the large sample size of our database, and does not say anything about the magnitude of this (in our case, we very confidently conclude that there is a negligible bias upward at the low end of the distribution, and a negligible bias downward at the upper end of the distribution). Further, the facility and stay-level R-squared statistics, the receiver operator characteristics (ROC) curve, and the calibration plot each show that discharge to community measure's risk adjustment mode is effective at controlling for variance in discharge to community rates determined by factors outside of the facility's control. Overall, therefore, we conclude that the risk model is an effective and well specified device for controlling for unwanted variance in discharge to community rates.

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*)
Not Applicable

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)*

To determine statistically significant and meaningful performance differences, we used the results from our reliability analyses. First, for comparisons between different facilities, we measured the percent of facilities whose discharge to communities change by various degrees when the facility's stays are resampled from the same underlying population of stays. Second, for comparisons for a single facility over time, we measured the percent of facilities whole risk adjusted rates change by various degrees from one quarter to the next. By looking at the tails of this distribution – which measure directly the statistically significant differences in risk adjusted discharge to community rates at various levels of certainty. For this, we analyzed Table 2a2.3.a, *Changes in Absolute Rates after Re-Sampling*; and Table 2a2.3.c, *Changes in Absolute Rates from Q3-2014 to Q4-2014*.

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*)

Comparisons between different facilities. Table 2a2.3.a shows that after resampling a facility's stays (from the same underlying distribution), 75% of facilities stay within 5 percentage points of the original risk adjusted discharge to community rate, 90% stay within 8 percentage points of the original risk adjusted rate, and 95% stay within 10 percentage points. That is, we are 75% confident that a difference in rates of 5 percentage points is statistically significant, 90% confident that a difference in risk adjusted rates of 8 percentage points is statistically significant, and 95% confident that a difference in risk adjusted rates of 10 percentage points is statistically significant.

Comparisons for a facility between consecutive quarters. Table 2a2.3.c shows that 85% of facilities stay within 5 percentage points of the original risk adjusted discharge to community rate, 90% stay within 6 percentage points of the original risk adjusted rate, and 98% stay within 10 percentage points. That is, we are 85% confident that a difference in rates of 5 percentage points is truly the case, 90% confident that a difference in risk adjusted rates of 6 percentage points is truly the case, and 98% confident that a difference in risk adjusted rates of 10 percentage points is truly the case.

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?*)

Trade standard for defining statistical significance is that 90% confidence or 95% confidence constitutes statistical significance, but much below 90% – more than 10% uncertainty about the conclusion – is insufficiently confident. Therefore, our analysis concludes that an 8 percent difference in risk adjusted discharge to community rates between two different facilities is sufficient to confidently conclude one is better than the other; and a 6 percentage point difference in risk adjusted rates for a given facility between two quarters is sufficient to confidently conclude the facility's performance has improved/gotten worse.

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*)
Not applicable

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*)
Not applicable

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*)
Not applicable

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*)
If the resident had no available risk adjustment data (clinical and demographic characteristic) on any MDS assessment within 18 days of the SNF admission, they were excluded from the measure. However, for the cases when some of the risk adjustment data was missing, we imputed a value for such covariates by first utilizing the facility mean for said covariate. In the event that was also missing for the entire facility, we imputed this value by calculating and utilizing the population mean.

To evaluate the influence of the missing data, we analyzed the proportion of stays for which each risk adjustor needed to be imputed, was imputed using the facility mean, and was imputed using the population mean. If the frequencies of missingness were low, and then if those missing values were imputed dominantly by the facility means, then the imputation approach will generate an unbiased prediction of the facility's discharge to community rate, and we may conclude that missingness is not a problem in our measure.

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*)

Table 2b7.2.a (part 1-3) presents the rates of missingness for each risk adjustor, and presents the percent of stays that were imputed using facility means, and those imputed using population means.

Table 2b7.2.a part 1 of 3: Rates of missingness of risk adjustors

	Percent of Stays where Covariate is Missing			
Risk adjustor	**Missing**	**Imputed by Facility Mean**	**Imputed by Population Mean**	**Not Imputed**
Pressure ulcer	0.12%	0.12%	0.00%	0.00%
ADL times light BIMS	0.21%	0.21%	0.00%	0.00%
ADL times mediate BIMS	0.21%	0.21%	0.00%	0.00%
ADL times severe BIMS	0.21%	0.21%	0.00%	0.00%
Age	0.00%	0.00%	0.00%	0.00%
Age squared	0.00%	0.00%	0.00%	0.00%
Antipsychotics	0.07%	0.07%	0.00%	0.00%
Any acute hospitalization within 30 days of admission	0.00%	0.00%	0.00%	0.00%
Any behavioral codes	0.63%	0.63%	0.00%	0.00%
Any rejection of care	1.01%	1.01%	0.00%	0.00%
Bathing with supervision	0.33%	0.33%	0.00%	0.00%
Bathing with physical help limited to transfer	0.33%	0.33%	0.00%	0.00%
Bathing with physical help in part of bathing activity	0.33%	0.33%	0.00%	0.00%
Bathing with total assistance	0.33%	0.33%	0.00%	0.00%
Bed mobility with supervision	0.07%	0.07%	0.00%	0.00%
Bed mobility with limited assistance	0.07%	0.07%	0.00%	0.00%
Bed mobility with extensive assistance	0.07%	0.07%	0.00%	0.00%
Bed mobility with total assistance	0.07%	0.07%	0.00%	0.00%
Occasionally bowel incontinent	0.26%	0.26%	0.00%	0.00%
Frequently or always bowel incontinent	0.26%	0.26%	0.00%	0.00%
BIMS - light	0.00%	0.00%	0.00%	0.00%
BIMS - mediate	0.00%	0.00%	0.00%	0.00%
BIMS - severe	0.00%	0.00%	0.00%	0.00%
Eating with supervision	0.08%	0.08%	0.00%	0.00%
Eating with limited assistance	0.08%	0.08%	0.00%	0.00%
Eating with extensive assistance	0.08%	0.08%	0.00%	0.00%
Eating with total assistance	0.08%	0.08%	0.00%	0.00%
Balance seat to stand - not steady but able to stabilize without staff	0.73%	0.73%	0.00%	0.00%
Balance seat to stand - not steady only able to stabilize without staff	0.73%	0.73%	0.00%	0.00%
Balance seat to stand - activity did not occur	0.73%	0.73%	0.00%	0.00%
Balance walking - not steady but able to stabilize without staff	0.92%	0.92%	0.00%	0.00%
Balance walking - not steady only able to stabilize without staff	0.92%	0.92%	0.00%	0.00%
Balance walking - activity did not occur	0.92%	0.92%	0.00%	0.00%
Balance turning around - not steady but able to stabilize without staff	1.81%	1.81%	0.00%	0.00%
Balance turning around - not steady only able to stabilize without staff	1.81%	1.81%	0.00%	0.00%
Balance turning around - activity did not occur	1.81%	1.81%	0.00%	0.00%
Balance on/off toilet - not steady but able to stabilize without staff	0.85%	0.85%	0.00%	0.00%
Balance on/off toilet - not steady only able to stabilize without staff	0.85%	0.85%	0.00%	0.00%
Balance on/off toilet - activity did not occur	0.85%	0.85%	0.00%	0.00%

Table 2b7.2.a part 2 of 3: Rates of missingness of risk adjustors

	Percent of Stays where Covariate is Missing			
Risk adjustor	Missing	Imputed by Facility Mean	Imputed by Population Mean	Not Imputed
Personal hygiene with supervision	0.12%	0.12%	0.00%	0.00%
Personal hygiene with limited assistance	0.12%	0.12%	0.00%	0.00%
Personal hygiene with extensive assistance	0.12%	0.12%	0.00%	0.00%
Personal hygiene with total assistance	0.12%	0.12%	0.00%	0.00%
Inattention (CAM)	1.57%	1.57%	0.00%	0.00%
Locomotion on unit with supervision	0.15%	0.15%	0.00%	0.00%
Locomotion on unit with limited assistance	0.15%	0.15%	0.00%	0.00%
Locomotion on unit with extensive assistance	0.15%	0.15%	0.00%	0.00%
Locomotion on unit with total assistance	0.15%	0.15%	0.00%	0.00%
Major Depression	2.38%	2.38%	0.00%	0.00%
Male	0.00%	0.00%	0.00%	0.00%
Married	3.45%	3.45%	0.00%	0.00%
Max Injection	0.05%	0.05%	0.00%	0.00%
I0200 Active Diagnoses: Anemia	0.02%	0.02%	0.00%	0.00%
I0600 Active Diagnoses: Heart Failure (CHF)	0.02%	0.02%	0.00%	0.00%
I0700 Active Diagnoses: Hypertension	0.03%	0.03%	0.00%	0.00%
I2000 Active Diagnoses: Pneumonia	0.02%	0.02%	0.00%	0.00%
I2100 Active Diagnoses: Septicemia	0.01%	0.01%	0.00%	0.00%
I2300 Active Diagnoses: Urinary Tract Infection (UTI)	0.02%	0.02%	0.00%	0.00%
I2400 Active Diagnoses: Viral Hepatitis	0.01%	0.01%	0.00%	0.00%
I2500 Wound Infection	0.01%	0.01%	0.00%	0.00%
I2900 Active Diagnoses: Diabetes Mellitus (DM)	0.02%	0.02%	0.00%	0.00%
I3200 Active Diagnoses: Hyperkalemia	0.01%	0.01%	0.00%	0.00%
I3300 Active Diagnoses: Hyperlipidemia	0.03%	0.03%	0.00%	0.00%
I3900 Active Diagnoses: Hip Fracture	0.02%	0.02%	0.00%	0.00%
I4000 Active Diagnoses: Other Fracture	0.02%	0.02%	0.00%	0.00%
I4200 Active Diagnoses: Alzheimers Disease	0.01%	0.01%	0.00%	0.00%
I4500 Active Diagnoses: Stroke (CVA or TIA or Stroke)	0.02%	0.02%	0.00%	0.00%
I4800 Active Diagnoses: Dementia	0.02%	0.02%	0.00%	0.00%
I4900 Active Diagnoses: Hemiplegia	0.02%	0.02%	0.00%	0.00%
I5000 Active Diagnoses: Paraplegia	0.01%	0.01%	0.00%	0.00%
I5250 Active Diagnoses: Huntingtons	0.01%	0.01%	0.00%	0.00%
I5600 Active Diagnoses: Malnutrition	0.01%	0.01%	0.00%	0.00%
I5700 Active Diagnoses: Anxiety Disorder	0.02%	0.02%	0.00%	0.00%
I5800 Active Diagnoses: Depression	0.02%	0.02%	0.00%	0.00%
I5900 Active Diagnoses: Manic Depression	0.01%	0.01%	0.00%	0.00%
I5950 Active Diagnoses: Psychotic	0.01%	0.01%	0.00%	0.00%
I6000 Active Diagnoses: Schizophrenia	0.01%	0.01%	0.00%	0.00%
I6200 Active Diagnoses: Asthma, COPD, Chronic Lung Disease	0.02%	0.02%	0.00%	0.00%

Table 2b7.2.a part 3 of 3: Rates of missingness of risk adjustors

	Percent of Stays where Covariate is Missing			
Risk adjustor	**Missing**	**Imputed by Facility Mean**	**Imputed by Population Mean**	**Not Imputed**
Shortness of breath or trouble breathing with exertion	0.15%	0.15%	0.00%	0.00%
Shortness of breath or trouble breathing when sitting at rest.	0.14%	0.14%	0.00%	0.00%
Shortness of breath or trouble breathing when lying flat	0.19%	0.19%	0.00%	0.00%
Prognosis	0.29%	0.29%	0.00%	0.00%
Oxygen therapy while a resident	0.15%	0.15%	0.00%	0.00%
Tracheostomy care while a resident	0.15%	0.15%	0.00%	0.00%
Ventilator/respirator care while a resident	0.15%	0.15%	0.00%	0.00%
Dialysis while a resident	0.15%	0.14%	0.00%	0.00%
Hospice care while a resident	0.15%	0.15%	0.00%	0.00%
RUGs: Moderate-high nursing, no therapy	0.00%	0.00%	0.00%	0.00%
RUGs: Moderate-high nursing, no therapy	0.00%	0.00%	0.00%	0.00%
RUG: lower nursing, therapy but have both	0.00%	0.00%	0.00%	0.00%
RUG: Moderate nursing/moderate high therapy (RVA)	0.00%	0.00%	0.00%	0.00%
RUG: Moderate nursing/moderate high therapy (RVB)	0.00%	0.00%	0.00%	0.00%
RUG: Moderate nursing/moderate high therapy (RVC)	0.00%	0.00%	0.00%	0.00%
RUG: Moderate nursing/moderate high therapy (RUA)	0.00%	0.00%	0.00%	0.00%
RUG: moderate nursing, high therapy (RUB)	0.00%	0.00%	0.00%	0.00%
RUG: moderate nursing, high therapy (RUC)	0.00%	0.00%	0.00%	0.00%
RUG: high nursing, low therapy	0.00%	0.00%	0.00%	0.00%
RUG: high nursing, high therapy	0.00%	0.00%	0.00%	0.00%
Makes self understood - usually	0.67%	0.67%	0.00%	0.00%
Makes self understood - sometimes	0.67%	0.67%	0.00%	0.00%
Swallowing disorder	0.27%	0.27%	0.00%	0.00%
Transfers with supervision	0.06%	0.06%	0.00%	0.00%
Transfers with limited assistance	0.06%	0.06%	0.00%	0.00%
Transfers with extensive assistance	0.06%	0.06%	0.00%	0.00%
Transfers with total assistance	0.06%	0.06%	0.00%	0.00%
Usually understands others	0.70%	0.70%	0.00%	0.00%
Sometimes understands others	0.70%	0.70%	0.00%	0.00%
Occasionally urinary incontinent	0.16%	0.16%	0.00%	0.00%
Frequently or always urinary incontinent	0.16%	0.16%	0.00%	0.00%
Vision impaired	1.30%	1.30%	0.00%	0.00%
Vision moderately, highly or severely impaired	1.30%	1.30%	0.00%	0.00%
Walk in corridor with supervision	0.21%	0.21%	0.00%	0.00%
Walk in corridor with limited assistance	0.21%	0.21%	0.00%	0.00%
Walk in corridor with extensive assistance	0.21%	0.21%	0.00%	0.00%
Walk in corridor with total assistance	0.21%	0.21%	0.00%	0.00%
K0300 Weight loss	1.33%	1.32%	0.02%	0.00%

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*)

Nearly all risk adjustors had extremely low rates of missingness, well below 0.1%. A handful had rates of missingness as high as 1.3%, which is still very low. For all covariates, whether their rates of missingness were extremely low (<0.1%) or only very low (around 1%), nearly all cases were imputed using the facility mean.

This means that the risk adjustment model will be unbiasedly predicting the facility's discharge to community rate in virtually all cases. Therefore, the imputation approach used to construct the variables in the risk adjustment model is not biased and performs extremely well in essentially all cases.

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)

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 clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)

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.**

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.

Since this measure relies solely on data items from the MDS 3.0 and all facilities are required to submit this data, this presents no additional burden with regards to data collection and availability of data. Furthermore, collecting and calculating the measure on a quarterly basis but spanning a 12 month period has helped assure fewer facilities with missing rates due to small sample size. This in turn has the effect of preventing large fluctuations from one measurement period to the next due to small sample size. Providers have asked for rates that span solely 1 quarter duration, however under this scenario the number of facilities with the inadequate denominator size of 30 increases, thus affecting the measure stability. However, even with a 12 month window (reported as rolling average each quarter) we still have a number of facilities that cannot have a reported rate or may have a measure one quarter but not another since their total number of admissions from a hospital (i.e. denominator size) is close to the minimum number required for reporting (i.e. 30 admissions).

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)*.

No fees are required for the utilization of the measure specifications, and the measure is not copyright.

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)
	Public Reporting AHCA/NCAL's Research and Data Website, Measure Downloads http://www.ahcancal.org/research_data/quality/Pages/Measure%20Downloads.aspx Quality Improvement with Benchmarking (external benchmarking to multiple organizations) AHCA/NCAL's Research and Data Website, Measure Downloads http://www.ahcancal.org/research_data/quality/Pages/Measure%20Downloads.aspx AHCA/NCAL LTC Trend TrackerSM http://www.ahcancal.org/research_data/trendtracker Quality Improvement (Internal to the specific organization) AHCA/NCAL LTC Trend TrackerSM http://www.ahcancal.org/research_data/trendtracker

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

For Public Reporting:
1a.AHCA/NCAL's Research and Data Website, Measure Downloads
(http://www.ahcancal.org/research_data/quality/Pages/Measure%20Downloads.aspx)
b.Purpose: Provide national data for other stakeholders to use and support Quality Assurance Performance Improvement (QAPI) Programs.
c.We do not track who downloads the data from our website, however are aware that the data are being used by at least two major national data vendors for the LTC sector, , two insurers, and two large ACOs.

Geographic area and number and percentage of accountable entities and patients included: All of the United States, including Puerto Rico and the U.S. Virgin Islands, covering approximately 16,000 SNFs

For Quality Improvement with Benchmarking (external benchmarking to multiple organizations):
1.AHCA/NCAL's Research and Data Website, Measure Downloads
(http://www.ahcancal.org/research_data/quality/Pages/Measure%20Downloads.aspx)
a.As described under the public reporting section.

2.AHCA/NCAL LTC Trend TrackerSM member data profiling and tracking tool (http://www.ahcancal.org/research_data/trendtracker)

a.Name of Program and Sponsor: AHCA/NCAL's Long Term Care Trend Tracker, sponsor: AHCA/NCAL's
b.Purpose: Enables providers to profile their performance on metrics and benchmark against peers, as well as examine ongoing Quality Assurance Performance Improvement (QAPI) Programs
c.Geographic area and number and percentage of accountable entities and patients included: All AHCA/NCAL SNFs in the United States, including Puerto Rico and the U.S. Virgin Islands, covering approximately 10,000 SNFs.

For Quality Improvement (Internal to the specific organization):
1.AHCA/NCAL LTC Trend TrackerSM member data profiling and tracking tool (http://www.ahcancal.org/research_data/trendtracker)
a.As described under internal quality reporting, above.

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?*)
Not applicable.

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.*)
Not applicable.

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**

The national mean risk adjusted discharge to community score has increased from 57.9% in the 4th quarter of 2012, to 59.3% in the 4th quarter of 2013, to 60.0 in the 4th quarter of 2014. This is an increase of 3.6% since the end of 2014.

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.**
Not Applicable

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.**
Not Applicable

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.
No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

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?

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Not Applicable

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.)
Not Applicable

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:** Discharge_to_Community_Appendix_Final.docx

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** American Health Care Association
Co.2 **Point of Contact:** Urvi, Patel, upatel@ahca.org, 202-842-4444-2858
Co.3 **Measure Developer if different from Measure Steward:** American Health Care Association
Co.4 **Point of Contact:** Urvi, Patel, upatel@ahca.org, 202-842-4444-2858

Additional Information

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

in measure development.
The following is a list of members who served on the post-acute care workgroup. This workgroup reviewed the measure specifications and provided advice on how to construct the measure for most of the different steps; numerator, denominator, risk adjustment and exclusions.

Barry Lazarus - HCR ManorCare
Holly Harmon - American Health Care Association
James Muller - American Health Care Association
Barbara Yody - Genesis
Tami Johnson - Kindred
Joanne Wisely - Genesis
Vincent Mor - Brown University
Bill Goulding - Aegis Therapies
Douglas Burr - Health Care Navigator

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2014
Ad.3 **Month and Year of most recent revision:** 07, 2015
Ad.4 **What is your frequency for review/update of this measure?** Annually
Ad.5 **When is the next scheduled review/update for this measure?** 01, 2017

Ad.6 **Copyright statement:** None
Ad.7 **Disclaimers:** None

Ad.8 **Additional Information/Comments:** None

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 #: 2886
De.2. **Measure Title:** Risk-Standardized Acute Admission Rates for Patients with Heart Failure
Co.1.1. **Measure Steward:** Centers for Medicare & Medicaid Services (CMS)
De.3. **Brief Description of Measure:** Rate of risk-standardized acute, unplanned hospital admissions among Medicare Fee-for-Service (FFS) patients 65 years and older with heart failure
1b.1. **Developer Rationale:** The goal of this measure is to evaluate and to improve the quality of care for patients with heart failure cared for by ACOs. These patients account for a significant proportion of Medicare beneficiaries and they experience high morbidity and costs associated with their disease. These patients need efficient, coordinated, and patient-centered care management. They also benefit from provider support and infrastructure that facilitate effective chronic disease management. This measure is focused on hospital admissions for acute illness as the outcome because these admissions are often sentinel events associated with high morbidity as well as physical and emotional stress; they also result in high costs for both the patient and the ACO. Research shows that effective health care can lower the risk of admission for this vulnerable group of patients. For example, efforts to improve coordination and navigation of the healthcare system, along with home-based interventions and exercise-based rehabilitation therapy among patients with heart failure may reduce the risk of hospitalization.

This measure is intended to incentivize ACOs to provide high-quality, coordinated care that focuses on the whole patient. ACOs were conceptualized and created to achieve the goals of improved care, improved population health, and lower cost. Consistent with this mission, we envision that the measure will incentivize providers participating in ACOs to collaborate in order to provide the best system of clinical care and to partner with health and non-health related organizations in their communities as appropriate to improve the health of their patient population.

References:

Centers for Medicare & Medicaid Services (CMS). Medicare Health Support. 2012; https://www.cms.gov/Medicare/Medicare-General-Information/CCIP/. Accessed March 27, 2014.

Brown RS, Peikes D, Peterson G, Schore J, Razafindrakoto CM. Six features of Medicare coordinated care demonstration programs that cut hospital admissions of high-risk patients. Health Affairs. 2012 Jun 2012;31(6):1156-1166.

McCarthy D, Cohen A, Johnson MB. Gaining Ground: Care Management Programs to Reduce Hospital Admissions and Readmissions Among Chronically Ill and Vulnerable Patients. The Commonwealth Fund, New York. 2013.

Patient Protection and Affordable Care Act, 42 U.S.C., §3022 (2010).

Zhang NJ, Wan TT, Rossiter LF, Murawski MM, Patel UB. Evaluation of chronic disease management on outcomes and cost of care for Medicaid beneficiaries. Health policy (Amsterdam, Netherlands). May 2008;86(2-3):345-354.

Inglis SC, Pearson S, Treen S, Gallasch T, Horowitz JD, Stewart S. Extending the horizon in chronic heart failure: effects of multidisciplinary, home-based intervention relative to usual care. Circulation. Dec 5 2006;114(23):2466-2473.

Austin J, Williams WR, Ross L, Hutchison S. Five-year follow-up findings from a randomized controlled trial of cardiac rehabilitation for heart failure. European journal of cardiovascular prevention and rehabilitation: official journal of the European Society of Cardiology, Working Groups on Epidemiology & Prevention and Cardiac Rehabilitation and Exercise Physiology. Apr 2008;15(2):162-

167.

Taylor RS, Sagar VA, Davies EJ, et al. Exercise-based rehabilitation for heart failure. The Cochrane database of systematic reviews. 2014;4:Cd003331.

S.4. **Numerator Statement:** The outcome measured for each patient is the number of acute, unplanned admissions per 100 person-years at risk for admission. Persons are considered at risk for admission if they are alive, enrolled in FFS Medicare, and not currently admitted. (See S.6, Numerator Details, for more information.)
S.7. **Denominator Statement:** The target population is ambulatory Medicare FFS patients aged 65 years and older with a diagnosis of heart failure.
S.10. **Denominator Exclusions:** The measure excludes:
1. Patients without continuous enrollment in Medicare Part A for the duration of the measurement period (or until death).
Rationale: We exclude these patients to ensure full data availability for outcome assessment (Part A during the measurement year).
2. Patients with left ventricular assist devices (LVADs).
Rationale: We exclude these patients because while they have a high risk of admission, they are low in prevalence and are clustered among a few ACOs.

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Administrative claims
S.26. **Level of Analysis:** Integrated Delivery System

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

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?** Not applicable.

New Measure -- Preliminary Analysis

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.

- The developer notes that through efficient, coordinated, and patient-centered care management, along with provider support and infrastructure that facilitate effective chronic disease management, ACOs can improve the quality of care for patients with heart failure.

Question for the Committee:

- *Did the developer provide at least one health care structure or process that an ACO can undertake to improve this outcome?*

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 developer provided data from ACO performance score using the 2012 Medicare Full Sample which showed

the crude US national Medicare FFS rate of acute, unplanned admissions per person-year among patients with heart failure was 85.5 per 100 person-years.

- Among ACOs, the mean RSAAR for calendar year 2012 was 81.9 per 100 person-years (standard deviation = 11.6). The median RSAAR was 81.5 per 100 person-years (interquartile range [IQR] 73.6 to 88.8). The minimum RSAAR score was 53.7 per 100 person-years; the 5th percentile was 64.6 per 100 person-years; the 95th percentile was 101.7 per 100 person-years; and maximum score was 120.7 per 100 person-years.
- They observed that 61 ACOs (53.5%) had RSAARs that were 'no different than the national rate' (of all Medicare FFS beneficiaries with heart failure). An additional 37 ACOs (32.5%) had 'better than the national rate' RSAAR scores and 16 (14.0%) were 'worse than the national rate.

Disparities

- The developer reports that they examined disparities in ACO performance based on the proportion of patients of low socioeconomic status (SES) being cared for by each ACO.
- The developer found that performance scores did not change appreciably after adjusting the models for patients' SES. The Spearman correlation comparing the ACO measure scores estimated with and without risk adjustment for the AHRQ SES Index was 0.990. Similarly, the Spearman correlation for the scores estimated with and without patients' Medicaid dual eligibility was 0.991. These results demonstrate that adjusting for SES at the patient level has little effect on the measure score.
- Overall, results indicate that SES status plays little role at the patient level, thus measure was not adjusted for patient-level SES. According to the developer, ACOs should and do influence a broad range of patient and community-level factors that can mitigate the risk of admission associated with low SES, and do not want to adjust for modifiable factors.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Given the developer disparities testing results, does the Committee agree that SDS adjustment is not warranted?*

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)

1. Importance to Measure and Report

1a. Evidence to Support Measure Focus

Comments: **Yes - preventing admissions as a quality measure actually makes more sense than preventing readmissions for an organization like an ACO.

**Measure is PRO and used lit review as well as environmental scan. Hopes to show influence of ACO can affect patient outcome

1b. Performance Gap

Comments: **Variability on this measure is greater than variability on the readmission measures, which likely suggests more opportunity for improvement. In terms of disparities, data suggest that differences in outcomes by neighborhood and individual SES are present and significant.

**Seeks to demonstrate variations in unplanned readmissions and care delivered by ACOs. Assumes rates are driven by ACO willingness (or incentives) to provide home-based interventions and mitigate SES factors.

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

Comments: **n/a

**NA, Outcome measure

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

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): Administrative claims

Specifications:

- This measure calculates the rate of risk-standardized acute, unplanned hospital admissions among Medicare Fee-for-Service (FFS) patients 65 years and older with heart failure
- This is a health outcome measure and the level of analysis is Integrated Delivery System.
- The numerator is the number of acute, unplanned admissions per 100 person-years at risk for admission. Persons are considered at risk for admission if they are alive, enrolled in FFS Medicare, and not currently admitted.
- The denominator is ambulatory Medicare FFS patients aged 65 years and older with a diagnosis of heart failure.

Questions for the Committee :

- *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

- Datasets used for testing included Medicare Parts A and B claims, the denominator file, the Medicare Provider Analysis and Review (MedPAR) file, and the American Community Survey to derive the AHRQ SES index.
- **Data element reliability**:
  - With regard to data element reliability, the developer notes that the measure has been developed to avoid the use of claims data elements that are thought to be coded inconsistently across hospitals or providers, instead using fields that are consequential for payment and which are audited by CMS.
  - In addition, the developer compared frequencies and odds ratios of variables from their risk model to assess the consistency of those variables across samples.
- **Performance score reliability:**
  - The developer defines performance score reliability as the degree to which repeated measurements of the same entity agree with each other.
  - In line with this thinking, the developer's approach to assessing score-level reliability was to consider the extent to which assessments of a hospital using different but randomly-selected subsets of patients produce similar measures of hospital performance. The developers refer to this as a "test-retest" approach; it may also be called a "split-half" method. This is generally considered an appropriate method of testing reliability.

Results of reliability testing

- **Data element reliability**:
  - Summarizing the results of this analysis, the developer notes that the mean age and frequency of risk-adjustment variables was similar among the two samples of 2012 data suggesting that the data elements are reliable across the samples.

- **Performance score reliability:**
  - The 2012 full Medicare sample was divided into two subsets of patients randomly. The developer calculated the measure score of all ACOs for each of the two subsets of patients. Each ACO was measured twice, but each measurement was make using distinct sets of measures. The interclass correlation coefficient (ICC) for the two subsets of patients was 0.81, which can be interpreted as excellent correlation, and thus reliable.

Guidance from the Reliability Algorithm

- Question 1. Submitted specifications are precise, unambiguous, and complete. Measure can be consistently implemented.
- Question 2. Empirical reliability testing was conducted using statistical tests with the measure as specified.
- Question 3. Empirical validity testing of patient-level data was conducted.
- Question 4. Reliability testing was conducted with computed performance measure scores for each measured entity.
- Question 5. Random split-half correlation was used to assess the proportion of variability due to real differences among the measured entities.
- Question 6. The ICC was 0.81 which is considered an excellent level of agreement.

Questions for the Committee:

- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*
- *Does the measure testing match the measure specifications?*

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

2b. Validity

2b1. Validity: Specifications

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

- This measure estimates the predicted number of admissions given the Accountable Care Organization's (ACO's) case mix, sample size, and actual admission rate. The outcome for this measure is the number of acute, unplanned admissions per 100 person-years at risk for admission. The outcome includes inpatient admissions to an acute care hospital for any cause during the measurement year, unless an admission is identified as "planned."

Question for the Committee:

- *Are the specifications clear?*

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

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.

- The developer tested the validity of the measure using three different methods:
  - Validity of the claims-based measures. The developer argues that other NQF endorsed mortality and readmission measures have been validated by comparing the claims to the medical records data elements. It is unclear if the risk adjustment validation approach that the developer cites is sufficiently similar to this measure and for this level of analysis and ambulatory patients.
  - The developer also notes that this measure has been validated by using established measure development guidelines. While an important step for measure development, this method of validity testing has generally not be considered sufficient for demonstrating measure validity.
  - Finally, the measure developer completed a systemic face validity assessment of this measure with 8

experts agreeing that this measure was a valid indicator of health care quality.

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

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:

- Out of the total Medicare FFS patients with heart failure (N=2,649,829), the developer excluded 66,909 due to non-continuous enrollment in part A in 2012, and also excluded 1,048 patients with left ventricular assist devices (LVAD).

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *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 ☒ No

Risk adjustment summary

- The developers provided a conceptual framework that was used to develop the risk adjustment model for this measure. This conceptual framework included 4 contextual domains that influence ACO performance including, physical environment, community resources, patient resources, and patient behavioral/personal preferences.
- The measure included demographic factors, and clinical risk factors present at the start of the measurement period.
  - The measure developers reviewed 189 diagnosis groups included in the hierarchical condition category (HCC), and calculated the prevalence of each CC in the year preceding the measurement period. After examining the bi-variate analysis, the developers reduced the list to 22 candidate variables including age.
  - The measure developers did not adjust for contextual factors that impact admissions; however, they did provide data demonstrating that including SDS adjustment did not make a meaningful difference to the measure score of the ACOs. The spearman correlation coefficient that estimated the difference in performance with and without SDS adjustment was 0.990. Thus, the results demonstrate that adjustment had little effect on the measure score.
- **Risk Model Diagnostics:**
  - To assess the overall performance of their risk-adjustment model, the developers computed two summary statistics, including:
    - *Risk model discrimination statistics* (the model's ability to explain how successful the fit is in explaining the variation of the data. In this case, the r-sq value was 0.123. In other words, the model was able to explain 12.3% of the total deviance.
    - Overfitting indices (model calibration) [presented as (γ0, γ1)]:
      - The developer states that if the γ0 in the validation samples are substantially far from zero and the γ1 is substantially far from one, there is potential evidence of over-fitting. The calibration value of close to 0 at one end and close to 1 to the other end indicates good calibration of the model.
        - 2012 Development Sample (Index): (**0,1**)
        - 2012 Validation Sample: **(-0.0020, 1.0002)**

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 note that the methodology to publicly report this measure has not been determined yet
- For other publically reported measures with the same methodology, CMS categories hospitals at "better than the national rate", "worse than the national rate" and "no different than the national rate".
  - For this measure, 61 ACOs (54%) performed no different than the national rate, 37 (32%) performed better then the national rate, and 16 (14%) performed worse than the national rate. The developers suggest that this demonstrates that there is a meaningful different in performance on this measure.

Question for the Committee:

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

2b6. Comparability of data sources/methods:

- While the developer did not decide to include SDS variables in their final model, they did compare measure results with and without SDS adjustment.

2b7. Missing Data

- N/A

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)

2. Scientific Acceptability of Measure Properties

2a1. & 2b1. Specifications

Comments: **Specifications appear valid.

**None

2a2. Reliability Testing

Comments: **Reliability seems good, which is actually surprising given the sample size issues inherent in working with ACO data rather than Medicare-wide data; question would be as smaller (rural, physician group, etc,) ACOs come on board whether cutoffs would need to be specified that would limit applicability.

**Excellent

2b2. Validity Testing

Comments: **See above. Measure seems valid assuming prior testing of validity of diagnosis codes for identifying heart failure admissions is acceptable.

**TEP showed agreement but not strong agreement

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **no

**Race may be a factor

Criterion 3. Feasibility

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.

The developer states:

- ALL measure data elements are in defined fields in electronic claims and routinely generated or collected by and used by healthcare personnel during the provision of care, coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims).
- There is no cost associated with data collection.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Is the data collection strategy ready to be put into operational use?*

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3. Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **This measure is highly feasible assuming new Medicare entrants are excluded (won't have a year preceding admission to accrue risk).

**None

Criterion 4: Usability and Use

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**

OR

Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details [Accountability program(s) – details]

The developer states:

- This measure was included by CMS in the November 2014 Physician Fee Schedule final rule, and finalized adding the measure to the Medicare Shared Savings Program quality measure set (see 79 FR 67912; https://www.gpo.gov/fdsys/pkg/FR-2014-11-13/pdf/2014-26183.pdf).
- The measure is planned for pay-for-reporting in the Medicare Shared Savings Program for 2015 and 2016 reporting periods (79 FR 67912, 67916) and for pay-for-performance in the Medicare Shared Savings Program beginning 2017 reporting period (79 FR 67912, 67916).

Improvement results N/A

Potential harms

The developer states:

- To minimize the potential of this measure to result in the denial of future care to high-risk individuals, they developed the patient cohort exclusions and risk-adjustment model to ensure providers who care for patients at higher risk of admission will not be disadvantaged in the measure. CMS is committed to monitoring this measure's use and assessing potential unintended consequences over time.

***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

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **Usable and an appropriate measure for ACOs. Again since no response box present for many questions above, noting here the low R-squared (though I don't really have a prior on what it should be in this population). As noted for other measures, the discussion around SES is really a philosophical one and in this case the measure developer did not feel it was appropriate to adjust for neighborhood or individual SES. Finally, the overlap with the AHRQ PQIs bears discussing.

**May results in increased penalties for ACO with safety net populations or rural populations

Criterion 5: Related and Competing Measures

Related or competing measures

- 0277 : Heart Failure Admission Rate (PQI 8)
- 0709 : Proportion of patients with a chronic condition that have a potentially avoidable complication during a calendar year.

Harmonization

- The measures listed above have different cohort populations and risk-adjustment models. NQF #0709 is not risk-adjusted; NQF #0277 is risk-adjusted for age and sex only, while this measure is fully risk-adjusted. The outcomes measured (NQF 0709: potentially avoidable complications; NQF 0277: heart failure admissions) are different from this measure's outcome of acute, all-cause admission rates.

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: Risk-Standardized Acute Admission Rates for Patients with Heart Failure

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: 1/29/2016

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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: Click here to name the health outcome

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

Patients with heart failure are vulnerable to complications that result from their underlying disease, as well as to a range of other acute illnesses, placing them at relatively high risk for hospitalization. Provision of coordinated care that is focused on improving health for the whole patient, across all stages of disease, and in the context of coexisting comorbidities and life circumstances should lower the risk of hospital admission for these patients.

To provide high-quality care for patients with chronic conditions, health systems must effectively prevent and manage the complications of chronic disease as well as other related and unrelated illnesses that frequently arise among patients with chronic disease. For more than a decade we have known that admission rates vary across the country, even after adjusting for differences in patient populations. To date, however, admission rates have been used as quality and accountability measures to only a limited degree. For example, it is only recently that the Centers for Medicare & Medicaid Services (CMS) has started to use admission scores developed by the Agency for Healthcare Research and Quality (AHRQ), known as Prevention Quality Indicators (PQIs), in several of its ambulatory programs. These admission scores, however, are narrowly focused and measure only disease-specific admissions among populations defined by the disease (for example, heart failure admissions among patients with heart failure). They do not capture the wide spectrum of hospital admissions for which patients with chronic conditions are at increased risk.

This measure of acute, unplanned admission rates among patients with heart failure will illuminate differences in the quality of care delivered by ACOs, and drive efforts to improve prevention and management strategies, including the efficiency and coordination of care.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

Research shows that effective health care can lower the risk of admission for patients with heart failure [1-4]. For example, efforts to improve coordination and navigation of the healthcare system, along with home-based interventions and exercise-based rehabilitation therapy among patients with heart failure, may reduce the risk of hospitalization [1, 5-8].

It is our vision that this measure will illuminate variation among ACOs in hospital admission rates and incentivize ACOs to develop efficient and coordinated chronic disease management strategies that anticipate and respond to patients' needs and preferences. This vision is consistent with ACOs' commitment to deliver patient-centered care that fulfills the goals of the Department of Health and Human Service's Triple Aim – improving population health, improving care, and lowering care costs.

Citations:

1. Patient Protection and Affordable Care Act, 42 U.S.C., §3022 (2010).

2. Centers for Medicare & Medicaid Services (CMS). Medicare Health Support. 2012; https://www.cms.gov/Medicare/Medicare-General-Information/CCIP/. Accessed March 27, 2014.

3. Brown RS, Peikes D, Peterson G, Schore J, Razafindrakoto CM. Six features of Medicare coordinated care demonstration programs that cut hospital admissions of high-risk patients. *Health Affairs.* 2012 Jun 2012;31(6):1156-1166.

4. McCarthy D, Cohen A, Johnson MB. Gaining Ground: Care Management Programs to Reduce Hospital Admissions and Readmissions Among Chronically Ill and Vulnerable Patients. *The Commonwealth Fund, New York.* 2013.

5. Zhang NJ, Wan TT, Rossiter LF, Murawski MM, Patel UB. Evaluation of chronic disease management on outcomes and cost of care for Medicaid beneficiaries. *Health policy (Amsterdam, Netherlands).* May 2008;86(2-3):345-354.

6. Inglis SC, Pearson S, Treen S, Gallasch T, Horowitz JD, Stewart S. Extending the horizon in chronic heart failure: effects of multidisciplinary, home-based intervention relative to usual care. *Circulation.* Dec 5 2006;114(23):2466-2473.

7. Austin J, Williams WR, Ross L, Hutchison S. Five-year follow-up findings from a randomized controlled trial of cardiac rehabilitation for heart failure. European journal of cardiovascular prevention and rehabilitation : official

journal of the European Society of Cardiology, Working Groups on Epidemiology & Prevention and Cardiac Rehabilitation and Exercise Physiology. Apr 2008;15(2):162-167.

8. Taylor RS, Sagar VA, Davies EJ, et al. Exercise-based rehabilitation for heart failure. The Cochrane database of systematic reviews. 2014;4:Cd003331.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

Not applicable. This is an outcome measure.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Not applicable. This is an outcome measure.

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Not applicable. This is an outcome measure.

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Not applicable. This is an outcome measure.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

Not applicable. This is an outcome measure.

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

Not applicable. This is an outcome measure.

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

Not applicable. This is an outcome measure.

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

Not applicable. This is an outcome measure.

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

Not applicable. This is an outcome measure.

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

Not applicable. This is an outcome measure.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

Not applicable. This is an outcome measure.

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

Not applicable. This is an outcome measure.

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Not applicable. This is an outcome measure.

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

Not applicable. This is an outcome measure.

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Not applicable. This is an outcome measure.

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

Not applicable. This is an outcome measure.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

Not applicable. This is an outcome measure.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

Not applicable. This is an outcome measure.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

Not applicable. This is an outcome measure.

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

Not applicable. This is an outcome measure.

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Not applicable. This is an outcome measure.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

Not applicable. This is an outcome measure.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

Not applicable. This is an outcome measure.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

Not applicable. This is an outcome measure.

1a.8 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.

Not applicable. This is an outcome measure.

1a.8.1 What process was used to identify the evidence?

Not applicable. This is an outcome measure.

1a.8.2. Provide the citation and summary for each piece of evidence.

Not applicable. This is an outcome measure.

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
Heart_Failure_ACO_Admission_Measure_NQF_Evidence_Form_01-29-16_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 evaluate and to improve the quality of care for patients with heart failure cared for by ACOs. These patients account for a significant proportion of Medicare beneficiaries and they experience high morbidity and costs associated with their disease. These patients need efficient, coordinated, and patient-centered care management. They also benefit from provider support and infrastructure that facilitate effective chronic disease management. This measure is focused on hospital admissions for acute illness as the outcome because these admissions are often sentinel events associated with high morbidity as well as physical and emotional stress; they also result in high costs for both the patient and the ACO. Research shows that effective health care can lower the risk of admission for this vulnerable group of patients. For example, efforts to improve coordination and navigation of the healthcare system, along with home-based interventions and exercise-based rehabilitation therapy among patients with heart failure may reduce the risk of hospitalization.

This measure is intended to incentivize ACOs to provide high-quality, coordinated care that focuses on the whole patient. ACOs were conceptualized and created to achieve the goals of improved care, improved population health, and lower cost. Consistent with this mission, we envision that the measure will incentivize providers participating in ACOs to collaborate in order to provide the best system of clinical care and to partner with health and non-health related organizations in their communities as appropriate to improve the health of their patient population.

References:

Centers for Medicare & Medicaid Services (CMS). Medicare Health Support. 2012; https://www.cms.gov/Medicare/Medicare-General-Information/CCIP/. Accessed March 27, 2014.

Brown RS, Peikes D, Peterson G, Schore J, Razafindrakoto CM. Six features of Medicare coordinated care demonstration programs that cut hospital admissions of high-risk patients. Health Affairs. 2012 Jun 2012;31(6):1156-1166.

McCarthy D, Cohen A, Johnson MB. Gaining Ground: Care Management Programs to Reduce Hospital Admissions and Readmissions Among Chronically Ill and Vulnerable Patients. The Commonwealth Fund, New York. 2013.

Patient Protection and Affordable Care Act, 42 U.S.C., §3022 (2010).

Zhang NJ, Wan TT, Rossiter LF, Murawski MM, Patel UB. Evaluation of chronic disease management on outcomes and cost of care for Medicaid beneficiaries. Health policy (Amsterdam, Netherlands). May 2008;86(2-3):345-354.

Inglis SC, Pearson S, Treen S, Gallasch T, Horowitz JD, Stewart S. Extending the horizon in chronic heart failure: effects of multidisciplinary, home-based intervention relative to usual care. Circulation. Dec 5 2006;114(23):2466-2473.

Austin J, Williams WR, Ross L, Hutchison S. Five-year follow-up findings from a randomized controlled trial of cardiac rehabilitation for heart failure. European journal of cardiovascular prevention and rehabilitation: official journal of the European Society of Cardiology, Working Groups on Epidemiology & Prevention and Cardiac Rehabilitation and Exercise Physiology. Apr 2008;15(2):162-

167.

Taylor RS, Sagar VA, Davies EJ, et al. Exercise-based rehabilitation for heart failure. The Cochrane database of systematic reviews. 2014;4:Cd003331.

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.*
We report the variation in ACO performance score using the 2012 Medicare Full Sample.
There were 2,581,892 patients in the 2012 Medicare Full Sample who met our inclusion and exclusion criteria for the measure cohort. Among these, there were 123,626 patients in 114 ACOs.

The crude US national Medicare FFS rate of acute, unplanned admissions per person-year among patients with heart failure was 85.5 per 100 person-years.

Among ACOs, the mean RSAAR for calendar year 2012 was 81.9 per 100 person-years (standard deviation = 11.6). The median RSAAR was 81.5 per 100 person-years (interquartile range [IQR] 73.6 to 88.8). The minimum RSAAR score was 53.7 per 100 person-years; the 5th percentile was 64.6 per 100 person-years; the 95th percentile was 101.7 per 100 person-years; and maximum score was 120.7 per 100 person-years.

We observed that 61 ACOs (53.5%) had RSAARs that were 'no different than the national rate' (of all Medicare FFS beneficiaries with heart failure). An additional 37 ACOs (32.5%) had 'better than the national rate' RSAAR scores and 16 (14.0%) were 'worse than the national rate.'

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.
Not applicable.

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 examined disparities in ACO performance based on the proportion of patients of low socioeconomic status (SES) being cared for by each ACO.

Identification of ACOs caring for few and many 'low SES' patients
We identified low SES patients using two variables: the Agency for Healthcare Research and Quality (AHRQ) SES Index and patient Medicare and Medicaid dual-eligibility status.

Using the AHRQ SES Index (described in the NQF Testing form, Section 2b4.3 and Appendix E of the attached technical report), which is a continuous variable, we created a dichotomous low-SES variable by assessing the distribution of SES scores across a broad sample of Medicare FFS beneficiaries and labeling patients with the lowest 20% of scores as "low SES" (see Testing Form, Section 1.8, for further details). We then categorized ACOs into quartiles based on the proportion of low SES patients in their cohort (first quartile (Q1) = 'few' low SES patients, fourth quartile (Q4) = 'many' low SES patients).

Similarly, we categorized ACOs by the proportion of Medicaid dual-eligible patients in their cohort into ACOs caring for 'few' (Q1) and 'many' (Q4) Medicaid dual-eligible patients.

Results:
AHRQ SES Index Analyses and Medicaid Dual-Eligibility Analyses
Using the AHRQ SES Index, for the 29 ACOs in Q1, the proportion of low SES patients ranged from 0 to 3.9%; for the 28 ACOs in the fourth quartile, the proportion of low SES patients ranged from 27.3% to 97.1%.

Among the 29 ACOs caring for few low SES patients (Q1), 2 (6.9%) performed worse than the national rate, 17 (58.6%) performed 'no

different than the national rate,' and 10 (34.5%) performed 'better than the national rate.' Among the 28 ACOs caring for many low SES patients (Q4), 7 (25.0%) performed 'worse than the national rate,' 16 (57.1%) performed 'no different than the national rate,' and 5 (17.9%) performed 'better than the national rate.' (See attached Technical Report, Table 10).

Using Medicaid dual eligibility as an indicator of low SES, among the 29 ACOs caring for few Medicaid dual-eligible patients (Q1), the proportion of Medicaid dual-eligible patients ranged from 3.2% to 9.7%; among the 29 ACOs caring for the most Medicaid dual-eligible patients (Q4) the proportion of Medicaid dual-eligible patients ranged from 22.5% to 70.9%.

Among the 29 ACOs with few Medicaid dual-eligible patients (Q1), 1 (3.4%) performed worse than the national rate, 14 (48.3%) performed no different than the national rate, and 14 (48.3%) performed better than the national rate. Among the 29 ACOs with many Medicaid dual-eligible patients (Q4), 7 (24.1%) performed 'worse than the national rate,' 17 (58.6%) performed 'no different than the national rate,' and 5 (17.2%) performed 'better than the national rate.' (See attached Technical Report, Table 10).

The distribution of RSAARs across ACOs caring for increasing proportions of low SES patients reveals two patterns: (1) ACOs in Q1 (few low SES patients) tend to have lower RSAARs than ACOs in Q4 (many low SES patients); (2) there is more variation in RSAARs among ACOs in Q4 as compared with ACOs in Q1-Q3. There are small differences in these patterns when analyses are performed using Medicaid dual-eligibility as an indicator of SES status (see Figure 17 of the attached technical report).
Socioeconomic Status Interpretation
Among a group of 114 ACOs, there is substantial variation in performance among ACOs caring for many (Q4) and few (Q1) low SES patients. ACOs serving many low SES patients more often perform worse than the national rate compared with ACOs serving few low SES patients. This was true using either the AHRQ SES index (25.0% vs. 6.9%, respectively) or Medicaid dual-eligibility status (24.1% vs. 3.4%, respectively) as an indicator of patients' SES. However, among ACOs serving many low SES patients, using the AHRQ SES index, 17.9% performed 'better than the national rate;' using Medicaid dual-eligibility status, 17.2% performed 'better than the national rate.'

We also found that performance scores did not change appreciably after adjusting the models for patients' SES. As demonstrated in the Testing Form, Section 2b4.11, the Spearman correlation comparing the ACO measure scores estimated with and without risk adjustment for the AHRQ SES Index was 0.990. Similarly, the Spearman correlation for the scores estimated with and without patients' Medicaid dual eligibility was 0.991. These results demonstrate that adjusting for SES at the patient level has little effect on the measure score.

We did not adjust the measure for patient-level SES. Conceptually, ACOs should and do influence a broad range of patient and community-level factors that can mitigate the risk of admission associated with low SES, and we do not want to adjust for modifiable factors. Empirically, our results indicate that SES status plays little role at the patient level.

References:
Wynn B. Analysis of the Joint Distribution of Disproportionate Share Hospital Payments. 2002.

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

Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian SV, Carson R. Choosing area based socioeconomic measures to monitor social inequalities in low birth weight and childhood lead poisoning: The Public Health Disparities Geocoding Project (US). J Epidemiol Community Health. 2003a Mar;57(3):186-99

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.**
Not applicable. Data on disparities are presented above.

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, 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.**
Heart failure is a complex, high-prevalence chronic disease that affects 14% of Medicare beneficiaries. Heart failure impacts people's functional status as well as their daily living. It is also a high-cost disease with Medicare heart failure beneficiaries accounting for 43% of total Medicare spending [1]. Patients with heart failure are vulnerable to complications that result from their underlying disease, as well as to a range of other acute illnesses, placing them at relatively high risk for hospitalization [2-3]. Provision of coordinated care that is focused on improving health for the whole patient, across all stages of disease, and in the context of coexisting comorbidities and life circumstances should lower the risk of hospital admission for these patients [2-12].
Specific to the heart failure cohort assessed for this measure, in the 2012 Medicare Full Sample, there were 2,581,892 patients who met criteria for heart failure, among which 123,626 (4.8%) were assigned to one of 114 ACOs. In these groups, the rate of acute, unplanned hospital admissions was 85.5 per 100 person-years among all Medicare FFS heart failure beneficiaries, and 83.2 per 100 person-years among heart failure beneficiaries assigned to an ACO. The average risk-standardized acute admission rate among ACOs was 81.9 (range of 53.7 to 120.7) per 100 person-years at risk for hospitalization. These rates illustrate the high morbidity associated with this condition, the variation in ACO performance, and the opportunity to reduce hospitalizations, improve care, and potentially lower costs.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
1. Dall TM, Blanchard TD, Gallo PD, Semilla AP. The economic impact of Medicare Part D on congestive heart failure. The American journal of managed care. May 2013;19(6 Suppl):s97-100.

2. Brown RS, Peikes D, Peterson G, Schore J, Razafindrakoto CM. Six features of medicare coordinated care demonstration programs that cut hospital admissions of high-risk patients. Health Affairs. 2012;31:1156-11662.

3. Levine S, Steinman BA, Attaway K, Jung T, Enguidanos S. Home care program for patients at high risk of hospitalization. American Journal of Managed Care. 2012;18:e269-276

4. Zhang NJ, Wan TT, Rossiter LF, Murawski MM, Patel UB. Evaluation of chronic disease management on outcomes and cost of care for medicaid beneficiaries. Health policy (Amsterdam, Netherlands). 2008;86:345-354

5. Sommers LS, Marton KI, Barbaccia JC, Randolph J. Physician, nurse, and social worker collaboration in primary care for chronically ill seniors. Archives of internal medicine. 2000;160:1825-1833

6. Dorr DA, Wilcox AB, Brunker CP, Burdon RE, Donnelly SM. The effect of technology-supported, multidisease care management on the mortality and hospitalization of seniors. Journal of the American Geriatrics Society. 2008;56:2195-2202

7. Chan CL, You HJ, Huang HT, Ting HW. Using an integrated coc index and multilevel measurements to verify the care outcome of patients with multiple chronic conditions. BMC health services research. 2012;12:405

8. Littleford A, Kralik D. Making a difference through integrated community care for older people. Journal of Nursing and Healthcare of Chronic Illness. 2010;2:178-186

9. Centers for Medicare & Medicaid Services (CMS). Medicare health support. 2012

10. RTI International, Telligen. Accountable care organization 2013 program analysis: Quality performance standards narrative measure specifications. 2012

11. McCarthy D, Cohen A, Johnson MB. Gaining ground: Care management programs to reduce hospital

admissions and readmissions among chronically ill and vulnerable patients. The Commonwealth Fund, New York. 2013

12. Friedberg MW, Rosenthal MB, Werner RM, Volpp KG, Schneider EC. Effects of a medical home and shared savings intervention on quality and utilization of care. JAMA Intern Med 2015 Aug; 175(8):1362-8.

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.)*
Not applicable.

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 : Congestive Heart Failure

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Health and Functional Status, Health and Functional Status : Development/Wellness, Health and Functional Status : Functional Status, Safety, Safety : Complications

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:** Heart_Failure_ACO_Admission_Measure_NQF_Data_Dictionary_01-29-16_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.
Not applicable.

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 measured for each patient is the number of acute, unplanned admissions per 100 person-years at risk for admission. Persons are considered at risk for admission if they are alive, enrolled in FFS Medicare, and not currently admitted. (See S.6, Numerator Details, for more information.)

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.)
The measure requires three years of data.

Outcome time window: We observe for the outcome of admission for one full calendar year.

Time period for cohort identification: The cohort is identified using two years of claims data prior to the measurement year.

Risk-adjustment look-back period: Risk-adjustment variables are identified using one year of data prior to the measurement year.

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.
Note: The numerator of the measure score is the predicted number of admissions given the Accountable Care Organization's (ACO's) case mix, sample size, and actual admission rate. We use this field to define the outcome.

Outcome Definition:
The outcome for this measure is the number of acute, unplanned admissions per 100 person-years at risk for admission. The outcome includes inpatient admissions to an acute care hospital for any cause during the measurement year, unless an admission is identified as "planned."

Identification of Planned Admissions:
The measure outcome includes only unplanned admissions. Although clinical experts agree that proper care in the ambulatory setting should reduce hospital admissions, variation in planned admissions (such as for elective surgery) does not typically reflect quality differences. We based the planned admission algorithm on the Centers for Medicare & Medicaid Services (CMS) Planned Readmission Algorithm Version 3.0, which CMS originally created to identify planned readmissions for the hospital-wide readmission measure. In brief, the algorithm identifies a short list of always planned admissions (i.e., those where the principal discharge diagnosis is major organ transplant, obstetrical delivery, or maintenance chemotherapy) as well as those admissions with a potentially planned procedure (e.g., total hip replacement or cholecystectomy) AND a non-acute principal discharge diagnosis code. To adapt the algorithm for this measure, we removed from the potentially planned procedure list two procedures, cardiac catheterization and amputation, because the need for these procedures might reflect progression of clinical conditions that potentially could have been managed in the ambulatory setting to avoid admissions for these procedures. For full details on the planned admission algorithm as adapted for this measure, please see Appendix A of the attached technical report.
Appendix A of the attached technical report contains the detailed algorithm used to identify planned admissions. Among 2,123,190 admissions in the 2012 Medicare Full Sample, 145,443 (6.9%) were planned admissions. For ACO patients, there were 102,740 admissions; of these, 7,991 (7.8%) were planned admissions. For non-ACO patients, there were 2,020,450 admissions; of these, 137,452 (6.8%) were planned admissions.
Please see Data Dictionary, sheet "S.6 ICD9-ICD10 Planned Algorithm," for the ICD-9 to ICD-10 crosswalk for the planned admission algorithm.

Outcome Attribution:
The outcome is attributed to the ACO to which the patient is assigned. Patients are assigned to ACOs according to the specific ACO program assignment algorithm. For example, for the Medicare Shared Savings Program, patient assignment is done retrospectively based on the plurality of care received at that ACO during the measurement year. Information on ACO patient assignment can be found here: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Downloads/Shared-Savings-Losses-Assignment-Spec-v2.pdf.

Citations:
Brown RS, Peikes D, Peterson G, Schore J, Razafindrakoto CM. Six features of Medicare coordinated care demonstration programs that cut hospital admissions of high-risk patients. Health Affairs. 2012 Jun 2012;31(6):1156-1166.

Center for Medicare and Medicaid Services. Medicare Shared Savings Program Shared Savings and Losses and Assignment Methodology Specifications. 2013; https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Downloads/Shared-Savings-Losses-Assignment-Spec-v2.pdf. Accessed July 30, 2014.

Centers for Medicare & Medicaid Services (CMS). Medicare Health Support. 2012; https://www.cms.gov/Medicare/Medicare-General-Information/CCIP/. Accessed March 27, 2014.

Horwitz LI, Grady JN, Cohen DB, Lin Z, Volpe M, Ngo CK, Masica AL, Long T, Wang J, Keenan M, Montague J, Suter LG, Ross JS, Drye EE, Krumholz HM, Bernheim SM. Development and validation of an algorithm to identify planned readmissions from claims data. J Hosp Med 2015 Oct; 10(10):670-7.

McCarthy D, Cohen A, Johnson MB. Gaining Ground: Care Management programs to reduce hospital admissions and readmissions among chronically ill and vulnerable patients. The Commonwealth Fund, New York. 2013.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The target population is ambulatory Medicare FFS patients aged 65 years and older with a diagnosis of heart failure.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

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)*
Note: The denominator of the measure score is the expected admission rate for the ACO; we use this box to describe the measure cohort.

The targeted patient population is Medicare FFS patients aged 65 years and older with a diagnosis of heart failure receiving ambulatory care during the measurement period. To be included in the cohort, patients must have one inpatient principal discharge diagnosis code of heart failure or two heart failure diagnosis codes in any position (inpatient and/or outpatient claims) within one or two years prior to the measurement period. We allowed for up to two years of claims to define the cohort since there is no specified optimal frequency of follow-up visits among ambulatory, stable patients (i.e., patients without a change in their symptoms may never be hospitalized and may only be seen annually). To be included in the cohort, patients must be enrolled full-time in both Part A and B during the year prior to the measurement period.

Heart failure is defined using the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnosis codes identified in Medicare Part A inpatient and outpatient claims data. Patients excluded from the cohort are identified using ICD-9-CM procedure codes in Medicare Part A inpatient and outpatient claims and the Medicare Denominator File. The ICD-9-CM codes that define the cohort and cohort exclusions are listed in the attached Excel file, sheets "S.9 Denominator Details – Cohort" and "S.11 Denominator Exclusions."

An ICD-9-CM to ICD-10-CM code crosswalk is attached in data field S.2b. (Data Dictionary or Code Table).

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
The measure excludes:
1. Patients without continuous enrollment in Medicare Part A for the duration of the measurement period (or until death).
Rationale: We exclude these patients to ensure full data availability for outcome assessment (Part A during the measurement year).
2. Patients with left ventricular assist devices (LVADs).
Rationale: We exclude these patients because while they have a high risk of admission, they are low in prevalence and are clustered among a few ACOs.

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)*
1. Patients without continuous enrollment in Medicare Part A for the duration of the measurement period (or until death).
Lack of continuous enrollment in Medicare Part A is determined by patient enrollment status in FFS Part A using the Medicare Denominator File. The enrollment indicators must be appropriately marked during the measurement period (Part A).

2. Patients with LVADs.
We identify patients as having an LVAD based on ICD-9-CM procedure codes in Medicare Part A or B assigned to the patient within the two years prior to the measurement year. The ICD-9-CM codes are listed below and are also found in the attached Excel file,

sheet "S.11 Denominator Exclusions."
ICD-9-CM Code/Description
37.60/Implantation of heart and circulatory assist system(s)
37.62/Insertion of temporary non-implantable extracorporeal circulatory assist device
37.65/Implant of single ventricular (extracorporeal) external heart assist system
37.66/Insertion of implantable heart assist system
37.68/Insertion of percutaneous external heart assist device

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)*
Not applicable. This measure is not stratified.

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)*
We use a two-level hierarchical negative binomial model to estimate risk-standardized acute, unplanned admissions per person-year at risk for admission. This approach accounts for the clustering of patients within ACOs and variation in sample size.

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" [1-2]. The risk-standardization model includes age and 22 clinical variables. We define clinical variables using condition categories (CCs), which are clinically meaningful groupings of more than 15,000 ICD-9 diagnosis codes [3]. A map showing the assignment of ICD-9 codes to CCs can be found in the attached Data Dictionary Excel file, sheet "S.14 CC to ICD-9." Data Dictionary, sheet "S.15 ICD9-ICD10 Pacemaker" contains the crosswalk of ICD-9 to ICD-10 codes for the pacemaker/cardiac resynchronization therapy/implantable cardiac device variable.

Model Variables
The risk-adjustment variables are:
1. Age
2. Pulmonary diseases (CC 107-110, 114-115)
3. Disability/Frailty (CC 21, 67-69, 100, 116, 148-149, 157, 177-178)
4. Other advanced organ failure (CC 77, 79)
5. Arrhythmia (CC 92-93)
6. Psychiatric Illness/Substance Abuse (CC 51-60)
7. Kidney disease (CC 128, 131-132)
8. Dialysis Status (CC 130)
9. Advanced cancer (CC 7-9, 11)
10. High risk cardiovascular conditions (CC 81-82, 89, 104)
11. Low risk cardiovascular conditions (CC 83-84, 94, 105-106)
12. Structural heart disease (CC 86-88)
13. Dementia (CC 49-50)
14. Diabetes with complications (CC 15-19, 119-120)
15. Gastrointestinal/genitourinary diseases (CC 29-31, 33-34, 133,176)
16. Hematologic diseases (CC 44, 46)
17. Infectious/immunologic diseases (CC 1, 3-5, 45, 85)
18. Liver disease (CC 25-28)
19. Neurological diseases (CC 48, 61, 65, 70-75, 95-99, 101-103,155)
20. Pacemaker/cardiac resynchronization therapy/implantable cardiac device (ICD-9-CM codes 00.50, 00.51, 00.52, 00.53, 00.54, V45.01, V53.31, V53.39, V45.02, V53.32, 37.7, 37.71, 37.72, 37.73, 37.74, 37.74, 37.76, 37.77, 37.78, 37.79 37.80, 37.81, 37.82, 37.83, 37.85, 37.86, 37.87, 37.89, 37.94, 37.95, 37.96, 37.97, 37.98, 37.99)
21. Iron deficiency anemia (CC 47)
22. Major organ transplant (CC 174)

23. Other organ transplant (CC 175)

Citations:
1. 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.
2. Normand S-LT, Shahian DM. Statistical and Clinical Aspects of Hospital Outcomes Profiling. Stat Sci. 2007; 22 (2): 206-226.
3. Pope, G.C., Kautter, J., Ellis, R.P., et al.: Risk Adjustment for Medicare Capitation Payments Using the CMS-HCC Model. Health Care Financing Review. 2004; 25(4):119-141.

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*)

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.)*
The risk-standardized acute admission rate (RSAAR) for each ACO is calculated as the number of "predicted" to the number of "expected" admissions per person-year, multiplied by the national rate of admissions per person-year among all Medicare FFS patients with heart failure – i.e., all eligible Medicare FFS patients with heart failure are used in the measure score calculation, and a score is generated for each ACO. For a full description of the modeling, please see the attached technical report (Section 3.5.5 and Appendix B of attached technical report).

In brief, the measure uses a hierarchical (two-level) statistical model that accounts for the clustering of patients within ACOs and accommodates the widely varying sizes of different ACOs. The measure uses a negative binomial model since our outcome is a count of the number of admissions. The first level of the model adjusts for patient factors. The relationship between patient risk factors and the outcome of admission is determined based on a national sample of patients with heart failure. Stated another way, since the effects that risk factors exert on the number of admissions are estimated based on data from all ACO and non-ACO patients in the nation, the 'expected' number of admissions for each ACO is based on the performance of a national group of providers.

The second level of the model estimates a random-intercept term that reflects the ACO's contribution to admission risk, based on its actual admission rate, the performance of other providers with similar case mix, and its sample size. The ACO-specific random intercept is used in the numerator calculation to derive ACO-specific number of "predicted" admissions per person-year.

The measure score is the ratio of predicted admissions over the expected admissions multiplied by the crude national rate. The predicted to expected ratio of admissions is analogous to an observed/expected ratio, but the numerator accounts for clustering and sample-size variation.

The expected number of admissions is calculated based on the ACO's case mix and an intercept derived from a national average of all patients included in the cohort.
The predicted number of admissions is calculated based on the ACO's case mix and the estimated ACO-specific intercept term.

We multiply the ratio for each ACO by a constant, the crude national rate of acute, unplanned admissions per person-years at risk

for hospitalization, for ease of interpretation.

To place ACOs in performance categories, for each ACO RSAAR, one can calculate a 95% interval estimate (IE), which is similar to a confidence interval, using standard bootstrapping methods (further described in the Testing Form, Section 2b5.1). Using the 95% IEs, one can assign ACOs to one of three performance categories: 'better than the national rate,' 'no different than the national rate,' and 'worse than the national rate.' The ACO is 'better than the national rate' if the 95% IE is completely below the United States (US) national rate among Medicare FFS patients with heart failure; 'no different than the national rate' if the 95% IE is included in the US national rate among Medicare FFS patients with heart failure; and 'worse than the national rate' if the 95% IE is above the US national rate among Medicare FFS patients with heart failure.

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.
Not applicable. This 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.
Not applicable. This 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.
Not applicable.

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

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.
Medicare administrative claims and enrollment data

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)*
Integrated Delivery System

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic, Other
If other: ACO

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.)*
Not applicable.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
Heart_Failure_ACO_Admission_Measure_NQF_Testing_Form_01-29-16_v1.0.docx

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Risk-Standardized Acute Admission Rates for Patients with Heart Failure
Date of Submission: 1/29/2016
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☒ Outcome (*including PRO-PM*)
☐ 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 decision making) is a basis for exclusion, there must be evidence that the exclusion

impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [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*).

To develop and to test the patient-level model, we used several 2010-2012 Medicare claims datasets as outlined below:

1. Medicare dataset used to identify the heart failure cohort and patient risk factors for admission:

We used the 2010-2011 Chronic Conditions Data Warehouse (CCW) 100% dataset which includes patients with at least one of the 27 CCW chronic conditions. We used the CCW 2010-2011 Medicare Part A and Part B files to define the cohort and CCW 2011 Medicare Part A and Part B files to identify each patient's risk factors for the outcome of acute, unplanned admissions per person-year at risk for admission. Our heart failure cohort is fully encompassed within this dataset of patients with at least one CCW chronic condition.

We used the 2011-2012 Denominator File to determine Medicare fee-for-service (FFS) enrollment, demographic, and death information for beneficiaries in our cohort in order to determine inclusion/exclusion criteria for the cohort.

2. Medicare dataset to capture the outcome (acute, unplanned admissions per person-years at risk for hospitalization):

We used the 2012 Medicare Provider Analysis and Review (MedPAR) 100% FFS dataset, containing Medicare Part A claims, to identify the outcome of admissions.

We used the 2012 Denominator File to determine Medicare FFS enrollment, demographic (including 5-digit zip code), and death information for beneficiaries in the heart failure cohort to determine person-years at risk for hospitalization.

3. Dataset to identify assignment of patients to Accountable Care Organizations (ACOs):

We used a file provided by a CMS contractor to identify which Medicare FFS beneficiaries were assigned to each of 114 Medicare Shared Savings Plan ACOs in the year 2012.

4. Dataset to determine socioeconomic status (SES):

We used the 2008-2012 American Community Survey data from the United States (US) Census Bureau to derive the Agency for Healthcare Research and Quality (AHRQ) SES index for each zip code in the United States (US).

5. Dataset to identify dual-eligibility status:

We used the 2012 Denominator File to identify dual-eligible Medicare FFS beneficiaries.

The datasets used for testing vary by testing type; see Section 1.7 for details.

1.3. What are the dates of the data used in testing? Click here to enter date range

We used data from 2010-2012. The dates of the data listed above are as follows:

1. CCW 100% Medicare Parts A and B dataset: 2010-2011

2. MedPAR dataset: 2012

3. ACO assignment data: 2012

4. US Census Bureau, American Community Survey dataset: 2008-2012

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

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

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

The number of measured entities (i.e., ACOs) varies by testing type; see Section 1.7 for details.

1.6. How many and which patients 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 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.

As set forth in Section 1.2 above, we use Medicare claims and enrollment data to identify the cohort, to define the outcome, and to accumulate risk-adjustment variables. For measure development and testing, we created datasets using 2010-2012 Medicare data, using 2012 as the measurement year. The datasets, dates, number of measured entities and number of patients used in each type of testing are as follows:

1) 2012 Medicare Full Sample
This sample includes the cohort of all Medicare FFS beneficiaries meeting our heart failure definition for the 2012 measurement year. The 2012 Medicare Full Sample included 2,581,892 patients with heart failure. Patients were mostly female (56.9%) with an average age of 80.4 years. There were 114 ACOs in the 2012 Medicare Full Sample. Among the 2,581,892 patients with heart failure, 123,626 (4.8%) were assigned to one of 114 ACOs.
-Dataset used for: testing measure exclusions (see Section 2b3), meaningful differences in performance (see Section 2b5), risk-adjustment model (Section 2b4.4b), and all ACO measure score calculations

For model development and testing, we randomly split the 2012 Medicare Full Sample into two subsets of patients: the 2012 Development Sample and 2012 Validation Sample (described below).

a) 2012 Development Sample
-This sample includes 1,290,946 patients with heart failure. Patients were mostly female (56.9%), with an average age of 80.5 years. There were 114 ACOs; 62,350 (4.8%) patients in the 2012 Development Sample were assigned to ACOs.
-Dataset used for: data element reliability (see Section 2a2.3), testing risk-adjustment model (see Section 2b4)

b) 2012 Validation Sample
-This sample includes 1,290,946 patients with heart failure. Patients were mostly female (56.9%), with an average age of 80.4 years. There were 114 ACOs; 61,276 (4.7%) patients in the 2012 Validation Sample were assigned to ACOs.
-Dataset used for: data element reliability (see Section 2a2.3), testing risk-adjustment model (see Section 2b4)

We also split the 2012 Medicare Full Sample into subsets of patients by randomly splitting each ACO's patients in half and then randomly splitting all non-ACO patients in half.

c) 2012 Reliability Sample 1
-2012 Reliability Sample 1 includes 1,290,999 patients with heart failure. Patients were mostly female (56.9%), with an average age of 80.5 years. 61,840 (4.8%) patients were assigned to ACOs.
-Dataset used for: measure score reliability (see Sections 2a2 and 2b2)

d) 2012 Reliability Sample 2
-2012 Reliability Sample 2 includes 1,290,893 patients with heart failure. Patients were mostly female (56.9%), with an average age of 80.4 years. 61,786 (4.8%) patients were assigned to ACOs.
-Dataset used for: measure score reliability (see Sections 2a2 and 2b2)

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 used two different indicators of Medicare beneficiaries' socioeconomic status (SES): (1) the SES score of the patient's five-digit zip code, adapted from the Agency for Healthcare Research and Quality (AHRQ) SES

Index, which was created for the purpose of characterizing the SES of Medicare beneficiaries and (2) the Medicaid dual-eligibility status of beneficiaries [1]. Although race was available (as black or other) in the Medicare data, we chose not to further evaluate it based on our conceptual model and input from our technical expert panel (TEP) and public comment.

The AHRQ SES Index is based on seven neighborhood variables previously shown to contribute to SES and to be associated with outcomes. They are: (1) median household income, (2) percentage of persons living below the federal poverty level, (3) percentage of persons who are aged >16 years and in the labor force but not employed, (4) median value of owner-occupied homes, (5) percentage of persons aged >25 years who completed at least a 12^{th}-grade education, (6) percentage of persons aged >25 years who completed at least four years of college, and (7) percentage of households that average one or more persons per room. The original AHRQ SES Index was derived using data from the 2000 U.S. Census Bureau and was calculated using U.S. Census Block data, which corresponded to Medicare beneficiaries' nine-digit zip code. For this measure, we used data from the U.S. Census Bureau, American Community Survey (2008-2012) and performed a principal component analysis to derive a composite SES index score for each five-digit zip code, which we then assigned to the patient based on their zip code of residence (i.e., the smallest unit by which we could identify Medicare beneficiaries' home address). The AHRQ SES Index is a continuous variable whereby lower scores indicate lower SES zip codes and higher scores indicate higher SES zip codes.

We created a dichotomous variable from the AHRQ SES index, stratifying zip code scores into 'low SES' and 'non-low SES.' Based on the distribution of the AHRQ SES index among the entire FFS Medicare population in the 5% Medicare FFS sample, we selected the lowest quintile to represent low SES. In this lowest quintile, 21.9% of beneficiaries were Medicaid dual-eligible, as compared with 13.7% in the second lowest quintile. We then categorized each patient as low or non-low SES based on the AHRQ score derived from their zip code of residence.

Additionally, we categorized ACOs based on the proportion of low SES patients in their cohort into quartiles (first quartile [Q1] indicating few low SES patients, fourth quartile [Q4] indicating many low SES patients). Similarly, we categorized ACOs by the proportion of Medicaid dual-eligible patients in their cohort into ACOs caring for 'few' (Q1) and 'many' (Q4) Medicaid dual-eligible patients. For more information on the derivation of the AHRQ SES index and the selection of a low SES thresholds for patients and ACOs, see the Appendix E of the attached Appendix.

We did not use race in our analyses since differences in risk of admission among groups of different race should be captured in our risk-adjustment model (which includes age and comorbidities). Any remaining differences in the risk for hospitalization among patients of different race may represent disparities in care delivery and quality of care.

Citations

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

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 in Medicare FFS patients, we aimed to utilize only those data elements from claims data that have both face validity and reliability. We avoided the use of fields that are thought to be coded inconsistently across facilities. Specifically, we used fields that are consequential for payment and which are audited. We identified such variables through empiric analyses and our understanding of the Centers for Medicare & Medicaid Services (CMS) auditing and billing policies. We sought to avoid variables which do not meet these standards.

In addition, CMS has in place several hospital auditing programs used to assess overall accuracy of claims-based coding, to ensure appropriate billing, and for overpayment recoupment. CMS routinely conducts data analysis to identify potential problem areas and to 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 assessed the reliability of the data elements by comparing model variable frequencies in our 2012 Development Sample and 2012 Validation Sample.

Measure Score Reliability

The reliability of a measurement can be defined as the degree to which repeated measurements of the same entity agree with one another. For our measures of facility performance, the measured entity is the ACO, and reliability is the extent to which repeated measurements of the same ACO give similar results [1].

To calculate measure score reliability, we randomly sampled half of the patients from each ACO and half of the patients who were not in ACOs from the 2012 Medicare Full Sample (2012 Reliability Sample 1 and 2012 Reliability Sample 2). We calculated the measure score for all the ACOs using data from ACO and non-ACO patients, and repeated the calculation using the second half of patients. Thus, each ACO was measured twice, but each measurement was 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 ACO, not of the patients. As a metric of agreement we calculated the intraclass correlation coefficient (ICC) [2], and assessed the values according to conventional standards [3]. The agreement of the two risk-standardized acute admission rates was quantified for ACOs in each sample using the ICC (2,1) by Shrout and Fleiss [2].

Citations

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

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

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

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:

Table 1. Risk variable frequencies for 2012 Development Sample and 2012 Validation Sample

Variable	Prevalence of risk factors (%)

	Development Sample N = 1,290,946	Validation Sample N = 1,290,946
Age		
65-70 years	10.5	10.5
70-80 years	35.0	35.1
80-90 years	40.1	40.1
≥90 years	14.4	14.4
High risk cardiovascular factors	32.5	32.5
Low risk cardiovascular factors	84.4	84.4
Arrhythmia	62.6	62.7
Structural heart disease	39.7	39.7
Advanced cancer	7.4	7.3
Dementia	25.7	25.7
Diabetes with complications	51.7	51.7
Dialysis status	3.0	3.0
Disability/Frailty	24.2	24.3
Gastrointestinal and genitourinary disorders (GI/GU)	32.1	32.2
Hematological disorders	16.0	16.1
Infectious and immune disorders	6.1	6.1
Kidney disease	38.2	38.2
Liver disease	2.3	2.4
Neurological disease	45.8	45.8
Psychiatric illness/Substance abuse	38.6	38.8
Pulmonary diseases	60.3	60.4
Other advanced organ failure	21.2	21.2
Iron deficiency anemia	54.1	54.0
Major organ transplant	0.3	0.3
Other organ transplant	0.8	0.8
Pacemaker/cardiac resynchronization therapy/implantable cardiac device	21.9	21.9

Measure Score Reliability:
The ICC between the two risk-standardized acute admission rates (RSAARs) was 0.81, which according to the conventional interpretation is “excellent” [1].

Citations
1. 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?*)

Data Element Reliability Results
Compared with the 2012 Development Sample, the mean age of patients and the frequency of risk-adjustment variables were similar in the 2012 Validation Sample. This suggests that the data elements are reliable across these samples.

Measure Score Reliability Results
The ICC demonstrates excellent agreement across samples, indicating 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)*

We demonstrated measure validity through: (1) reliance on relevant prior validity testing conducted for other claims-based measures; (2) use of established measure development guidelines; and (3) assessment by external groups and a TEP.

1. Validity of Claims-Based Measures
Our team has demonstrated the validity of using claims data for risk adjustment in lieu of medical record data in estimating facility-level measure scores for a number of hospital outcome measures endorsed by the National Quality Forum (NQF). CMS has validated six NQF-endorsed measures currently in public reporting (acute myocardial infarction [AMI], heart failure, and pneumonia mortality and readmission) with models that used medical record-abstracted data for risk adjustment. Specifically, we conducted claims model validation by building comparable models using abstracted medical record data for risk adjustment for patients with heart failure (National Heart Failure data), AMI (Cooperative Cardiovascular Project data), and pneumonia (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]. These findings support this measure's validity; however, we acknowledge that the use of claims data for risk adjustment has been validated for hospital outcomes measure and not for outcome measures among ambulatory patients.

2. 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], 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].

3. Validity as Assessed by External Groups

Throughout measure development, we obtained expert and stakeholder input through: holding regular discussions with the external experts in our working group, consulting our national TEP, and holding a 30-day public comment period. We obtained expert and stakeholder feedback for development of two related measures for patients with diabetes or multiple chronic conditions.

Yale New Haven Health Services Corporation—Center for Outcomes Research and Evaluation (CORE) clinicians and statistical experts comprised the working group. The working group members have expertise in quality measurement, clinical management of patients with heart failure, statistical modeling, healthcare disparities, and healthcare policy. Through regular in-person meetings and teleconferences, the working group discussed all aspects of measure development, including the cohort and outcome definitions and risk adjustment.

In addition to the working group and in alignment with the CMS Measures 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, patients, and individuals with experience in quality improvement, performance measurement, and healthcare disparities. We held four structured TEP conference calls consisting of presentation of key issues, our proposed approach, and relevant data, followed by open discussion among TEP members.

List of TEP Members

1. Lawrence M. Becker, BS, Xerox Corporation (Director, Strategic Partnerships, Alliances and Analytics); Rochester, NY
2. Alex Blum, MD, MPH, Evergreen Health Cooperative (Chief Medical Officer); Baltimore, MD
3. Sanjay Doddamani, MD, Geisinger Health System (System-wide Chief of Advanced Cardiac Disease – Heart Failure); Danville, PA
4. Kevin Fiscella, MD, MPH, University of Rochester Medical Center (Professor of Family Medicine); Rochester, NY
5. Elbert Huang, MD, MPH, University of Chicago (Associate Professor of Medicine, Director of the Center for Translational and Policy Research of Chronic Diseases, and Associate Director of the Chicago Center for Diabetes Translation Research); Chicago, IL
6. Bruce Leff, MD, Johns Hopkins University School of Medicine (Professor of Medicine, Division of Geriatric Medicine); The Johns Hopkins University Bloomberg School of Public Health (Faculty, Health Services Research Development Center and Lipitz Center for Integrated Health Care); Baltimore, MD
7. Andy Miller, MD, MPH, Healthcare Quality Strategies, Inc. (Medical Director); East Brunswick, NJ; Colorado Foundation for Medical Care (CMO, Integrating Care for Populations & Communities National Coordinating Center); Englewood, CO
8. Ami Parekh, MD, JD, University of California, San Francisco (Medical Director for Health System Innovation); San Francisco, CA
9. Christine Ritchie, MD, University of California, San Francisco (Professor of Medicine, Division of Geriatrics); San Francisco, CA
10. Two patient representatives.

We systematically assessed the face validity of the measure score as an indicator of quality by soliciting the TEP members' agreement with the following statement: "The RSAARs obtained from the heart failure measure as specified can be used to distinguish between better and worse quality ACOs."

TEP members indicated their agreement with the face validity of the measure on a six-point scale:
1=Strongly disagree
2=Moderately disagree

3=Somewhat disagree
4=Somewhat agree
5=Moderately agree
6=Strongly agree

Process Used to Identify International Classification of Diseases, Tenth Revision (ICD-10) Codes
This application includes ICD-10 codes that correspond to all International Classification of Diseases, Ninth Revision (ICD-9) codes included in the specifications. The goal was to convert this measure into a new code set, fully consistent with the intent of the original measure.

• ICD-10 diagnosis codes used to the cohort were identified using the 2013 ICD-9-CM to ICD-10-CM General Equivalence Mapping (GEM) files made available by CMS. We then internally performed clinician review of this crosswalk.

• ICD-10 diagnosis codes used to define the pacemaker/CRT/ICD risk variable defined with ICD-9-CM codes were identified using the 2013 ICD-9-CM to ICD-10-CM General Equivalence Mapping (GEM) files made available by CMS. We then internally performed clinician review of this crosswalk.

• ICD-10 diagnosis and procedure codes used to define the Planned Admission Algorithm were identified from the 2014 version of the AHRQ Clinical Classification Software (CCS) categories specified for ICD-10, followed by clinician review. The algorithm also includes some individual ICD-9-CM codes. To create the crosswalk for the ICD-9-level codes, we used the 2013 ICD-9-CM to ICD-10-CM GEM files made available by CMS, followed by clinician review.

Citations
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
The results of the TEP rating of agreement with the validity statement were as follows:
N=8
Mean rating=4.9

All TEP members who responded to the survey indicated they agreed with the statement; 6 of the 8 indicated that they moderately or strongly agreed. Five TEP members did not respond to the TEP survey.

Frequency of Ratings of Agreement:

Rating	# (%) of Responses
1 (Strongly disagree)	0 (0.0)
2 (Moderately disagree)	0 (0.0)
3 (Somewhat disagree)	0 (0.0)
4 (Somewhat agree)	2 (25.0)
5 (Moderately agree)	5 (62.5)
6 (Strongly agree)	1 (12.5)

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?*)

There was strong support expressed by the members of the TEP and in public comment for the validity of the measure. There were no strong concerns about the measure. One of 13 commenters felt the outcome was not an indicator of quality. See public comment document for further details: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/CallforPublicComment.html

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*)

We determined the exclusions to be appropriate based on clinical and methodological considerations, such as whether we had sufficient data for patient subsets or could adequately adjust for the risk of admission in certain patient subpopulations. We examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion.

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*)

Table 2 provides the number of patients excluded from the heart failure cohort. Out of the total number of Medicare FFS patients with heart failure (N = 2,649,829), we excluded 66,909 due to non-continuous enrollment in Medicare Part A in 2012 because we were not able to adequately capture the outcome for these patients. None of these excluded patients were assigned to ACOs by Medicare's patient assignment algorithm.

In addition, we excluded 1,048 patients with heart failure who had a history of a left ventricular assist device (LVAD); 49 of those excluded patients were assigned to ACOs. Because these patients were at high risk for admission and clustered among a small group of ACOs, we would not be able to adequately risk adjust for them in the measure.

Since the number of excluded patients assigned to ACOs was very low, we did not perform a frequency distribution analysis across ACOs.

The final cohort included 2,581,892 patients.

Table 2. Patients excluded from sample for each exclusion criterion

Exclusion	Number excluded from Medicare FFS heart failure cohort	Number of patients excluded from ACOs
1. Non-continuous enrollment in Part A in 2012	66,909	0
2. Patients with left ventricular assist devices	1,048	49

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*)

We needed to exclude patients without continuous enrollment to capture the outcome as well as exclude a clinical subpopulation of patients at high risk of admission to ensure adequate risk adjustment. We excluded very few patients based on clinical criteria. As a result, the measure captures the majority of Medicare FFS patients 65 years and older with a diagnosis of heart failure who are continuously enrolled (96.0%).

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 23 risk factors**

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

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

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

Not applicable. This measure is risk-adjusted.

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*)

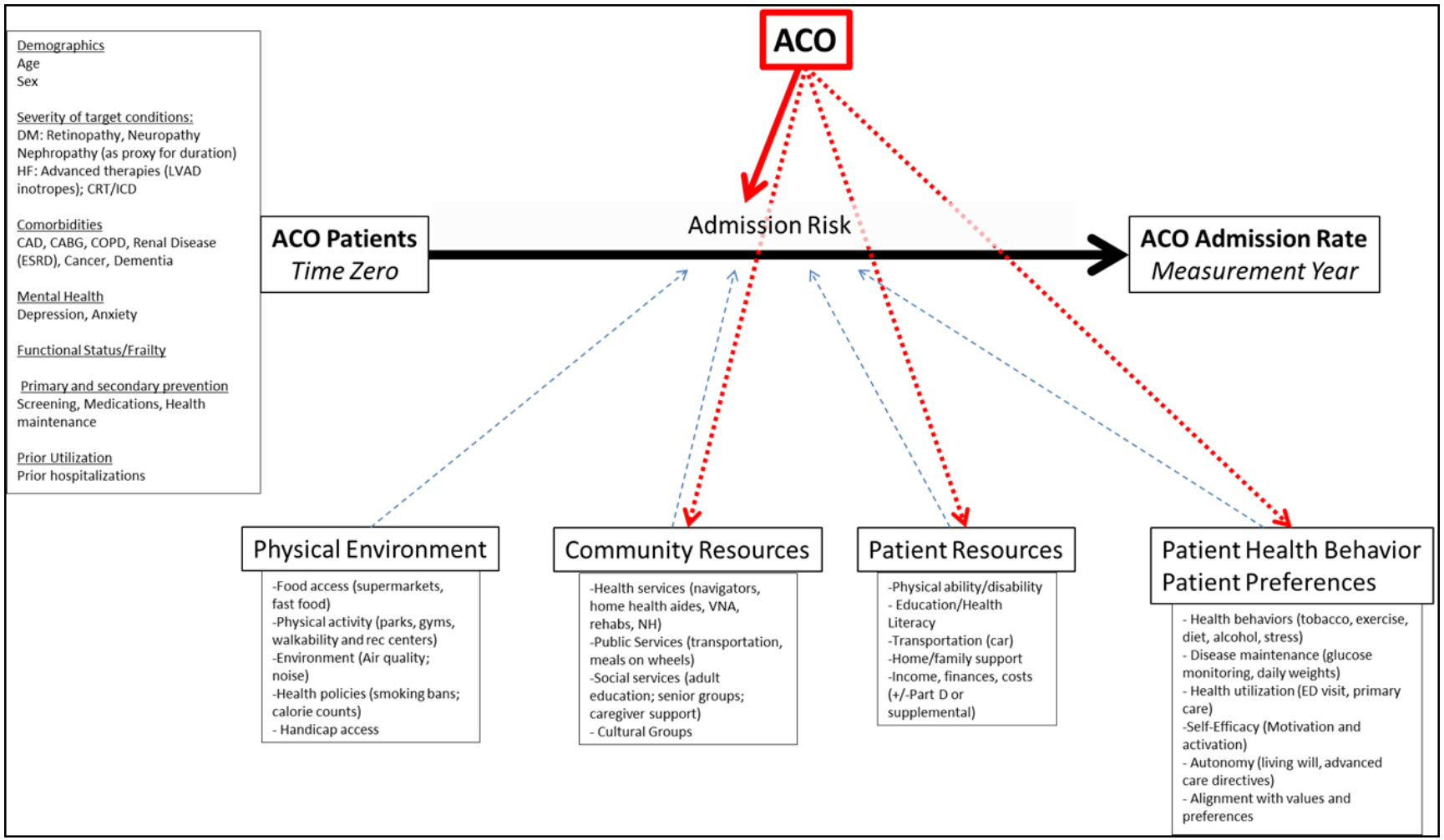
We selected the risk-adjustment model variables based on the existing literature, clinical judgment, empirical analyses, and input from our TEP and other experts. We considered factors that may impact the rate of admission, including patient-level factors (e.g., demographics, SES, clinical risk factors on admission); we also considered the impact of other non-clinical factors such as health behaviors and community resources. In this work, we were guided by a conceptual framework that was informed by a literature review and environmental scan, outlining the relationships between potential clinical and contextual factors and rates of admissions among chronic disease populations cared for by ACOs. Importantly, many factors other than traditional medical care delivered in the office or

hospital settings will impact health outcomes for patients with chronic disease. For example, ACOs practicing in communities where patients have limited access to transportation, healthy foods and recreational facilities, may have less success in promoting healthy behaviors among patients with heart failure; this may in-turn impact quality outcomes. Recognition of and attention to the health environment may be important for achieving the goals of better care, better health, and lower costs and thus, shared savings.

The conceptual model (Figure 1) was presented and endorsed by the TEP engaged during the development of this measure. The model recognizes patient-level demographic and clinical factors, along with 4 contextual domains that may influence ACO performance: (1) Physical environment (e.g., green spaces; safe streets); (2) Community resources (e.g., home health; senior services); (3) Patient resources (e.g., social support; transportation; income); and (4) Patient behavior/personal preferences (e.g., exercise; diet; advanced care directives; preference for intervention).

The model also recognizes the capacity of ACOs to mitigate the effects of many contextual factors on rates of admissions, encompassing both SES and non-SES variables, and supporting our decision not to adjust for contextual factors. Adjusting for contextual factors would obscure important differences in ACO quality and could serve as a disincentive for ACOs to engage with such factors. We did, however, conduct analyses of SES factors to further inform the committee's deliberation.

Figure 1. Conceptual model of factors affecting risk of hospital admission



We describe our approach to risk-adjustment for the demographic factors, clinical risk factors, and contextual domains, in turn, below:

1. Demographic factors
We used clinical and conceptual criteria to adjust this measure for age but not sex or race. Age is a clinically recognized risk factor for acute admissions. In contrast, sex or race differences in risk of admission should be captured in our risk-adjustment model (which includes age and comorbidities). Any remaining differences in the risk for hospitalization among patients of different sex or race may represent disparities in care delivery and quality of care [1,2]. We did examine the effects of including sex in the models, since the relationship between sex and acute, unplanned admissions has not been tested in this setting, finding that sex was not significant after adjusting for age and clinical comorbidities.

2. Clinical risk factors
We used clinical, conceptual, and statistical criteria to select clinical risk factors for adjustment. This measure adjusts for clinical risk factors that are present at the start of the measurement period, but not for conditions that arise during the measurement period.

Development of Candidate Clinical Variables:
To select candidate variables for risk adjustment, we used Part A and Part B data from one year prior to the measurement year for 100% of the Medicare FFS patients included in the cohort (2012 Medicare Full Sample). We reviewed 189 diagnostic groups included in the Hierarchical Condition Category (HCC) clinical classification system. We defined comorbidities using 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, sheet "S.14 CC-ICD-9 Map." To select candidate variables, two clinicians reviewed all 189 CCs and excluded those that were not relevant to the Medicare population or that were not clinically relevant to the all-cause acute admission outcome (e.g., attention deficit disorder, female infertility). The remaining 181 clinically relevant CCs were selected as candidate variables.

Among the 181 clinically relevant CCs, we calculated the prevalence of the CC in the year preceding the measurement period (i.e., 2011), the number of hospital admissions per patient-year during the measurement period (i.e., 2012) among patients with and without the CC, and the rate ratio for the number of hospital admissions associated with each CC. Based on these statistical findings, we reduced the list of CCs to 115 from the initial list of 181 clinically relevant CCs. We reviewed the results of the bivariate analyses of the 115 CCs and collapsed the 115 CCs into 22 candidate variables, plus age. Additionally, we developed a variable that captured the use of three cardiac devices that may reflect severity of heart failure and the risk of admission (implantable cardiac defibrillator [ICD], cardiac resynchronization therapy [CRT], and permanent pacemaker).

Candidate Clinical Variables
The selected candidate variables were:
1. Age (categorized) (65-70, 70-80, 80-90, 90+)
2. High risk cardiovascular factors (CC 81-82, 89, 104)
3. Low risk cardiovascular factors (CC 83, 84, 94, 105-106)
4. Arrhythmia (CC 92, 93)

5. Structural heart disease (CC 86-88)
6. Advanced cancer (CC 7-9, 11)
7. Dementia (CC 49-50)
8. Diabetes with complications (CC 15-19, 119-120)
9. Dialysis (CC 130)
10. Disability/Frailty (CC 21, 67-68, 100, 116, 148-149, 157, 177-178)
11. Gastrointestinal and genitourinary disorders (CC: 29-30, 31, 33-34, 133, 176)
12. Hematological disorders (CC 44, 46)
13. Infectious/Immune disorders (CC 1, 3-5, 45, 85)
14. Kidney disease (CC 128, 131-132)
15. Liver disease (CC 25-28)
16. Neurological disorders (CC 48, 61, 65, 70-75, 95-99, 101-103, 155)
17. Psychiatric illness/Substance abuse (CC 51-60)
18. Pulmonary disease (CC 107-110,114-115)
19. Other advanced organ failure (CC 77, 79)
20. CRT/ICD/Pacemaker (ICD-9-CM codes 00.50, 00.51, 00.52, 00.53, 00.54, V45.01, V53.31, V53.39, V45.02. V53.32, 37.7, 37.71, 37.72, 37.73, 37.74, 37.74, 37.76, 37.77, 37.78, 37.79 37.80, 37.81, 37.82, 37.83, 37.85, 37.86, 37.87, 37.89, 37.94, 37.95, 37.96, 37.97, 37.98, 37.99)
21. Iron deficiency anemia (CC 47)
22. Major organ transplant (CC 174)
23. Other organ transplant (CC 175)
24. Hip fracture/major fracture (CC 158-159)

Final variable selection

In order to select the final set of variables, we ranked the variables in terms of their importance for the model by comparing the Akaike Information Criterion (AIC) values using the 2012 Development Sample. We selected variables starting with the 24 candidate variables. We removed one variable and determined the best combination of 23 variables that resulted in the smallest AIC compared with other combinations of 23 variables. Based on the best 23 variables, we removed one more variable and determined the best 22 variables. We repeated these steps until we reached one variable. Each of the final 24 models represents the best model (combination of variables) given different numbers of variables.

The attached Data Dictionary, sheet "S.15 Risk Model Specs" indicates the final risk variables selected, the codes used to define the risk variables for our statistical model, and their frequencies in the 2012 Development Sample and 2012 Validation Sample.

3. Socioeconomic status

Based on a conceptual model that was informed by a literature review and environmental scan, we did not adjust for contextual factors which may impact acute admissions, including variables related to SES. ACOs should and do influence a broad range of patient and community level factors that can mitigate the risk of admission associated with the contextual environment.

However, to inform the committee's consideration of the decision not to adjust for SES, we performed focused analyses using SES variables. These analyses are informative for future

measure use, but the decision not to adjust for SES in this measure was not based on the results of these statistical analyses.

To assess the potential effect of SES on ACO performance, we first included SES as a patient-level covariate in the models. As there are no standardized methods for assessing a Medicare beneficiary's SES, we used two different indicators of SES: (1) the SES score of the patient's 5-digit zip code, adapted from the AHRQ SES Index [3], which was developed for the purpose of characterizing the SES of Medicare beneficiaries and (2) the Medicaid dual-eligibility status of beneficiaries. We created a dichotomous variable from the AHRQ SES score, defining patients as low SES if they had an AHRQ Score of 0 to 45 and *non*-low SES if they had an AHRQ score of >45. This cut-point represented the lowest quintile of AHRQ SES scores among the 5% Medicare FFS Sample. In this lowest quintile, 21.9% of patients were Medicaid dual eligible. For further details on how we calculated the AHRQ SES score and developed a dichotomous variable we refer to the attached technical report, Appendix E. Additionally, we performed ACO-level analyses based on the proportion of low SES patients being cared for by an ACO. These methods and results are reported in the NQF Submission form.

4. Contextual Domains
The four contextual domains, which include SES factors, may influence the clinical health status of patients as well as the outcome of acute admissions, impacting ACOs' ability to prevent acute admissions. However, when evaluating provider quality, we do not want to adjust for them, since these affects may be mediated by ACOs, and the measure score should ideally reflect successful efforts to mitigate their impact on admission rates. This approach is consistent with the ACO program design - as part of their mission, ACOs are encouraged to develop strategic partnerships with community-based organizations and businesses in order to improve population health and reduce the risk of admission. It is also supported by growing evidence that integrated health systems can identify and mitigate the degree to which non-health factors impact health outcomes (e.g., by connecting patients with available health-related services) [4].

Citations:
1. Rathore SS, Foody JM, Wang Y, et al. Race, quality of care, and outcomes of elderly patients hospitalized with heart failure. JAMA : the journal of the American Medical Association. May 21 2003;289(19):2517-2524.

2. Deswal A, Petersen NJ, Urbauer DL, Wright SM, Beyth R. Racial variations in quality of care and outcomes in an ambulatory heart failure cohort. American heart journal. Aug 2006;152(2):348-354.

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

4. Alley DE, Asomugha CN, Conway PH, Sanghavi DM. Addressing Social Needs through Medicare and Medicaid, N Engl J Med 2016; 374:8-11.

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

Based on the smallest AIC among the 24 combinations we retained 23 variables in the final model. Of the 24 candidate variables, the only variable that was not included was the hip and other major fractures variable, which was not statistically significant in the model.

The following variables were selected as the final risk-adjustment variables:
1. Age (categorized) (65-70, 70-80, 80-90, 90+)
2. Pulmonary diseases (CC 107-110, 114-115)
3. Disability/frailty (CC 21, 67-69, 100, 116, 148-149, 157, 177-178)
4. Other advanced organ failure (CC 77, 79)
5. Arrhythmia (CC 92-93)
6. Psychiatric illness/Substance abuse (CC 51-60)
7. Kidney disease (CC 128, 131-132)
8. Dialysis status (CC 130)
9. Advanced cancer (CC 7-9, 11)
10. High risk cardiovascular factors (CC 81, 82, 89, 104)
11. Low risk cardiovascular factors (CC 83-84, 94, 105-106)
12. Structural heart disease (CC 86-88)
13. Dementia (CC 49-50)
14. Diabetes with complications (CC 15-19, 119-120)
15. Gastrointestinal/genitourinary disorders (CC 29-31, 33-34, 133, 176)
16. Hematologic disorders (CC 44, 46)
17. Infectious and immune disorders (CC 1, 3-5, 45, 85)
18. Liver disease (CC 25-28)
19. Neurological diseases (CC 48, 61, 65, 70-75, 95-99, 101-103, 155)
20. Pacemaker/CRT/ICD (ICD-9-CM codes 00.50, 00.51, 00.52, 00.53, 00.54, V45.01, V53.31, V53.39, V45.02. V53.32, 37.7, 37.71, 37.72, 37.73, 37.74, 37.74, 37.76, 37.77, 37.78, 37.79 37.80, 37.81, 37.82, 37.83, 37.85, 37.86, 37.87, 37.89, 37.94, 37.95, 37.96, 37.97, 37.98, 37.99)
21. Iron deficiency anemia (CC 47)
22. Major organ transplant (CC 174)
23. Other organ transplant (CC 175)

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)

We performed multiple analyses to assess the effect of sex and SES on model performance. These analyses are informative for future measure use, but the decision not to adjust for sex or SES in this measure was based on conceptual/clinical factors and not on the results of these statistical analyses (see 2b4.3.).

To assess the effect of sex and SES on model performance, we compared deviance R-squared values with and without the variables for sex and SES included as patient-level variables in the model. We compared the correlation between measure scores with and without sex and SES included in the models, using the Spearman correlation.

For the SES analyses, we also assessed ACO performance among groups of ACOs caring for similar proportions of low SES patients. To do this, we categorized ACOs into quartiles (Q1 indicating ACOs with few low SES patients, Q4 indicating ACOs with many low SES patients). We used boxplots to compare the distribution of RSAARs across ACOs by low SES quartiles.

The SES analyses were performed using both the AHRQ SES index (i.e., low SES, binary variable described above) and Medicaid dual-eligibility status as a proxy for patients' SES status. Statistical analyses were performed using SAS version 9.3 (SAS Institute Inc., Cary, NC).

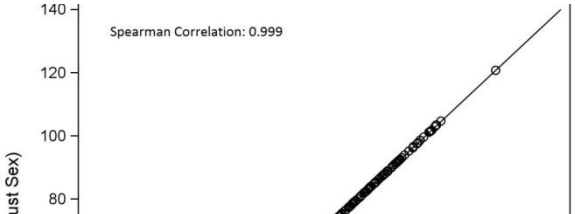
The results of the patient-level analyses indicate that adjustment for sex and for low-SES status as a patient variable in the models did not affect measure performance.

Specifically, related to SES, performance scores did not change appreciably after adjusting the models for patients' SES. As demonstrated in the Testing Form, Section 2b4.11, the Spearman correlation comparing the ACO measure scores estimated with and without risk adjustment for the AHRQ SES Index was 0.990. Similarly, the Spearman correlation for the scores estimated with and without patients' Medicaid dual eligibility was 0.991. These results demonstrate that adjusting for SES at the patient level has little effect on the measure score.

Sex

The deviance R squared values for the two models, one adjusted for the 23 clinical variables *and* sex, and one adjusted for 23 clinical variables *without* adjusting for sex, were 0.123 and 0.122, respectively, meaning adjustment for sex explained the same

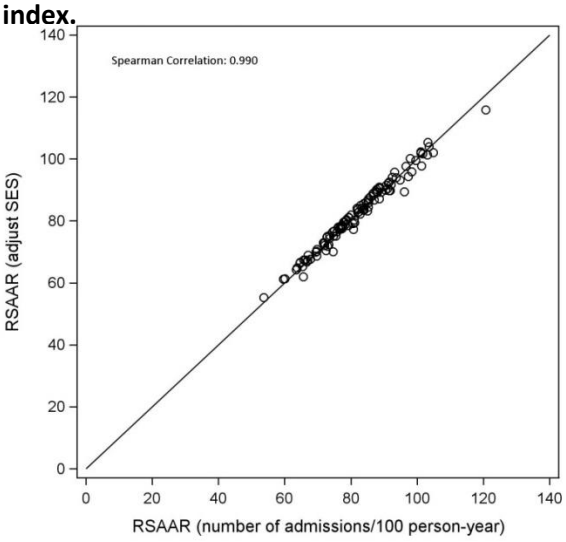
Figure 2. Plot of acute, unplanned admission rates with and without adjustment for sex.



AHRQ SES Index

The deviance R squared values for the two models – one adjusted for the 23 clinical variables *and* Low SES, and one adjusted for the 23 clinical variables *without* adjusting for Low SES, were 0.123 and 0.122 – respectively, meaning adjustment for Low SES explained the same variation and did not provide incremental benefit. Comparing the RSAAR with and without Low SES included in the model resulted in a high degree of correlation (Spearman correlation = 0.990). The graph demonstrates that, compared with *not* adjusting for Low SES, adjusting for Low SES results in some ACOs having slightly lower RSAAR scores (below the line) and other ACOs having higher RSAAR scores (above the line). (Figure 3)

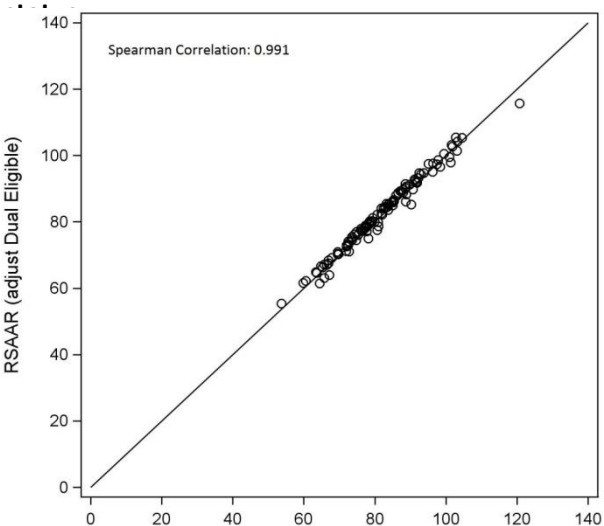
Figure 3. Plot of acute, unplanned admission rates with and without adjustment for AHRQ SES index.



Medicaid Dual-Eligibility Status

The deviance R squared values for the two models, one adjusted for the 23 clinical variables *and* Medicaid dual-eligibility status, and one adjusted for the 23 clinical variables *without* adjusting for Medicaid dual-eligibility status, were 0.124 and 0.122, respectively, meaning that the deviance from the overall model fit was not explained by adjustment for dual-eligibility status. Comparing the RSAAR with and without Medicaid dual-eligibility status included in the model resulted in a high degree of correlation (Spearman correlation = 0.991). The graph demonstrates that, compared with *not* adjusting for Medicaid dual-eligibility status, adjusting for Medicaid

Figure 4. Plot of acute, unplanned admission rates with and without adjustment for dual-eligibility



In assessing the relationship between the proportions of low SES patients enrolled in an ACO and ACO measure performance, we found that ACOs serving many low SES patients more often perform worse than the national rate compared with ACOs serving few low SES patients. This was true using either the AHRQ SES index (25.0% vs. 6.9%, respectively) or Medicaid dual-eligibility status (24.1% vs. 3.4%) as an indicator of patients' SES. However, among ACOs serving many low SES patients, using the AHRQ SES index, 17.9% of ACOs performed better than the national rate and over half (57.1%) performed no different than the national rate, demonstrating that most ACOs serving high proportions of low SES patients can and do perform well on the measure. Similarly, using Medicaid dual-eligibility status as an indicator, 17.2% of ACOs performer better than the national rate and over half (58.6%) performed no different than the national rate.

Figure 5. Boxplots of risk-standardized acute admission rates (RSAARs), comparing ACOs with varying proportions of low SES patients with heart failure (based on AHRQ SES Index; Quartile 1 [Q1]: ACOs with few low SES patients; Quartile 4 [Q4]: ACOs with many low SES patients)

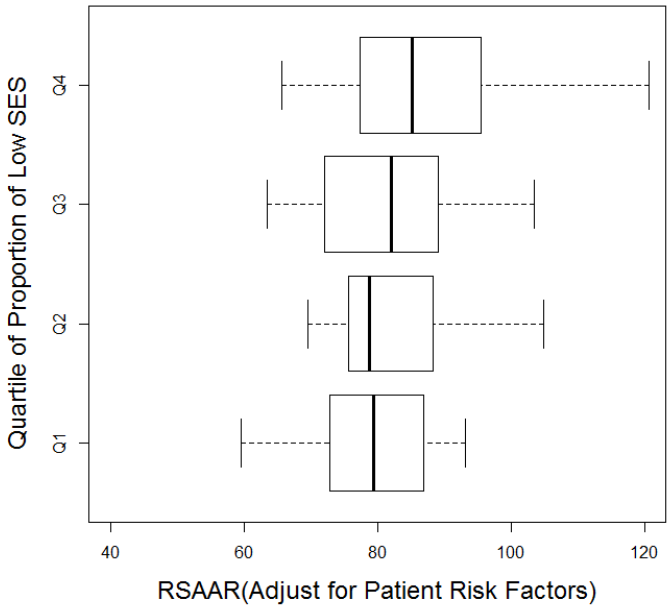
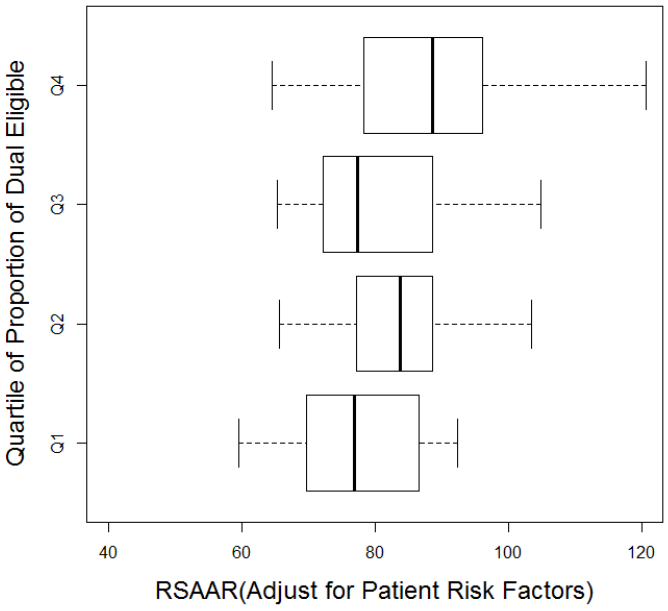


Figure 6. Boxplots of risk-standardized acute admission rates (RSAARs), comparing ACOs with varying proportions of Medicaid dual-eligible patients with heart failure (Quartile 1 [Q1]: ACOs with few Medicaid dual-eligible patients; Quartile 4 [Q4]: ACOs with many Medicaid dual- eligible patients)



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*)

We assessed adequacy of the patient-level risk-adjustment model (described above). We evaluated model performance first in the 2012 Development Sample. We then validated the model performance in the 2012 Validation Sample.

The measure uses the number of acute unplanned hospital admissions per person-year at risk for admission. Because the outcome is a count of hospital admissions – rather than a binary outcome, such as whether or not a patient has been admitted – several routinely used metrics of model performance cannot be applied (for example, we cannot use a c-statistic).

Using the 2012 Development Sample, we computed two summary statistics for assessing the risk-adjustment model performance: goodness-of-fit statistics (deviance R squared) and overfitting indices. We then compared the model performance in the 2012 Development Sample with its performance in the 2012 Validation Sample.

Deviance R squared

Our measure uses a negative binomial function because the outcome is a count of hospital admissions with over-dispersion. We calculated deviance R squared using the deviance residual defined by Cameron [1]. The deviance R squared evaluates how successful the fit is in explaining the variation of the data. Deviance R squared can take on any value between 0 and 1, with a value closer to 1 indicating that a greater proportion of deviance is accounted for by the model. For example, a deviance R squared value of 0.12 means that the fit explains 12% of the total deviance.

Overfitting indices

Overfitting refers to the phenomenon in which a model accurately describes the relationship between the predictive variables and the outcome in the development dataset, but fails to provide valid predictions in new patients.

Model performance among patients at different risk of admission

In order to determine whether the model performs well across groups of patients at different risk of admission, the sample was divided into quartiles of predicted admission rate (highest, second highest, lowest, and second lowest). We then assessed the model probability of the number of admissions compared with the observed probability of the number of admissions.

Generally, residuals measure the departure of fitted values from actual values of the dependent variable, but they cannot be applied to count data. For linear models, a residual is easily defined as the difference between actual and fitted values. For nonlinear models, the definition of a residual is not unique. Specifically, for count data, the raw residual (the observed value minus the fitted value) is heteroskedastic and asymmetric. Therefore, there is no residual that has zero mean, constant variance, and symmetric distribution. For fully parametric models such as negative binomial models, we can compare *predicted* probabilities with *observed* probabilities of each count of admissions. For each patient, we can calculate the *predicted* probability of being

admitted to the hospital *n* times (0, 1, 2, …*n*) given this patient's risk factors for hospitalization. For example, a patient has a single predicted admission rate of 2.5 admissions per person-years of exposure; however, given the assumed negative binomial distribution of the risk of admissions, we can also express the patient's risk of admission as the probabilities of observing 0, 1, 2,…10 hospital admissions. Therefore, for each patient, we can calculate a set of predicted probabilities of observing different counts of admissions. The *predicted* probability for a group of patients is the average probability of observing 0, 1, 2,…*n* hospital admissions, given these patients' risk factors for admission. The *observed* probability of each count of admissions for a group of patients is the proportion of these patients admitted to the hospital 0, 1, 2,…*n* times.

Citations

1. Cameron AC, Windmeijer FAG. R-Squared Measures for Count Data Regression Models with Applications to Health-Care Utilization. Journal of Business & Economic Statistics. 1996;14(2):209-220.

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*)**:**

2012 Development Sample results (deviance R squared): 0.122
2012 Validation Split Sample results (deviance R squared): 0.123

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

2012 Development Sample calibration results (overfitting index): (0, 1)
2012 Validation Sample calibration results (overfitting index): (-0.0020, 1.0002)

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

Below are plots of observed vs. predicted values for the number of hospital admissions among four groups of patients: lowest (A) and second lowest (B) predicted admissions; and second highest (C) and highest (D) predicted admissions in the 2012 Development Sample.

Figure 7. Observed vs. predicted probabilities for the number of hospital admissions among lowest predicted admission group. Lowest predicted admission group (23 to 53 admissions per 100 person-years, median: 41 and interquartile range: 35 to 47.

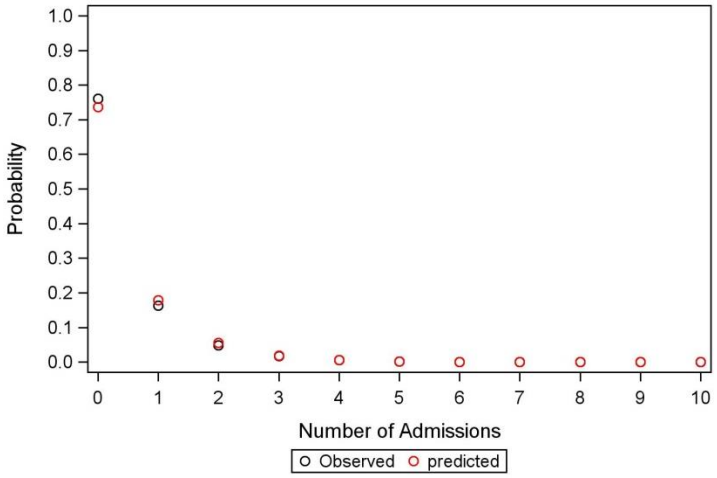


Figure 8. Observed vs. predicted probabilities for the number of hospital admissions among second lowest admission group (53 to 80 admissions per 100 person-years, median: 65, IQR: 58 to 72.

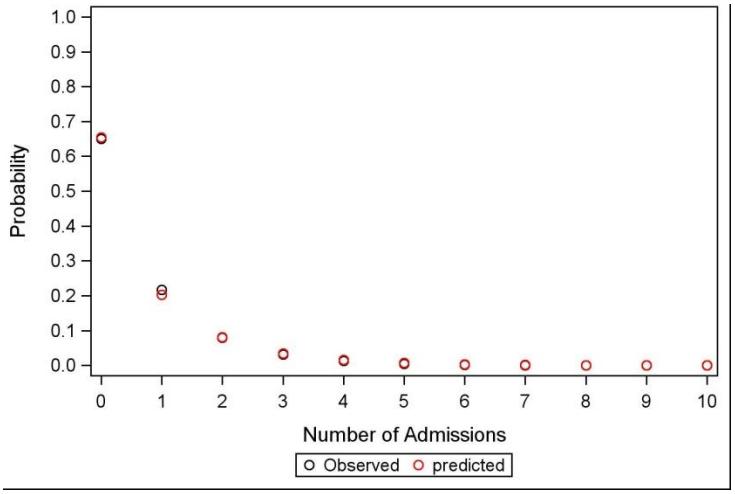


Figure 9. Observed vs. predicted probabilities for the number of hospital admissions among second highest admission group (80 to 130 admissions per 100 person-years, median: 99 IQR: 89 to 113.

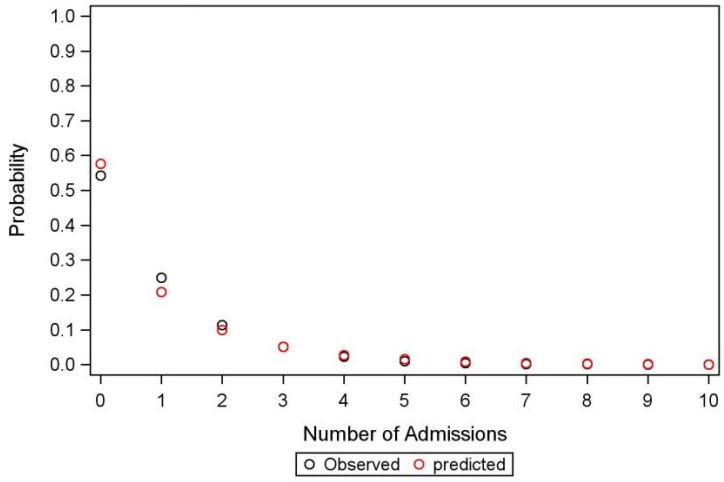
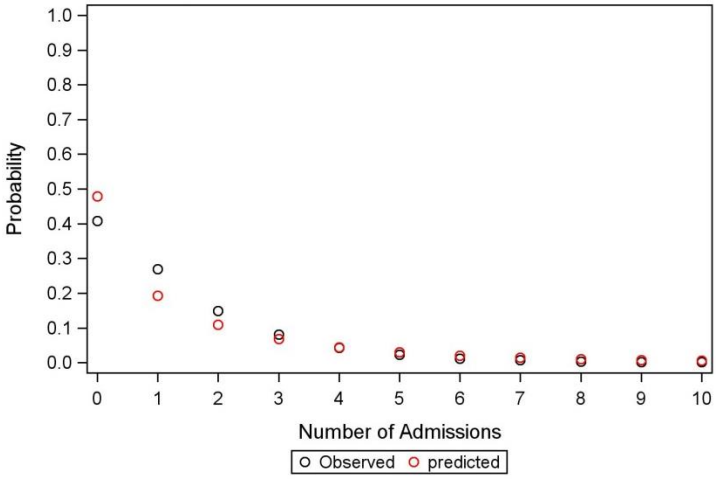


Figure 10. Observed vs. predicted probabilities for the number of hospital admissions among highest admission group (129 to 1,168 admissions per 100 person-years, median: 183, IQR: 151 to 238.



2b4.9. Results of Risk Stratification Analysis:

Not applicable. This measure is not risk stratified.

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*)

Model performance was similar in the development and validation datasets, with strong model discrimination and fit. The over-fitting index of γ_0 close to zero and γ_1 close to one indicates good calibration of the model. Additionally, the risk decile plots of all four risk groups show that the model performs well across a broad range of risk. In the highest risk group, we observed that the observed and predicted probabilities of the number of zero, one, or two admissions differed slightly. However, these differences were small and somewhat expected among the highest risk group of patients.

Citations

1. Cameron AC, Windmeijer FA. R-squared measures for count data regression models with applications to health-care utilization. Journal of Business & Economic Statistics. 1996;14(2):209-220.

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*)

Not applicable.

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)*

The method for discriminating facility-level performance for public reporting has not been determined. For publicly reported readmission measures of hospital outcomes developed with similar methodology, CMS currently estimates an interval estimate for each risk-standardized rate to characterize the amount of uncertainty associated with the rate, compares the interval estimate to the national crude rate for the outcome, and categorizes hospitals as .better than the national rate,' 'worse than the national rate,' or 'no different than that national rate.' We used that approach here. However, the approach to discriminating performance that would be used for this measure in public reporting has not been determined. In order to determine interval estimates (IEs), we used bootstrapping methods. In brief, we randomly sampled 114 ACOs with replacement. This is done by randomly selecting an ACO from the 114 ACOs, then placing the selected ACO back into the pool, until we reached 114 ACOs, with some ACOs being selected more than once. Performance scores were calculated for each random sample of 114 ACOs. If some ACOs were selected more than once in a bootstrapped sample, we treated them as distinct so that we had random effects to estimate the variance components. This process was repeated many times until 3,000 results were obtained for each ACO. Using the 95% IE estimates, we assigned each ACO to one of three performance categories: 'better than the national rate,' 'no different than national rate,' and 'worse than national rate.' Each ACO was compared to all Medicare FFS beneficiaries who met our heart failure cohort criteria, so that each ACO was evaluated against the US national admission rate among Medicare FFS patients with heart failure. The ACO was 'better than the national rate' if the 95% IE was completely below the US national Medicare FFS rate among patients with heart failure; 'no different than the national rate' if the 95% IE included the US national Medicare FFS rate among patients with heart failure; and 'worse than the national rate' if the 95% IE was above the US national Medicare FFS rate among patients with heart failure.

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*)

61 (54%) ACOs performed no different than the national rate, 37 (32%) performed better than the national rate, and 16 (14%) performed worse than the national rate of admissions per person-years at risk for hospitalization among the US national Medicare FFS heart failure patient population.

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?*)

These results suggest there are meaningful differences in the quality of care received for patients in the 114 ACOs in the ambulatory setting.

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.***

Items 2b6.1-2b6.3 skipped, as this measure has only one set of specifications.

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*)

Not applicable.

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)

Not applicable.

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*)

Not applicable.

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*)

Not applicable.

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*)

Not applicable.

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*)

Not applicable.

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.**
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)*.
Not applicable. There are no fees, licensing, or other requirements to use any aspect of the measure as specified.

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)
Payment Program	

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

Measure is currently not in use.

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?*)

This measure is not currently publicly reported or used in an accountability application because it only recently completed development. However, in the November 13, 2014 Physician Fee Schedule final rule, CMS finalized adding the measure to the Medicare Shared Savings Program quality measure set (see 79 FR 67912; https://www.gpo.gov/fdsys/pkg/FR-2014-11-13/pdf/2014-26183.pdf).

The measure is planned for pay-for-reporting in the Medicare Shared Savings Program for 2015 and 2016 reporting periods (79 FR 67912, 67916) and for pay-for-performance in the Medicare Shared Savings Program beginning 2017 reporting period (79 FR 67912, 67916).

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 will be used in one or more CMS programs as noted above in 4a.2. The measure has been finalized for use in the Medicare Shared Savings Program. The measure will be pay-for-reporting initially for the 2015 and 2016 reporting periods and then as pay-for-performance beginning in the 2017 reporting period.

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**

Not applicable.

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.

The measure is not currently used in a quality improvement program, but the primary goal of the measure is to provide ACOs with information necessary to implement focused quality improvement.

This measure was evaluated by a group of clinical experts and a technical expert panel (TEP) throughout the measure development process. We received input and feedback on key methodological, clinical, and other measure decisions as well as on its utility in guiding focused quality improvement within ACOs.

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.**

In designing the measure, we sought to minimize the potential of this measure to result in the denial of future care to high-risk individuals. We developed the patient cohort exclusions and risk-adjustment model to ensure providers who care for patients at higher risk of admission will not be disadvantaged in the measure. CMS is committed to monitoring this measure's use and assessing potential unintended consequences over time.

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)**

0277 : Heart Failure Admission Rate (PQI 8)

0709 : Proportion of patients with a chronic condition that have a potentially avoidable complication during a calendar year.

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?

No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**

The measures listed above are NQF-endorsed. There are several differences between our measure and these two NQF measures. 1. The cohort populations are different. The NQF measures focus on patients aged 18-65 years and 18+ years, respectively, for the two measures; thus, the cohorts have limited overlap. 2. The risk-adjustment models are different. NQF #0709 is not risk-adjusted; NQF #0277 is risk-adjusted for age and sex only, while our measures are fully risk-adjusted. 3. The outcomes measured

(NQF 0709: potentially avoidable complications; NQF 0277: heart failure admissions) are different from our outcome of acute, all-cause admission rates.

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.)
Not applicable.

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:** Heart_Failure_ACO_Admission_Measure_NQF_Appendix_01-29-16_v1.0.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services (CMS)
Co.2 **Point of Contact:** Vinitha, Meyyur, Vinitha.meyyur@cms.hhs.gov, 410-786-8819-
Co.3 **Measure Developer if different from Measure Steward:** Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (CORE)
Co.4 **Point of Contact:** Elizabeth, Drye, Elizabeth.drye@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.
TEP Members:

CORE convened a TEP of clinicians, patients, purchasers, and experts in quality improvement to provide input on key methodological decisions.

Lawrence M. Becker, BS – Xerox Corporation (Director, Strategic Partnerships, Alliances and Analytics)
Alex Blum, MD, MPH – Evergreen Health Cooperative (Chief Medical Officer)
Sanjay Doddamani, MD – Geisinger Health System (System-wide Chief of Advanced Cardiac Disease HF)
Kevin Fiscella, MD, MPH – University of Rochester Medical Center (Professor of Family Medicine)
Elbert Huang, MD, MPH – University of Chicago (Associate Professor of Medicine, Director of the Center for Translational and Policy Research of Chronic Diseases, and Associate Director of the Chicago Center for Diabetes Translation Research)
Bruce Leff, MD – Johns Hopkins University School of Medicine (Professor of Medicine, Division of Geriatric Medicine); The Johns Hopkins University Bloomberg School of Public Health (Faculty, Health Services Research Development Center and Lipitz Center for Integrated Health Care)
Andy Miller, MD, MPH – Healthcare Quality Strategies, Inc. (Medical Director);Colorado Foundation for Medical Care (CMO, Integrating Care for Populations & Communities National Coordinating Center)
Ami Parekh, MD, JD – University of California, San Francisco (Medical Director for Health System Innovation)

Christine Ritchie, MD – University of California, San Francisco (Professor of Medicine, Division of Geriatrics)
Two patients with chronic conditions (anonymous)

CORE Measure Development Team:
Faseeha Altaf, MPH – Research Project Coordinator
Haikun Bao, PhD – Lead Analyst, diabetes measure
Susannah Bernheim, MD, MHP – Director of CMS Projects; Clinical Investigator
Kanchana Bhat, MPH – Senior Project Manager
Ying Dai, PhD – Lead Analyst
Weiwei Zhang, MPH – Supporting Analyst
Elizabeth Drye, MD, SM – Project Director; Project Lead, MCCs measure
Elizabeth Eddy, BA – Research Project Coordinator
Leora Horwitz, MD, MHS – Clinical Investigator
Erin Joyce, BA – Research Assistant
Zhenqiu Lin, PhD – Managing Analyst
Harlan Krumholz, MD, SM – Director, CORE
Kasia Lipska, MD, MHS – Project Lead
Julia Montague, MPH – Research Project Coordinator II/Project Manager
Craig Parzynski, MS – Supporting Analyst
Joseph Ross, MD, MHS – Clinical Investigator, CORE
Erica Spatz, MD, MHS – Project Lead
La'Mont Sutton, MPH – Research Associate
Vera Zhang, MPH – Supporting Analyst

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:**
Ad.3 **Month and Year of most recent revision:**
Ad.4 **What is your frequency for review/update of this measure?** Not applicable.
Ad.5 **When is the next scheduled review/update for this measure?**

Ad.6 **Copyright statement:** Not applicable.
Ad.7 **Disclaimers:** Not applicable.

Ad.8 **Additional Information/Comments:** Not applicable.

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 #: 2887
De.2. **Measure Title:** Risk-Standardized Acute Admission Rates for Patients with Diabetes
Co.1.1. **Measure Steward:** Centers for Medicare & Medicaid Services (CMS)
De.3. **Brief Description of Measure:** Rate of risk-standardized acute, unplanned hospital admissions among Medicare fee-for-service (FFS) patients 65 years and older with diabetes
1b.1. **Developer Rationale:** The goal of this measure is to evaluate and to improve the quality of care for patients with diabetes cared for by ACOs. These patients account for a significant proportion of Medicare beneficiaries and they experience high morbidity and costs associated with their disease. These patients need efficient, coordinated, and patient-centered care management. They also benefit from provider support and infrastructure that facilitate effective chronic disease management. This measure is focused on hospital admissions for acute illness as the outcome because these admissions are often sentinel events associated with high morbidity as well as physical and emotional stress; they also result in high costs for both the patient and the ACO. Research shows that effective health care can lower the risk of admission for this vulnerable group of patients.

This measure is intended to incentivize ACOs to provide high-quality, coordinated care that focuses on the whole patient. ACOs were conceptualized and created to achieve the goals of improved care, improved population health and lower cost. Consistent with this mission, we envision that the measure will incentivize providers participating in ACOs to collaborate to provide the best system of clinical care and to partner with health and non-health related organizations in their communities as appropriate to improve the health of their patient population.

References:

Centers for Medicare & Medicaid Services (CMS). Medicare Health Support. 2012; https://www.cms.gov/Medicare/Medicare-General-Information/CCIP/. Accessed March 27, 2014.

Chen JY, Tian H, Taira Juarez D, et al. The effect of a PPO pay-for-performance program on patients with diabetes. The American journal of managed care. Jan 2010;16(1):e1119.

Brown RS, Peikes D, Peterson G, Schore J, Razafindrakoto CM. Six features of Medicare coordinated care demonstration programs that cut hospital admissions of high-risk patients. Health Affairs. 2012 Jun 2012;31(6):1156-1166.

Leong A, Dasgupta K, Bernatsky S, Lacaille D, Avina-Zubieta A, Rahme E. Systematic review and meta-analysis of validation studies on a diabetes case definition from health administrative records. PloS one. 2013;8(10):e75256.

McCarthy D, Cohen A, Johnson MB. Gaining Ground: Care Management Programs to Reduce Hospital Admissions and Readmissions Among Chronically Ill and Vulnerable Patients. The Commonwealth Fund, New York. 2013.

Patient Protection and Affordable Care Act, 42 U.S.C., §3022 (2010).

Sadur CN, Moline N, Costa M, et al. Diabetes management in a health maintenance organization. Efficacy of care management using cluster visits. Diabetes care. Dec 1999;22(12):2011-2017.

Numerator Statement: The outcome measured for each patient is the number of acute, unplanned admissions per 100 person-years at risk for admission. Persons are considered at risk for admission if they are alive, enrolled in FFS Medicare, and not currently

admitted. (See S.6, Numerator Details, for more information.)
S.7. **Denominator Statement:** The target population is ambulatory Medicare FFS patients aged 65 years and older with a diagnosis of diabetes.
S.10. **Denominator Exclusions:** The measure excludes:
1. Patients without continuous enrollment in Medicare Part A for the duration of the measurement period (or until death).
Rationale: We exclude these patients to ensure full data availability for outcome assessment (Part A during the measurement year).

Measure Type: Outcome
S.23. **Data Source:** Administrative claims
S.26. **Level of Analysis:** Integrated Delivery System

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

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? Not applicable

New Measure -- Preliminary Analysis

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.

The developer provides the following evidence for this outcome measure:

- The developer notes that specific system-based interventions such as participation in group outpatient visits with a diabetes nurse education has been associated with lower all-cause hospitalization rates among diabetes patients.

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

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 developer provided data from ACO performance score using the 2012 Medicare Full Sample which showed the mean risk-standardized acute admission rate (RSAAR) among ACOs for year 2012 is 39.6, median is 39.1.
- They observed that 51 ACOs (44.7%) had RSAARs that were 'no different than the national rate' and 45 ACOs (39.5%) had RSAAR scores 'better than the national rate,' and 18 ACOs (15.8%) were 'worse than the national rate.'

Disparities

- The developer reports that they examined disparities in ACO performance based on the proportion of patients of low socioeconomic status (SES); using two variables, from the Agency for Healthcare Research and Quality (AHRQ)

SES Index and patient Medicare and Medicaid dual-eligibility status.

- The developer found that performance scores did not change appreciably after adjusting the models for patients' SES. The Spearman correlation comparing the ACO measure scores estimated with and without risk adjustment for the AHRQ SES Index was 0.981. Similarly, the Spearman correlation for the scores estimated with and without patients' Medicaid dual eligibility was 0.976. These results demonstrate that adjusting for SES at the patient level has little effect on the measure score.
- Overall, results indicate that SES status plays little role at the patient level, thus measure was not adjusted for patient-level SES. According to the developer, ACOs should and do influence a broad range of patient and community-level factors that can mitigate the risk of admission associated with low SES, and do not want to adjust for modifiable factors.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Given the developer disparities testing results, does the Committee agree that SDS adjustment is not warranted?*

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)

1. Importance to Measure and Report

1a. Evidence to Support Measure Focus

Comments: **The evidence is appropriate. The outcome is straight forward. There are actions that could potentially impact this measure, although it is likely multi-factorial.

1b. Performance Gap

Comments: **Data was provided. There does appear to be a performance gap at the extremes (5/95). Data on disparities would suggest, at least on its face, that the curve is shifted towards lower performance with more disparities.

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

Comments: **N/A

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

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): Administrative claims

Specifications:

- This measure calculates the rate of risk-standardized acute, unplanned hospital admissions among Medicare fee-for-service (FFS) patients 65 years and older with diabetes
- This is a health outcome measure and the level of analysis is Integrated Delivery System.
- The Numerator is the number of acute, unplanned admissions per 100 person-years at risk for admission. Persons are considered at risk for admission if they are alive, enrolled in FFS Medicare, and not currently admitted.
- The Denominator is ambulatory Medicare FFS patients aged 65 years and older with a diagnosis of diabetes.

Questions for the Committee :

○ *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.

Method(s) of reliability testing

- Datasets used for testing included Medicare Parts A and B claims, the denominator file, the Medicare Provider Analysis and Review (MedPAR) file, and the American Community Survey to derive the AHRQ SES index.
- **Data element reliability**:
  - With regard to data element reliability, the developer notes that the measure has been developed to avoid the use of claims data elements that are thought to be coded inconsistently across hospitals or providers, instead using fields that are consequential for payment and which are audited by CMS.
  - In addition, the developer compared frequencies and odds ratios of variables from their risk model to assess the consistency of those variables across samples.
- **Performance score reliability:**
  - The developer defines performance score reliability as the degree to which repeated measurements of the same entity agree with each other.
  - In line with this thinking, the developer's approach to assessing score-level reliability was to consider the extent to which assessments of a hospital using different but randomly-selected subsets of patients produce similar measures of hospital performance. The developers refer to this as a "test-retest" approach; it may also be called a "split-half" method. This is generally considered an appropriate method of testing reliability.

Results of reliability testing

- **Data element reliability**:
  - Summarizing the results of this analysis, the developer notes that the mean age and frequency of risk-adjustment variables was similar among the two samples of 2012 data suggesting that the data elements are reliable across the samples.
- **Performance score reliability:**
  - The 2012 full Medicare sample was divided into two subsets of patients randomly. The developer calculated the measure score of all ACOs for each of the two subsets of patients. Each ACO was measured twice, but each measurement was make using distinct sets of measures. The interclass correlation coefficient (ICC) for the two subsets of patients was 0.889, which can be interpreted as excellent correlation, and thus reliable.

Guidance from the Reliability Algorithm

- Question 1. Submitted specifications are precise, unambiguous, and complete. Measure can be consistently implemented.
- Question 2. Empirical reliability testing was conducted using statistical tests with the measure as specified.
- Question 3. Empirical validity testing of patient-level data was conducted.
- Question 4. Reliability testing was conducted with computed performance measure scores for each measured entity.
- Question 5. Random split-half correlation was used to assess the proportion of variability due to real differences among the measured entities.
- Question 6. The ICC was 0.889 which is considered an excellent level of agreement.

Questions for the Committee:

○ *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

○ *Does the measure testing match the measure specifications?*

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

2b. Validity

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**

- This measure estimates the predicted number of admissions given the Accountable Care Organization's (ACO's) case mix, sample size, and actual admission rate. The outcome for this measure is the number of acute, unplanned admissions per 100 person-years at risk for admission. The outcome includes inpatient admissions to an acute care hospital for any cause during the measurement year, unless an admission is identified as "planned."

Question for the Committee:

○ *Are the specifications clear?*

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.

- The developer tested the validity of the measure using three different methods:
  - Validity of the claims-based measures. The developer argues that other NQF endorsed mortality and readmission measures have been validated by comparing the claims to the medical records data elements. It is unclear if the risk adjustment validation approach that the developer cites is sufficiently similar to this measure and for this level of analysis and ambulatory patients.
  - The developer also notes that this measure has been validated by using established measure development guidelines. While an important step for measure development, this method of validity testing has generally not be considered sufficient for demonstrating measure validity.
  - Finally, the measure developer completed a systemic face validity assessment of this measure with 9 experts and two patients agreeing that this measure was a valid indicator of health care quality.

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**

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:

- Out of the total Medicare FFS patients with diabetes (N=6,746,776), the developer excluded 225,314 due to non-continuous enrollment in part A in 2012.

Questions for the Committee:

○ *Are the exclusions consistent with the evidence?*

○ *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** ☒ **No**

Risk adjustment summary

- The developers provided a conceptual framework that was used to develop the risk adjustment model for this measure. This conceptual framework included 4 contextual domains that influence ACO performance including, physical environment, community resources, patient resources, and patient behavioral/personal preferences.
- The measure included demographic factors, and clinical risk factors present at the start of the measurement period.
    - The measure developers reviewed 189 diagnosis groups included in the hierarchical condition category (HCC), and calculated the prevalence of each CC in the year preceding the measurement period. After examining the bi-variate analysis, the developers reduced the list to 22 candidate variables including age.
    - The measure developers did not adjust for contextual factors that impact admissions; however, they did provide data demonstrating that including SDS adjustment did not make a meaningful difference to the measure score of the ACOs. The spearman correlation coefficient that estimated the difference in performance with and without SDS adjustment was 0.981. Thus, the results demonstrate that adjustment had little effect on the measure score.
- **Risk Model Diagnostics:**
    - To assess the overall performance of their risk-adjustment model, the developers computed two summary statistics, including:
        - *Risk model discrimination statistics* (the model's ability to explain how successful the fit is in explaining the variation of the data. In this case, the r-sq value was 0.218. In other words, the model was able to explain 21.8% of the total deviance.
        - Overfitting indices (model calibration) [presented as (γ0, γ1)]:
            - The developer states that if the γ0 in the validation samples are substantially far from zero and the γ1 is substantially far from one, there is potential evidence of over-fitting. The calibration value of close to 0 at one end and close to 1 to the other end indicates good calibration of the model.
                - 2012 Development Sample (Index): (**0,1**)
                - 2012 Validation Sample: **(0.0017, 1.0031)**

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 note that the methodology to publicly report this measure has not been determined yet
- For other publically reported measures with the same methodology, CMS categories hospitals at "better than the national rate", "worse than the national rate" and "no different than the national rate".
    - For this measure, 51 ACOs (44.7%) performed no different than the national rate, 45 (39.5%) performed better then the national rate, and 18 (15.8%) performed worse than the national rate. The developers suggest that this demonstrates that there is a meaningful different in performance on this measure.

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

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)

2. Scientific Acceptability of Measure Properties

2a1. & 2b1. Specifications

Comments: **Specifications are appropriate.

2a2. Reliability Testing

Comments: **The split-half method is appropriate. The correlation of 0.889 is OK.

2b2. Validity Testing

Comments: **Data is appropriate. TEP panel response is somewhat tepid.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **The face validity of the SES projections of where ACOs would land (worse, at, or better than expected) seem to suggest that SES - with duals or SES - makes a difference. I understand that the correlation does not suggest the same.
Also, the fact that almost as many ACOs performed better than expected than performed at expected is somewhat concerning.
Finally, the R-squared of 0.218 seems low to me.

Criterion 3. Feasibility

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.

The developer states:

- ALL measure data elements are in defined fields in electronic claims and routinely generated or collected by and used by healthcare personnel during the provision of care, coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims).
- There is no cost associated with data collection.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Is the data collection strategy ready to be put into operational use?*

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

Committee pre-evaluation comments
Criteria 3: Feasibility

3. Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **There are not feasibility issues.

Criterion 4: Usability and Use

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**
OR
Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

The developer states:

- This measure was included by CMS in the November 13, 2014 Physician Fee Schedule final rule, and finalized adding the measure to the Medicare Shared Savings Program quality measure set (see 79 FR 67912; https://www.gpo.gov/fdsys/pkg/FR-2014-11-13/pdf/2014-26183.pdf).
- The measure is planned for pay-for-reporting in the Medicare Shared Savings Program for 2015 and 2016 reporting periods (79 FR 67912, 67916) and for pay-for-performance in the Medicare Shared Savings Program beginning 2017 reporting period (79 FR 67912, 67916).

Improvement results: N/A

Potential harms

The developer states:

- To minimize the potential of this measure to result in the denial of future care to high-risk individuals, they developed the patient cohort exclusions and risk-adjustment model to ensure providers who care for patients at higher risk of admission will not be disadvantaged in the measure. CMS is committed to monitoring this measure's use and assessing potential unintended consequences over time.

***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

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **The measure is usable. It would be an excellent measure to be used by ACOs or those responsible for large populations of patients (health plans, delivery systems, primary care practices) to spur quality improvement. The use of this measure for pay-for-performance (which as noted is already mentioned in a proposed rule from CMS) is suspect given the noted validity issues.

Criterion 5: Related and Competing Measures

Related or competing measures

0272 : Diabetes Short-Term Complications Admission Rate (PQI 01)
0274 : Diabetes Long-Term Complications Admission Rate (PQI 03)
0638 : Uncontrolled Diabetes Admission Rate (PQI 14)

Harmonization

The developer provides the following information:

- The measures listed above have different outcomes, target populations and risk-adjustment models.
- The existing measures are either not adjusted or adjusted for age and sex. Measure #2887 is fully adjusted for a broad range of clinical factors that contribute to the risk for admission, allowing for fair comparisons of ACO performance.
- Existing measures include adults with ages 18 to 75 or 18 to 65 years of age. For Measure #2887, the target population includes all Medicare FFS beneficiaries with a diagnosis of diabetes, who are 65 years or older, so focus is on older, complex adults with diabetes.

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: Risk-Standardized Acute Admission Rates for Patients with Diabetes

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: 1/29/2016

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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: Click here to name the health outcome

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

Patients with diabetes are vulnerable to complications that result from their underlying disease, as well as to a range of other acute illnesses, placing them at relatively high risk for hospitalization. Provision of coordinated care that is focused on improving health for the whole patient, across all stages of disease, and in the context of coexisting comorbidities and life circumstances should lower the risk of hospital admission for these patients.

To provide high-quality care for patients with chronic conditions, health systems must effectively prevent and manage the complications of chronic disease as well as other related and unrelated illnesses that frequently arise among patients with chronic disease. For more than a decade, we have known that admission rates vary across the country, even after adjusting for differences in patient populations. To date, however, admission rates have been used as quality and accountability measures to only a limited degree. For example, it is only recently that the Centers for Medicare & Medicaid Services (CMS) has started to use admission scores developed by the Agency for Healthcare Research and Quality (AHRQ), known as Prevention Quality Indicators (PQIs), in several of its programs. These admission scores, however, are narrowly focused and measure only disease-specific admissions among populations defined by the disease (e.g., heart failure admissions among patients with heart failure). They do not capture the wide spectrum of hospital admissions for which patients with chronic conditions are at increased risk.

This measure of acute, unplanned admission rates will illuminate differences in the quality of care delivered by ACOs and drive efforts to improve prevention and management strategies, including the efficiency and coordination of care.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

Research shows that effective health care can lower the risk of admission for patients with diabetes [1-7]. For example, specific system-based interventions such as seeing a physician involved in a pay-for-performance program for diabetes care or participation in group outpatient visits with a diabetes nurse educator have been associated with lower all-cause hospitalization rates among these patients [8]. It is our vision that this measure will illuminate variation in hospital admission rates and incentivize ACOs to develop efficient and coordinated chronic disease management strategies that anticipate and respond to patients' needs and preferences. This vision is consistent with ACOs' commitment to deliver patient-centered care that fulfills the goals of the Department of Health and Human Service's Triple Aim – improving population health, improving care, and lowering care costs.

References:

1. Patient Protection and Affordable Care Act, 42 U.S.C., §3022 (2010).

2. Centers for Medicare & Medicaid Services (CMS). Medicare Health Support. 2012; https://www.cms.gov/Medicare/Medicare-General-Information/CCIP/. Accessed March 27, 2014.

3. Chen JY, Tian H, Taira Juarez D, et al. The effect of a PPO pay-for-performance program on patients with diabetes. *The American journal of managed care.* Jan 2010;16(1):e1119.

4. Sadur CN, Moline N, Costa M, et al. Diabetes management in a health maintenance organization. Efficacy of care management using cluster visits. *Diabetes care.* Dec 1999;22(12):2011-2017.

5. Brown RS, Peikes D, Peterson G, Schore J, Razafindrakoto CM. Six features of Medicare coordinated care demonstration programs that cut hospital admissions of high-risk patients. *Health Affairs.* 2012 Jun 2012;31(6):1156-1166.

6. McCarthy D, Cohen A, Johnson MB. Gaining Ground: Care Management Programs to Reduce Hospital Admissions and Readmissions Among Chronically Ill and Vulnerable Patients. *The Commonwealth Fund, New York.* 2013.

7. Leong A, Dasgupta K, Bernatsky S, Lacaille D, Avina-Zubieta A, Rahme E. Systematic review and meta-analysis of validation studies on a diabetes case definition from health administrative records. *PloS one.* 2013;8(10):e75256.

8. Levine S, Steinman BA, Attaway K, Jung T, Enguidanos S. Home care program for patients at high risk of hospitalization. *American Journal of Managed Care.* 2012 Aug 2012;18(8):e269-276.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

Not applicable. This is an outcome measure.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Not applicable. This is an outcome measure.

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Not applicable. This is an outcome measure

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Not applicable. This is an outcome measure.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

Not applicable. This is an outcome measure.

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

Not applicable. This is an outcome measure.

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

Not applicable. This is an outcome measure.

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

Not applicable. This is an outcome measure.

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

Not applicable. This is an outcome measure.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

Not applicable. This is an outcome measure.

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

Not applicable. This is an outcome measure.

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Not applicable. This is an outcome measure.

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

Not applicable. This is an outcome measure.

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Not applicable. This is an outcome measure.

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

Not applicable. This is an outcome measure.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

Not applicable. This is an outcome measure.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

Not applicable. This is an outcome measure.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

Not applicable. This is an outcome measure.

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

Not applicable. This is an outcome measure.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Not applicable. This is an outcome measure.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

Not applicable. This is an outcome measure.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

Not applicable. This is an outcome measure.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

Not applicable. This is an outcome measure.

1a.8 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.

Not applicable. This is an outcome measure.

1a.8.1 What process was used to identify the evidence?

Not applicable. This is an outcome measure.

1a.8.2. Provide the citation and summary for each piece of evidence.

Not applicable. This is an outcome measure.

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
Diabetes_ACO_Admission_Measure_NQF_Evidence_Form_01-29-16_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 evaluate and to improve the quality of care for patients with diabetes cared for by ACOs. These patients account for a significant proportion of Medicare beneficiaries and they experience high morbidity and costs associated with their disease. These patients need efficient, coordinated, and patient-centered care management. They also benefit from provider support and infrastructure that facilitate effective chronic disease management. This measure is focused on hospital admissions for acute illness as the outcome because these admissions are often sentinel events associated with high morbidity as well as physical and emotional stress; they also result in high costs for both the patient and the ACO. Research shows that effective health care can lower the risk of admission for this vulnerable group of patients.

This measure is intended to incentivize ACOs to provide high-quality, coordinated care that focuses on the whole patient. ACOs were conceptualized and created to achieve the goals of improved care, improved population health and lower cost. Consistent with this mission, we envision that the measure will incentivize providers participating in ACOs to collaborate to provide the best system of clinical care and to partner with health and non-health related organizations in their communities as appropriate to improve the health of their patient population.

References:

Centers for Medicare & Medicaid Services (CMS). Medicare Health Support. 2012; https://www.cms.gov/Medicare/Medicare-General-Information/CCIP/. Accessed March 27, 2014.

Chen JY, Tian H, Taira Juarez D, et al. The effect of a PPO pay-for-performance program on patients with diabetes. The American journal of managed care. Jan 2010;16(1):e1119.

Brown RS, Peikes D, Peterson G, Schore J, Razafindrakoto CM. Six features of Medicare coordinated care demonstration programs that cut hospital admissions of high-risk patients. Health Affairs. 2012 Jun 2012;31(6):1156-1166.

Leong A, Dasgupta K, Bernatsky S, Lacaille D, Avina-Zubieta A, Rahme E. Systematic review and meta-analysis of validation studies on a diabetes case definition from health administrative records. PloS one. 2013;8(10):e75256.

McCarthy D, Cohen A, Johnson MB. Gaining Ground: Care Management Programs to Reduce Hospital Admissions and Readmissions Among Chronically Ill and Vulnerable Patients. The Commonwealth Fund, New York. 2013.

Patient Protection and Affordable Care Act, 42 U.S.C., §3022 (2010).

Sadur CN, Moline N, Costa M, et al. Diabetes management in a health maintenance organization. Efficacy of care management using cluster visits. Diabetes care. Dec 1999;22(12):2011-2017.

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.
We report the variation in ACO performance scores using the 2012 Medicare Full Sample (see Section 1.7 for information about sample).

There were 6,521,462 patients in the 2012 Medicare Full Sample who met our inclusion and exclusion criteria for the measure cohort. Among these, there were 341,193 patients in 114 ACOs.

The crude US national Medicare FFS rate of acute, unplanned admissions per person-year among patients with diabetes was 41.4 admissions per 100 person-years.

Among ACOs, the mean RSAAR for calendar year 2012 was 39.6 admissions per 100 person-years (standard deviation = 7.3). The median RSAAR was 39.1 admissions per 100 person-years (interquartile range [IQR] 34.8 to 43.9). The minimum RSAAR was 23.9 per 100 person-years; the 5th percentile was 28.6 admissions per 100 person-years; the 95th percentile was 53.0 admissions per 100 person-years; and maximum RSAAR was 68.1 admissions per 100 person-years.

We observed that 51 ACOs (44.7%) had RSAARs that were 'no different than the national rate.' An additional 45 ACOs (39.5%) had RSAAR scores 'better than the national rate,' and 18 ACOs (15.8%) were 'worse than the national rate.'

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.**
Not applicable.

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 examined disparities in ACO performance based on the proportion of patients of low socioeconomic status (SES) being cared for by each ACO.

Identification of ACOs caring for few and many 'low SES' patients
We identified low SES patients using two variables: the Agency for Healthcare Research and Quality (AHRQ) SES Index and patient Medicare and Medicaid dual-eligibility status.
Using the AHRQ SES Index (described in the NQF Testing form, Section 2b4.3 and Appendix E of the attached technical report), which is a continuous variable, we created a dichotomous low-SES variable by assessing the distribution of SES scores across a broad sample of Medicare FFS beneficiaries, labeling patients with the lowest 20% of scores as "low SES" (see Testing Form, Section 1.8, for further details). We then categorized ACOs into quartiles based on the proportion of low SES patients in their cohort (first quartile (Q1) = 'few' low SES patients, fourth quartile (Q4) = 'many' low SES patients).

Similarly, we categorized ACOs by the proportion of Medicaid dual-eligible patients in their cohort into ACOs caring for 'few' (first quartile) and 'many' (fourth quartile) Medicaid dual-eligible patients.

Results: AHRQ SES and Medicaid Dual-Eligibility Analyses

Using the AHRQ SES Index, for the 29 ACOs in Q1, the proportion of low SES patients ranged from 0.0% to 4.5%; for the 28 ACOs in Q4, the proportion of low SES patients ranged from 28.7% to 96.6%.

Among the 29 ACOs caring for few low SES patients (Q1), 1 (3.4%) performed 'worse than the national rate,' 15 (51.7%) performed 'no different than the national rate,' and 13 (44.8%) performed 'better than the national rate.' Among the 28 ACOs caring for many low SES patients (Q4), 9 (32.1%) performed 'worse than the national rate,' 11 (39.3%) performed 'no different than the national rate,' and 8 (28.6%) performed 'better than the national rate.' (See attached Technical Report, Table 10).
Using Medicaid dual eligibility as an indicator of low SES, among the 29 ACOs caring for few Medicaid dual-eligible patients (Q1), the proportion of Medicaid dual-eligible patients ranged from 2.4 to 7.5%; among the 28 ACOs caring for the most Medicaid dual-eligible patients (Q4) the proportion of Medicaid dual-eligible patients ranged from 16.3 to 78.7%.

Among the 29 ACOs with few Medicaid dual-eligible patients (Q1), 1 (3.4 %) performed worse than the national rate, 12 (41.4%) performed 'no different than the national rate,' and 16 (55.2%) performed 'better than the national rate.' Among the 28 ACOs with many Medicaid dual-eligible patients (Q4), 8 (28.6%) performed 'worse than the national rate,' 13 (46.4%) performed 'no different than the national rate,' and 7 (25.0%) performed 'better than the national rate.' (See attached Technical Report, Table 9).
The distribution of RSAARs across ACOs caring for increasing proportions of low SES patients reveals two patterns: (1) ACOs in Q1 (few low SES patients) tend to have lower RSAARs than ACOs in Q4 (many low SES patients); (2) there is more variation in RSAARs among ACOs in Q4 as compared with ACOs in Q1-Q3. There are small differences in these patterns when analyses are performed using Medicaid dual eligibility as an indicator of SES status (see Figure 15 of the attached technical report).

Socioeconomic Status Interpretation
Among a group of 114 ACOs, there is substantial variation in performance among ACOs caring for many (fourth quartile) and few (first quartile) low SES patients. ACOs serving many low SES patients more often perform worse than the national rate compared with ACOs serving few low SES patients. This was true using either the AHRQ SES index (32.1% vs. 3.4%, respectively) or Medicaid dual-eligibility status (28.6% vs. 3.4 %, respectively) as an indicator of patients' SES. However, among ACOs serving many low SES patients, using the AHRQ SES index, 8 ACOs (28.6%) performed 'better than the national rate;' using Medicaid dual-eligibility status as an indicator, 7 ACOs (25.0%) performed 'better than the national rate.'

We also found that performance scores did not change appreciably after adjusting the models for patients' SES. As demonstrated in the Testing Form, Section 2b4.4b, the Spearman correlation comparing the ACO measure scores estimated with and without risk adjustment for the AHRQ SES Index was 0.981. Similarly, the Spearman correlation for the scores estimated with and without patients' Medicaid dual eligibility was 0.976. These results demonstrate that adjusting for SES at the patient level has little effect on the measure score.

We did not adjust the measure for patient-level SES. Conceptually, ACOs should and do influence a broad range of patient and community-level factors that can mitigate the risk of admission associated with low SES, and we do not want to adjust for modifiable factors. Empirically, our results indicate that SES status plays little role at the patient level.

References:

Wynn B. Analysis of the Joint Distribution of Disproportionate Share Hospital Payments. 2002.

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

Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian SV, Carson R. Choosing area based socioeconomic measures to monitor social inequalities in low birth weight and childhood lead poisoning: The Public Health Disparities Geocoding Project (US). J Epidemiol Community Health. 2003a Mar;57(3):186-99

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.**
Not applicable. Data on disparities are presented above.

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
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.

Diabetes is a complex, high-prevalence chronic disease that affects 18% of Medicare beneficiaries. It has a strong impact on functional status and daily living of affected people, and Medicare diabetes beneficiaries are responsible for 32% of Medicare spending [1]. Patients with diabetes are vulnerable to complications that result from their underlying disease, as well as to a range of other acute illnesses, placing them at relatively high risk for hospitalization [2-3]. Provision of coordinated care that is focused on improving health for the whole patient, across all stages of disease, and in the context of coexisting comorbidities and life circumstances should lower the risk of hospital admission for these patients [2-12].

Specific to the diabetes cohort assessed for this measure, in the 2012 Medicare Full Sample, there were 6,521,462 patients who met the criteria for diabetes, among which 341,193 (5.2%) were assigned to one of 114 ACOs. The crude rate of acute, unplanned hospital admissions was 41.4 per 100 person-years among all Medicare FFS diabetes beneficiaries, and 38.9 per 100 person-years among diabetes beneficiaries assigned to an ACO. The average RSAAR among ACOs was 39.6 (range of 23.9 to 68.1) per 100 person-years at risk for hospitalization. These rates illustrate the high morbidity associated with this condition, the variation in ACO performance, and the opportunity to reduce hospitalizations, improve care and potentially lower costs.

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

1. Ashkenazy, R., and M. J. Abrahamson. "Medicare coverage for patients with diabetes." Journal of general internal medicine 21, no. 4 (2006): 386-392.

2. Brown RS, Peikes D, Peterson G, Schore J, Razafindrakoto CM. Six features of medicare coordinated care demonstration programs that cut hospital admissions of high-risk patients. Health Affairs. 2012;31:1156-11662.

3. Levine S, Steinman BA, Attaway K, Jung T, Enguidanos S. Home care program for patients at high risk of hospitalization. American Journal of Managed Care. 2012;18:e269-276

4. Zhang NJ, Wan TT, Rossiter LF, Murawski MM, Patel UB. Evaluation of chronic disease management on outcomes and cost of care for edicaid beneficiaries. Health policy (Amsterdam, Netherlands). 2008;86:345-354

5. Sommers LS, Marton KI, Barbaccia JC, Randolph J. Physician, nurse, and social worker collaboration in primary care for chronically ill seniors. Archives of internal medicine. 2000;160:1825-1833

6. Dorr DA, Wilcox AB, Brunker CP, Burdon RE, Donnelly SM. The effect of technology-supported, multidisease care management on the mortality and hospitalization of seniors. Journal of the American Geriatrics Society. 2008;56:2195-2202

7. Chan CL, You HJ, Huang HT, Ting HW. Using an integrated coc index and multilevel measurements to verify the care outcome of patients with multiple chronic conditions. BMC health services research. 2012;12:405

8. Littleford A, Kralik D. Making a difference through integrated community care for older people. Journal of Nursing and Healthcare of Chronic Illness. 2010;2:178-186

9. Centers for Medicare & Medicaid Services (CMS). Medicare health support. 2012

10. RTI International, Telligen. Accountable care organization 2013 program analysis: Quality performance standards narrative measure specifications. 2012

11. Sadur CN, Moline N, Costa M, Michalik D, Mendlowitz D, Roller S, Watson R, Swain BE, Selby JV, Javorski WC. Diabetes management in a health maintenance organization. Efficacy of care management using cluster visits. Diabetes care. 1999;22:2011-2017

12. McCarthy D, Cohen A, Johnson MB. Gaining ground: Care management programs to reduce hospital admissions and readmissions among chronically ill and vulnerable patients. The Commonwealth Fund, New York. 2013

13. Edwards, Samuel T., et al. "Home-Based Primary Care and the Risk of Ambulatory Care–Sensitive Condition Hospitalization Among Older Veterans With Diabetes Mellitus." JAMA internal medicine 174.11 (2014): 1796-1803.

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.)*
Not applicable.

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):*
Endocrine : Diabetes

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Care Coordination : Readmissions, Health and Functional Status, Health and Functional Status : Development/Wellness, Health and Functional Status : Functional Status, Overuse, Prevention, Safety, Safety : Complications

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)
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:** Diabetes_ACO_Admission_Measure_NQF_Data_Dictionary_01-29-16_v1.0-635896799914719697.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Not applicable

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 measured for each patient is the number of acute, unplanned admissions per 100 person-years at risk for admission. Persons are considered at risk for admission if they are alive, enrolled in FFS Medicare, and not currently admitted. (See S.6, Numerator Details, for more information.)

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.)*
The measure requires three years of data.

Outcome time window: We observe for the outcome of admission for one full calendar year.

Time period for cohort identification: The cohort is identified using two years of claims data prior to the measurement year.

Risk-adjustment look-back period: Risk-adjustment variables are identified using one year of data prior to the measurement year.

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.
Note: The numerator of the measure score is the predicted number of admissions given the Accountable Care Organization's (ACO's) case mix, sample size, and actual admission rate. We use this field to define the outcome.

Outcome Definition:
The outcome for this measure is the number of acute, unplanned admissions per 100 person-years at risk for admission. The outcome includes inpatient admissions to an acute care hospital for any cause during the measurement year, unless an admission is identified as "planned."

Identification of Planned Admissions:
The measure outcome includes only unplanned admissions. Although clinical experts agree that proper care in the ambulatory setting should reduce hospital admissions, variation in planned admissions (such as for elective surgery) does not typically reflect quality differences. We based the planned admission algorithm on the Centers for Medicare & Medicaid Services (CMS) Planned Readmission Algorithm Version 3.0, which CMS originally created to identify planned readmissions for the hospital-wide readmission measure. In brief, the algorithm identifies a short list of always planned admissions (i.e., those where the principal discharge diagnosis is major organ transplant, obstetrical delivery, or maintenance chemotherapy) as well as those admissions with a potentially planned procedure (e.g., total hip replacement or cholecystectomy) AND a non-acute principal discharge diagnosis code. To adapt the algorithm for this measure, we removed cardiac catheterization and amputation from the potentially planned procedure list. The need for these procedures might reflect progression of clinical conditions that potentially could have been managed in the ambulatory setting to avoid admissions for these procedures. For full details on the planned admission algorithm as adapted for this measure, please see Appendix A of the attached technical report.
Appendix A of the attached technical report contains the detailed algorithm used to identify planned admissions. Among 2,940,537 admissions in the 2012 Medicare Full Sample, 353,191 (12.0%) were planned admissions. For ACO patients, there were 148,708 admissions; of these, 20,000 (13.5%) were planned admissions. For non-ACO patients, there were 2,791,829 admissions; of these, 333,192 (12.0%) were planned admissions.
Please see Data Dictionary, sheet "S.6 ICD9-ICD10 Planned Algorithm," for the ICD-9 to ICD-10 crosswalk for the planned admission algorithm.

Outcome Attribution:
The outcome is attributed to the ACO to which the patient is assigned. Patients are assigned to ACOs according to the specific ACO program assignment algorithm. For example, for the Medicare Shared Savings Program, patient assignment is done retrospectively based on the plurality of care received at that ACO during the measurement year. Information on ACO patient assignment can be found here: Information on ACO patient assignment can be found here: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Downloads/Shared-Savings-Losses-Assignment-Spec-v2.pdf..

Citations:
Brown RS, Peikes D, Peterson G, Schore J, Razafindrakoto CM. Six features of Medicare coordinated care demonstration programs that cut hospital admissions of high-risk patients. Health Affairs. 2012 Jun 2012;31(6):1156-1166.

Center for Medicare and Medicaid Services. Medicare Shared Savings Program Shared Savings and Losses and Assignment Methodology Specifications. 2013; https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Downloads/Shared-Savings-Losses-Assignment-Spec-v2.pdf. Accessed July 30, 2014.

Centers for Medicare & Medicaid Services (CMS). Medicare Health Support. 2012; https://www.cms.gov/Medicare/Medicare-General-Information/CCIP/. Accessed March 27, 2014.

Horwitz LI, Grady JN, Cohen DB, Lin Z, Volpe M, Ngo CK, Masica AL, Long T, Wang J, Keenan M, Montague J, Suter LG, Ross JS, Drye EE, Krumholz HM, Bernheim SM. Development and validation of an algorithm to identify planned readmissions from claims data. J

Hosp Med 2015 Oct; 10(10):670-7.

McCarthy D, Cohen A, Johnson MB. Gaining Ground: Care Management programs to reduce hospital admissions and readmissions among chronically ill and vulnerable patients. The Commonwealth Fund, New York. 2013.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The target population is ambulatory Medicare FFS patients aged 65 years and older with a diagnosis of diabetes.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

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)*
Note: The denominator of the measure score is the expected admission rate for the ACO; we use this box to describe the measure cohort.

The targeted patient population is Medicare FFS patients aged 65 years and older with a diagnosis of diabetes receiving ambulatory care during the measurement period. To be included in the cohort, patients must have one inpatient or two outpatient diabetes diagnosis codes in any position within one or two years prior to the measurement period. We allowed for up to two years of claims to define the cohort since there is no specified optimal frequency of follow-up visits among ambulatory, stable patients (i.e., patients without a change in their symptoms may never be hospitalized and may only be seen annually). To be included in the cohort, patients must be enrolled full-time in both Part A and B during the year prior to the measurement period.

Diabetes is defined using the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnosis codes identified in Medicare Part A and Part B inpatient and outpatient claims data. Patients excluded from the cohort are identified using ICD-9-CM procedure codes in Medicare Part A inpatient and outpatient claims and the Medicare Denominator File. The ICD-9-CM codes that define the cohort are listed in the attached Excel file, sheets "S.9 Denominator Details – Cohort."
An ICD-9-CM to ICD-10-CM code crosswalk is attached in data field S.2b. (Data Dictionary or Code Table).

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
The measure excludes:
1. Patients without continuous enrollment in Medicare Part A for the duration of the measurement period (or until death).
Rationale: We exclude these patients to ensure full data availability for outcome assessment (Part A during the measurement year).

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)*
1. Patients without continuous enrollment in Medicare Part A for the duration of the measurement period (or until death).
Rationale: We exclude these patients to ensure full data availability for outcome assessment (Part A during the measurement year).

Lack of continuous enrollment in Medicare Part A is determined by patient enrollment status in FFS Part A using the Medicare Denominator File. The enrollment indicators must be appropriately marked during the measurement period (Part A).

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)*
Not applicable. This measure is not stratified.

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)

We use a two-level hierarchical negative binomial model to estimate risk-standardized acute, unplanned admissions per person-year at risk for admission. This approach accounts for the clustering of patients within ACOs and variation in sample size.

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" [1, 2]. The risk-standardization model includes age and 22 clinical variables. We define clinical variables using condition categories (CCs), which are clinically meaningful groupings of more than 15,000 ICD-9 diagnosis codes [3]. A map showing the assignment of ICD-9 codes to CCs can be found in the attached Data Dictionary Excel file, sheet "S.14 CC to ICD-9." Data Dictionary, sheet "S.15 ICD10 Crosswalk-Risk model" contains the crosswalk of ICD-9 to ICD-10 codes for the diabetes severity index variable.

Model Variables

The risk-adjustment variables are:

1. Age
2. High Risk cardiovascular (CV) factors (CC 81, 82, 89, 104)
3. Low risk CV factors (CC 83, 84, 94, 105, 106)
4. Arrhythmia (CC 92, 93)
5. Advanced Cancer (CC 7, 8, 9, 11)
6. Dementia (CC 49, 50)
7. Heart failure (CC 80)
8. Dialysis (CC 130)
9. Disability/Frailty (CC 21, 67, 68, 100, 116, 148, 149, 157, 177, 178, 69)
10. Gastrointestinal and Genitourinary disorders (GI/GU) (CC 29, 30, 31, 33, 34, 133, 176)
11. Hematological disorders (CC 44, 46)
12. Infectious and immune disorders (CC 1, 3, 4, 5, 45, 85)
13. Kidney disease (CC 128, 131, 132)
14. Liver disease (CC 25, 26, 27, 28)
15. Neurological disorders (CC 48, 61, 65, 70, 72, 73, 74,75, 95, 96, 97, 98, 99, 101, 102, 103, 155)
16. Psychiatric Illness/Substance abuse (CC 51, 52, 53, 54, 55, 56, 57, 58, 59, 60)
17. Pulmonary disease (CC 107, 108, 109, 110, 114, 115)
18. Other advanced organ failure (CC 77, 79)
19. Diabetes severity index (number of complications associated with diabetes based on ICD-9 codes; see Testing form 2b.4.3 for details and Excel file, sheet "S.15 Diabetes Severity Index" for the list of ICD-9 codes.)
20. Iron deficiency anemia (CC 47)
21. Major organ transplant (CC 174)
22. Other organ transplant (CC 175)
23. Hip fracture/Major fracture (CC 158, 159)

Citations:

1. 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.

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

3. Pope, G.C., Kautter, J., Ellis, R.P., et al.: Risk Adjustment for Medicare Capitation Payments Using the CMS-HCC Model. Health Care Financing Review. 2004; 25(4):119-141.

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*)

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.)*
The risk-standardized acute admission rate (RSAAR) for each ACO is calculated as the number of "predicted" to the number of "expected" admissions per person-year, multiplied by the national rate of admissions among all Medicare FFS patients with diabetes – i.e., all eligible Medicare FFS patients with diabetes are used in the measure score calculation, and a score is generated for each ACO. For a full description of the modeling, please see the attached technical report (Section 3.5.5 and Appendix B of attached technical report).

In brief, the measure uses a hierarchical (two-level) statistical model that accounts for the clustering of patients within ACOs and accommodates the widely varying sizes of different ACOs. The measure uses a negative binomial model since our outcome is a count of the number of admissions. The first level of the model adjusts for patient factors. The relationship between patient risk factors and the outcome of admission is determined based on a national sample of patients with diabetes. Stated another way, since the effects that risk factors exert on the number of admissions are estimated based on data from all ACO and non-ACO patients in the nation, the 'expected' number of admissions for each ACO is based on the performance of a national group of providers.
The second level of the model estimates a random-intercept term that reflects the ACO's contribution to admission risk, based on its actual admission rate, the performance of other providers with similar case mix, and its sample size. The ACO-specific random intercept is used in the numerator calculation to derive ACO specific number of "predicted" admissions per person-year.

The measure score is the ratio of predicted admissions over the expected admissions multiplied by the crude national rate. The predicted to expected ratio of admissions is analogous to an observed/expected ratio, but the numerator accounts for clustering and sample-size variation.

The expected number of admissions is calculated based on the ACO's case mix and national average intercept.

The predicted number of admissions is calculated based on the ACO's case mix and the estimated ACO-specific intercept term.

We multiply the ratio for each ACO by a constant, the crude national rate of acute, unplanned admissions per person-years at risk for hospitalization, for ease of interpretation.

To place ACOs in performance categories, for each ACO RSAAR, one can calculate a 95% interval estimate (IE), which is similar to a confidence interval, using standard bootstrapping methods (further described in the Testing Form, Section 2b5.1). Using the 95% IEs, one can assign ACOs to one of three performance categories: 'better than the national rate,' 'no different than the national rate,' and 'worse than the national rate.' The ACO is 'better than the national rate' if the 95% IE is completely below the United States (US) national rate among Medicare FFS patients with diabetes; 'no different than the national rate' if the 95% IE is included in the US national rate among Medicare FFS patients with diabetes; and 'worse than the national rate' if the 95% IE is above the US national rate among Medicare FFS patients with diabetes.

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.
This 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.
This 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.
Not applicable

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

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.
Medicare administrative claims and enrollment data

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)*
Integrated Delivery System

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic, Other
If other: ACO

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.)*
Not applicable.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
Diabetes_ACO_Admission_Measure_NQF_Testing_Form_01-29-16_1.0.docx

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

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Risk-Standardized Acute Admission Rates for Older Patients with Diabetes
Date of Submission: 1/29/2016
Type of Measure:

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

To develop and to test the patient-level model, we used several 2010-2012 Medicare claims datasets as outlined below:

1. Medicare dataset used to identify the diabetes cohort and patient risk factors for admission:

We used the 2010-2011 Chronic Conditions Data Warehouse (CCW) 100% dataset which includes patients with at least one of the 27 CCW chronic conditions. We used the CCW 2010-2011 Medicare Part A and Part B files to define the cohort and CCW 2011 Medicare Part A and Part B files to identify each patient's risk factors for the outcome of acute, unplanned admissions per person-year at risk for hospitalization. Our diabetes cohort is fully encompassed within this dataset of patients with at least one CCW chronic condition.

We used the 2011-2012 Denominator File to determine Medicare Fee-for-Service (FFS) enrollment, demographic, and death information for beneficiaries in our cohort in order to determine inclusion/exclusion criteria for the cohort.

2. Medicare dataset to capture the outcome (acute, unplanned admissions per person-years at risk for hospitalization):

We used the 2012 Medicare Provider Analysis and Review (MedPAR) 100% FFS dataset, containing Medicare Part A claims, to identify the outcome of admissions.

We used the 2012 Denominator File to determine Medicare FFS enrollment, demographic (including 5-digit zip code), and death information for beneficiaries in the diabetes cohort to determine person-years at risk for hospitalization.

3. Dataset to identify assignment of patients to Accountable Care Organizations (ACOs):

We used a file provided by a CMS contractor to identify which Medicare FFS beneficiaries were assigned to each of 114 Medicare Shared Savings Program ACOs in the year 2012.

4. Dataset to determine socioeconomic status:

We used the 2008-2012 American Community Survey data from the United States (US) Census Bureau to derive the Agency for Healthcare Research and Quality (AHRQ) socioeconomic status (SES) index for each zip code in the US.

5. Dataset to identify dual-eligibility status:

We used the 2012 Denominator File to identify dual-eligible Medicare FFS beneficiaries.

The datasets used for testing vary 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 2010-2012. The dates of the data listed above are as follows:

1. CCW 100% Medicare Parts A and B dataset: 2010-2011

2. MedPAR dataset: 2012

3. ACO assignment data: 2012

4. US Census Bureau, American Community Survey dataset: 2008-2012

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

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

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

The number of measured entities (i.e., ACOs) varies by testing type; see Section 1.7 for details.

1.6. How many and which patients 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 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.

As set forth in Section 1.2 above, we use Medicare claims and enrollment data to identify the cohort, to define the outcome, and to accumulate risk-adjustment variables. For measure development and testing, we created datasets using 2010-2012 Medicare data, using 2012 as the measurement year. The datasets, dates, number of measured entities and number of patients used in each type of testing are as follows:

1) 2012 Medicare Full Sample
This sample includes the cohort of all Medicare FFS beneficiaries meeting our diabetes definition for the 2012 measurement year. The 2012 Medicare Full Sample includes 6,521,462 diabetes beneficiaries. Patients were mostly female (54.7%) with an average age of 76.4 years. There were 114 ACOs in the 2012 Medicare Full Sample. Among the 6,521,462 diabetes beneficiaries, 341,193 (5.2%) were assigned to one of 114 ACOs.
-Dataset used for: testing measure exclusions (see Section 2b3), meaningful differences in performance (see Section 2b5), risk-adjustment model (Section 2b4.4b), and all ACO measure score calculations

For model development and testing, we randomly split the 2012 Medicare Full Sample into two equal subsets of patients: the 2012 Development Sample and 2012 Validation Sample (described below).

a) 2012 Development Sample
-This sample includes 3,260,731 patients with diabetes. Patients were mostly female (54.7%), with an average age of 76.4 years. There were 114 ACOs; 170,390 (5.2%) of patients in the 2012 Development Sample were assigned to ACOs.
-Dataset used for: data element reliability (see Section 2a2.3), testing risk-adjustment model (see Section 2b4)

b) 2012 Validation Sample
-This sample includes 3,260,731 patients with diabetes. Patients were mostly female (54.7%), with an average age of 76.4 years. There were 114 ACOs; 170,803 (5.2%) of patients in the 2012 Validation Sample were assigned to ACOs.
-Dataset used for: data element reliability (see Section 2a2.3), testing risk-adjustment model (see Section 2b4)

We also split the 2012 Medicare Full Sample into subsets of patients by randomly splitting each ACO's patients in half and then randomly splitting all non-ACO patients in half.

c) 2012 Reliability Sample 1
-2012 Reliability Sample 1 includes 3,260,759 patients with diabetes. Patients were mostly female (54.7%), with an average age of 76.4 years. 5.2% of patients were assigned to ACOs.
-Dataset used for: measure score reliability (see Sections 2a2 and 2b2)

d) 2012 Reliability Sample 2
-2012 Reliability Sample 2 includes 3,260,703 patients with diabetes. Patients were mostly female (54.7%), with an average of 76.4 years; 5.2% of patients were assigned to ACOs.
-Dataset used for: measure score reliability (see Sections 2a2 and 2b2)

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 used two different indicators of Medicare beneficiaries' socioeconomic status (SES): (1) the SES score of the patient's five-digit zip code, adapted from the AHRQ SES Index, which was created for the purpose of characterizing the SES of Medicare beneficiaries and (2) the Medicaid dual-eligibility status of beneficiaries [1]. Although race was available (as black or other) in the Medicare data, we chose not to further evaluate it based on our conceptual model and input from our Technical Expert Panel (TEP) and public comment.

The AHRQ SES Index is based on seven neighborhood variables previously shown to contribute to SES and to be associated with outcomes. They are: (1) median household income, (2) percentage of persons living below the federal poverty level, (3) percentage of persons who are aged >16 years and in the labor force but not employed, (4) median value of owner-occupied homes, (5) percentage of persons aged >25 years who completed at least a 12^{th}-grade education, (6) percentage of persons aged >25 years who completed at least four years of college, and (7) percentage of households that average one or more persons per room. The original AHRQ SES Index was derived using data from the 2000 US Census Bureau and was calculated using US Census Block data, which corresponded to Medicare beneficiaries' nine-digit zip code. For this measure, we used data from the US Census Bureau, American Community Survey (2008-2012) and performed a principal component analysis to derive a composite SES index score for each five-digit zip code, which we then assigned to the patient based on their zip code of residence (i.e., the smallest unit by which we could identify Medicare beneficiaries' home address). The AHRQ SES Index is a continuous variable whereby lower scores indicate lower SES zip codes and higher scores indicate higher SES zip codes.

We created a dichotomous variable from the AHRQ SES index, stratifying zip code scores into 'low SES' and 'non-low SES.' Based on the distribution of the AHRQ SES index among the entire FFS Medicare population in the 5% Medicare FFS sample, we selected the lowest quintile to represent low SES. In this lowest quintile, 21.9% of beneficiaries were Medicaid dual-eligible, as compared with 13.7% in the second lowest quintile. We then categorized each patient as low or non-low SES based on the AHRQ score derived from their zip code of residence.

Additionally, we categorized ACOs based on the proportion of low SES patients in their cohort into quartiles (first quartile [Q1] indicating few low SES patients, fourth quartile [Q4] indicating many low SES patients). Similarly, we categorized ACOs by the proportion of Medicaid dual-eligible patients in their cohort into ACOs caring for 'few' (Q1) and 'many' (Q4) Medicaid dual-eligible patients. For more information on the derivation of the AHRQ SES index and the selection of a low SES thresholds for patients and ACOs, see the Appendix E of the attached Appendix.

We did not use race in our analyses since differences in risk of admission among groups of different race should be captured in our risk-adjustment model (which includes age and comorbidities). Any remaining differences in the risk for hospitalization among patients of different race may represent disparities in care delivery and quality of care.

Citations

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

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 in Medicare FFS patients, we aimed to utilize only those data elements from claims data that have both face validity and reliability. We avoided the use of fields that are thought to be coded inconsistently across facilities. Specifically, we used fields that are consequential for payment and which are audited. We identified such variables through empiric analyses and our understanding of the Centers for Medicare & Medicaid Services (CMS) auditing and billing policies. We sought to avoid variables which do not meet these standards.

In addition, CMS has in place several hospital auditing programs used to assess overall accuracy of claims-based coding, to ensure appropriate billing, and for overpayment recoupment. CMS routinely conducts data analysis to identify potential problem areas and to 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 assessed the reliability of the data elements by comparing model variable frequencies in our 2012 Development Sample and 2012 Validation Sample.

Measure Score Reliability
The reliability of a measurement can be defined as the degree to which repeated measurements of the same entity agree with one another. For our measures of facility performance, the measured entity is the ACO, and reliability is the extent to which repeated measurements of the same ACO give similar results [1].

To calculate measure score reliability, we randomly sampled half of the patients from each ACO and half of the patients who were not in ACOs from the 2012 Medicare Full Sample (2012 Reliability Sample 1 and 2012 Reliability Sample 2). We calculated the measure score for all the ACOs using data from ACO and non-ACO patients, and repeated the calculation using the second half of patients. Thus, each ACO was measured twice, but each measurement was 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 ACO, not of the patients. As a metric of agreement, we calculated the intraclass correlation coefficient (ICC) [2], and assessed the values according to conventional standards [3]. The agreement of the two risk-standardized acute admission rates was quantified for ACOs in each sample using the ICC (2,1) by Shrout and Fleiss [2].

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

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

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

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:

Table 1. Risk variable frequencies for 2012 Development Sample and 2012 Validation Sample

	Prevalence of risk factors (%)	
	2012 Development Sample (N = 3,260,731)	**2012 Validation Sample (N = 3,260,731)**
Age mean (std)	76.4 (7.2)	76.4 (7.2)
High risk cardiovascular (CV) factors	549,201 (16.8%)	549,985 (16.9%)
Low risk CV factors	1,858,261 (57.0%)	1,859,678 (57.0%)
Arrhythmia	948,179 (29.1%)	948,271 (29.1%)
Advanced cancer	181,130 (5.6%)	181,857 (5.6%)
Dementia	464,377 (14.2%)	464,201 (14.2%)
Heart failure	751,268 (23.0%)	751,873 (23.1%)
Dialysis	49,847 (1.5%)	50,116 (1.5%)
Disability/Frailty	440,293 (13.5%)	439,847 (13.5%)
Gastrointestinal and Genitourinary disorders (GI/GU)	761,670 (23.4%)	759,637 (23.3%)
Hematological disorders	275,930 (8.5%)	277,354 (8.5%)
Infectious and immune disorders	109,381 (3.4%)	109,963 (3.4%)
Kidney disease	737,899 (22.6%)	738,477 (22.6%)
Liver disease	55,928 (1.7%)	5,5667 (1.7%)
Neurological disorders	850,552 (26.1%)	849,391 (26.0%)
Psychiatric illness/Substance abuse	871,240 (26.7%)	870,820 (26.7%)
Pulmonary disease	1,209,109 (37.1%)	1,211,483 (37.2%)
Other advanced organ failure	248,333 (7.6%)	248,488 (7.6%)
Diabetes severity index mean(std)	1.7 (1.4)	1.7 (1.4)
Iron deficiency anemia	1,199,402 (36.8%)	1,198,405 (36.8%)
Major organ transplant	7,451 (0.2%)	7,646 (0.2%)
Other organ transplant	19,942 (0.6%)	19,989 (0.6%)
Hip fracture/Major fracture	111,737 (3.4%)	111,706 (3.4%)

Measure Score Reliability:

The ICC between the two risk-standardized acute admission rates (RSAARs) was 0.889, which according to the conventional interpretation is "excellent" [1].

Citations

1. 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?*)

Data Element Reliability Results
Compared with the 2012 Development Sample, the mean age of patients and the frequency of risk-adjustment variables were similar in the 2012 Validation Sample. This suggests that the data elements are reliable across these samples.

Measure Score Reliability Results
The ICC demonstrates excellent agreement across samples, indicating 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)*

We demonstrated measure validity through: (1) reliance on relevant prior validity testing conducted for other claims-based measures; (2) use of established measure development guidelines; and (3) assessment by external groups and a technical expert panel (TEP).

Validity of Claims-Based Measures
Our team has demonstrated the validity of using claims data for risk adjustment in lieu of medical record data in estimating facility-level measure scores for a number of hospital outcome measures endorsed by the National Quality Forum (NQF). CMS has validated six NQF-endorsed measures currently in public reporting (acute myocardial infarction [AMI], heart failure, and pneumonia mortality and readmission) with models that used medical record-abstracted data for risk adjustment. Specifically, we conducted claims model validation by building comparable models using abstracted medical record 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]. These findings support this measure's validity; however, we acknowledge that the use of claims data for risk adjustment has been validated for hospital outcomes measure and not for outcome measures among ambulatory patients.

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], 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 through: holding regular discussions with the external experts in our working group; consulting our national Technical Expert Panel (TEP); and holding a 30-day public comment period. We obtained expert and stakeholder feedback for development of two related measures for patients with heart failure or multiple chronic conditions.

Yale New Haven Health Services Corporation—Center for Outcomes Research and Evaluation (CORE) clinicians and statistical experts comprised the working group. The working group members have expertise in quality measurement, clinical management of patients with diabetes, statistical modeling, healthcare disparities, and healthcare policy. Through regular in-person meetings and teleconferences, the working group discussed all aspects of measure development, including the cohort and outcome definitions and risk adjustment.

In addition to the working group and in alignment with the CMS Measures 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, patients, and individuals with experience in quality improvement, performance measurement, and healthcare disparities. We held four structured TEP conference calls consisting of presentation of key issues, our proposed approach, and relevant data, followed by open discussion among TEP members.

List of TEP Members

1. Lawrence M. Becker, BS, Xerox Corporation (Director, Strategic Partnerships, Alliances and Analytics); Rochester, NY
2. Alex Blum, MD, MPH, Evergreen Health Cooperative (Chief Medical Officer); Baltimore, MD
3. Sanjay Doddamani, MD, Geisinger Health System (System-wide Chief of Advanced Cardiac Disease – Heart Failure); Danville, PA
4. Kevin Fiscella, MD, MPH, University of Rochester Medical Center (Professor of Family Medicine); Rochester, NY
5. Elbert Huang, MD, MPH, University of Chicago (Associate Professor of Medicine, Director of the Center for Translational and Policy Research of Chronic Diseases, and Associate Director of the Chicago Center for Diabetes Translation Research); Chicago, IL
6. Bruce Leff, MD, Johns Hopkins University School of Medicine (Professor of Medicine, Division of Geriatric Medicine); The Johns Hopkins University Bloomberg School of Public Health (Faculty, Health Services Research Development Center and Lipitz Center for Integrated Health Care); Baltimore, MD
7. Andy Miller, MD, MPH, Healthcare Quality Strategies, Inc. (Medical Director); East Brunswick, NJ; Colorado Foundation for Medical Care (CMO, Integrating Care for Populations & Communities National Coordinating Center); Englewood, CO
8. Ami Parekh, MD, JD, University of California, San Francisco (Medical Director for Health System Innovation); San Francisco, CA
9. Christine Ritchie, MD, University of California, San Francisco (Professor of Medicine, Division of Geriatrics); San Francisco, CA
10. Two patient representatives.

We systematically assessed the face validity of the measure score as an indicator of quality by soliciting the TEP members' agreement with the following statement: "The RSAARs obtained from the diabetes measure as specified can be used to distinguish between better and worse quality ACOs."

TEP members indicated their agreement with the face validity of the measure on a six-point scale:

1=Strongly disagree
2=Moderately disagree
3=Somewhat disagree
4=Somewhat agree
5=Moderately agree
6=Strongly agree

Process Used to Identify International Classification of Diseases, Tenth Revision (ICD-10) Codes
This application includes ICD-10 codes that correspond to all International Classification of Diseases, Ninth Revision (ICD-9) codes included in the specifications. The goal was to convert this measure into a new code set, fully consistent with the intent of the original measure.

• ICD-10 diagnosis codes used to the cohort were identified using the 2013 ICD-9-CM to ICD-10-CM General Equivalence Mapping (GEM) files made available by CMS. We then internally performed clinician review of this crosswalk.

• ICD-10 diagnosis codes used to define diabetes severity index variable (the only ICD-9-CM risk-adjustment variable) were identified using the 2013 ICD-9-CM to ICD-10-CM GEM files made available by CMS. We then internally performed clinician review of this crosswalk.

• ICD-10 diagnosis and procedure codes used to define the Planned Admission Algorithm were identified from the 2014 version of the AHRQ Clinical Classification Software (CCS) categories specified for ICD-10, followed by clinician review. The algorithm also includes some individual ICD-9-CM codes. To create the crosswalk for the ICD-9-level codes, we used the 2013 ICD-9-CM to ICD-10-CM GEM files made available by CMS, followed by clinician review.

Citations
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
The results of the TEP rating of agreement with the validity statement were as follows:
N=8
Mean rating=4.9

All TEP members who responded to the survey indicated they agreed with the statement; 6 of the 8 indicated that they moderately or strongly agreed. Five TEP members did not respond to the TEP survey.

Frequency of Ratings of Agreement:

Rating	# (%) of Responses
1 (Strongly disagree)	0 (0.0)
2 (Moderately disagree)	0 (0.0)
3 (Somewhat disagree)	0 (0.0)
4 (Somewhat agree)	2 (25.0)
5 (Moderately agree)	5 (62.5)
6 (Strongly agree)	1 (12.5)

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?*)

There was strong support expressed by the members of the TEP and in public comment for the validity of the measure. There were no strong concerns about the measure. One of 13 commenters felt the outcome was not an indicator of quality. See public comment document for further details: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/CallforPublicComment.html.

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*)

We determined the exclusions to be appropriate based on clinical and methodological considerations, such as whether we had sufficient data for patient subsets or could adequately adjust for the risk of admission in certain patient subpopulations. We examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion.

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*)

Table 2 provides the number of patients excluded from the diabetes cohort. Out of the total number of Medicare FFS patients with diabetes (N = 6,746,776), we excluded 225,314 (3.3%) due to non-continuous enrollment in Medicare Part A in 2012 because we were not able to adequately capture the outcome for these patients. Among these excluded patients, 225,310 (99.99%) were non-ACO patients and 4 were ACO patients.

Since the number of excluded patients assigned to ACOs was very low, we did not perform a frequency distribution analysis across ACOs.

The final cohort included 6,521,462 patients.

Table 2. Patients excluded from sample for each exclusion criterion

Exclusion	Number excluded from Medicare FFS diabetes cohort	Number of patients excluded from ACOs
Non-continuous enrollment in Part A in 2012	225,314	4

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*)

We needed to exclude patients without continuous enrollment because we could not capture the outcome for these patients. We excluded very few patients based on this criterion. As a result, the measure captures the majority of Medicare FFS patients 65 years and older who meet its criteria for a diagnosis of diabetes (96.7%).

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 23 risk factors**

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

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

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

Not applicable. This measure is risk-adjusted.

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)

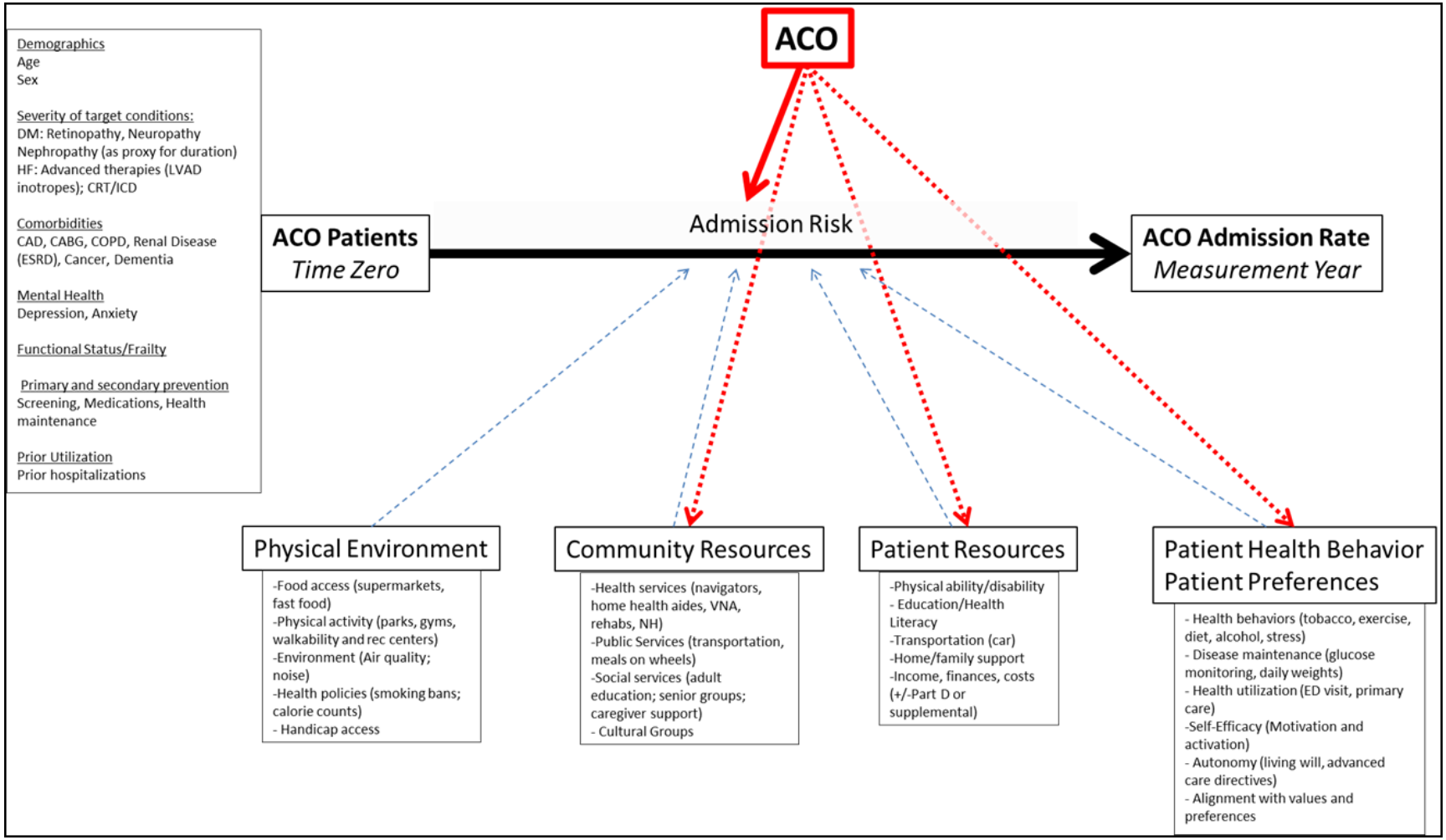
We selected the risk-adjustment model variables based on the existing literature, clinical judgment, empirical analyses, and input from our TEP and other experts. We considered factors that may impact the rate of admission, including patient-level factors (e.g., demographics, SES, clinical risk factors on admission); we also considered the impact of other non-clinical factors such as health behaviors and community resources.

In this work, we were guided by a conceptual framework that was informed by a literature review and environmental scan, outlining the relationships between potential clinical and contextual factors and rates of admissions among chronic disease populations cared for by ACOs. Importantly, many factors other than traditional medical care delivered in the office or hospital settings will impact health outcomes for patients with chronic disease. For example, ACOs practicing in communities where patients have limited access to transportation, healthy foods and recreational facilities, may have less success in promoting healthy behaviors among patients with diabetes; this may in-turn impact quality outcomes. Recognition of and attention to the health environment may be important for achieving the goals of better care, better health and lower costs and thus, shared savings.

The conceptual model (Figure 1) was presented and endorsed by the TEP engaged during the development of this measure. The model recognizes patient-level demographic and clinical factors, along with 4 contextual domains that may influence ACO performance: (1) Physical environment (e.g., green spaces; safe streets); (2) Community resources (e.g., home health; senior services); (3) Patient resources (e.g., social support; transportation; income); and (4) Patient behavior/personal preferences (e.g., exercise; diet; advanced care directives; preference for intervention).

The model also recognizes the capacity of ACOs to mitigate the effects of many contextual factors on rates of admissions, encompassing both SES and non-SES variables, and supporting our decision not to adjust for contextual factors. Adjusting for contextual factors would obscure important differences in ACO quality and could serve as a disincentive for ACOs to engage with such factors. We did, however, conduct analyses of SES factors to further inform the committee's deliberation.

Figure 1. Conceptual model of factors affecting risk of hospital admission



We describe our approach to risk-adjustment for the demographic factors, clinical risk factors, and contextual domains, in turn, below:

1. Demographic factors
We used clinical and conceptual criteria to adjust this measure for age but not sex or race. Age is a clinically recognized risk factor for acute admissions. In contrast, sex or race differences in risk of admission should be captured in our risk-adjustment model (which includes age and comorbidities). Any remaining differences in the risk for hospitalization among patients of different sex or race may represent disparities in care delivery and quality of care. [1,2] We did examine the effects of including sex in the models, since the relationship between sex and acute, unplanned admissions has not been tested in this setting, finding that sex was not significant after adjusting for age and clinical comorbidities.

2. Clinical risk factors
We used clinical, conceptual, and statistical criteria to select clinical risk factors for adjustment. This measure adjusts for clinical risk factors that are present at the start of the measurement period, but not for conditions that arise during the measurement period.

Development of Candidate Clinical Variables
To select candidate variables for risk adjustment, we used Part A and Part B data from one year prior to the measurement year for 100% of the Medicare FFS patients included in the cohort (2012 Medicare Full Sample). We reviewed 189 diagnostic groups included in the Hierarchical Condition Category (HCC) clinical classification system. We defined comorbidities using 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, sheet "S.14 CC-ICD-9 Map." To select candidate variables, two clinicians reviewed all 189 CCs and excluded those that were not relevant to the Medicare population or that were not clinically relevant to the all-cause acute admission outcome (e.g., attention deficit disorder, female infertility). The remaining 181 clinically relevant CCs were selected as candidate variables.

Among the 181 clinically relevant CCs, we calculated the prevalence of the CC in the year preceding the measurement period (i.e., 2011), the number of hospital admissions per patient-year during the measurement period (i.e., 2012) among patients with and without the CC, and the rate ratio for the number of hospital admissions associated with each CC. Based on these statistical findings, we reduced the list of CCs to 92 from the initial list of 181 clinically relevant CCs. We reviewed the results of the bivariate analyses of the 92 CCs and collapsed the 92 CCs into 22 candidate variables, plus age. Additionally, we included the diabetes complications severity index in the model. This variable captures the number of complications associated with diabetes that each patient has: retinopathy, nephropathy, neuropathy, cerebrovascular, cardiovascular, peripheral vascular disease, and metabolic complications. The index takes on values from zero to seven, according to the number of complications present, and has been validated in claims data [4-5].

Candidate Clinical Variables
The selected candidate variables were:
1. Age
2. High risk cardiovascular factors (CC 81-82, 89, 104)

3. Low risk cardiovascular factors (CC 83-84, 94, 105-106)
4. Arrhythmia (CC 92-93)
5. Advanced cancer (CC 7-9, 11)
6. Dementia (CC 49-50)
7. Heart failure (CC 80)
8. Dialysis (CC 130)
9. Disability/Frailty (CC 21, 67-68, 100, 116, 148-149, 157, 177-178, 69)
10. Gastrointestinal and genitourinary disorders (GI/GU) (CC: 29-30, 31, 33-34, 133, 176)
11. Hematological disorders (CC 44, 46)
13. Infectious and immune disorders (CC: 1, 3-5, 45, 85)
13. Kidney disease (CC 128, 131-132)
14. Liver disease (CC 25-28)
15. Neurological disorders (CC 48, 61, 65, 70, 72-75, 95-99, 101-103, 155)
16. Psychiatric illness/Substance abuse (CC 51-60)
17. Pulmonary disease (CC 107-110, 114-115)
18. Other advanced organ failure (CC 77, 79)
19. Diabetes severity index (number of complications associated with diabetes based on ICD-9 codes; see Excel attachment, sheet "S.15 Diabetes Severity Index")
20. Iron deficiency anemia (CC 47)
21. Major organ transplant (CC 174)
22. Other organ transplant (CC 175)
23. Hip fracture/Major fracture (CC 158-159)
24. Structural heart disease (CC 86-88)

Final variable selection:
In order to select the final set of variables, we ranked the variables in terms of their importance for the model by comparing the Akaike Information Criterion (AIC) values using the 2012 Development Sample. We selected variables starting with the 24 candidate variables. We removed one variable and determined the best combination of 23 variables that resulted in the smallest AIC compared with other combinations of 23 variables. Based on the best 23 variables, we removed one more variable and determined the best 22 variables. We repeated these steps until we reached one variable. Each of the final 23 variables represents the best model (combination of variables) given different numbers of variables.

The attached Data Dictionary, sheet "S.15 Risk Model Specifications" indicates the final risk variables selected, the codes used to define the risk variables for our statistical model, and their frequencies in the 2012 Development Sample and 2012 Validation Sample.

3. Socioeconomic status
Based on a conceptual model that was informed by a literature review and environmental scan, we did not adjust for contextual factors which may impact acute admissions, including variables related to SES. ACOs should and do influence a broad range of patient-level and community-level factors that can mitigate the risk of admission associated with the contextual environment.

However, to inform the committee's consideration of the decision not to adjust for SES, we performed focused analyses using SES variables. These analyses are informative for future

measure use, but the decision not to adjust for SES in this measure was not based on the results of these statistical analyses.

To assess the potential effect of SES on ACO performance, we first included SES as a patient-level covariate in the models. As there are no standardized methods for assessing a Medicare beneficiary's SES, we used two different indicators of SES: (1) the SES score of the patient's 5-digit zip code, adapted from the AHRQ SES Index [3], which was developed for the purpose of characterizing the SES of Medicare beneficiaries and (2) the Medicaid dual-eligibility status of beneficiaries. We created a dichotomous variable from the AHRQ SES score, defining patients as low SES if they had an AHRQ Score of 0 to 45 and non-low SES if they had an AHRQ score of >45. This cut-point represented the lowest quintile of AHRQ SES scores among the 5% Medicare FFS Sample. In this lowest quintile, 21.9% of patients were Medicaid dual eligible. For further details on how we calculated the AHRQ SES score and developed a dichotomous variable we refer to the attached technical report, Appendix E. Additionally, we performed ACO-level analyses based on the proportion of low SES patients being cared for by an ACO. These methods and results are reported in the NQF Submission form.

4. Contextual Domains
The four contextual domains, which include SES factors, may influence the clinical health status of patients as well as the outcome of acute admissions, impacting ACOs' ability to prevent acute admissions. However, when evaluating provider quality, we do not want to adjust for them, since these affects may be mediated by ACOs, and the measure score should ideally reflect successful efforts to mitigate their impact on admission rates. This approach is consistent with the ACO program design – as part of their mission, ACOs are encouraged to develop strategic partnerships with community-based organizations and businesses in order to improve population health and reduce the risk of admission. It is also supported by growing evidence that integrated health systems can identify and mitigate the degree to which non-health factors impact health outcomes (e.g., by connecting patients with available health-related services) [4].

Citations

1. Rathore SS, Foody JM, Wang Y, et al. Race, quality of care, and outcomes of elderly patients hospitalized with heart failure. JAMA : the journal of the American Medical Association. May 21 2003;289(19):2517-2524.

2. Deswal A, Petersen NJ, Urbauer DL, Wright SM, Beyth R. Racial variations in quality of care and outcomes in an ambulatory heart failure cohort. American heart journal. Aug 2006;152(2):348-354.

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

4. Alley DE, Asomugha CN, Conway PH, Sanghavi DM. Addressing Social Needs through Medicare and Medicaid, N Engl J Med 2016; 374:8-11.

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

Based on the smallest AIC among the 24 combinations, we retained 23 variables in the final model. Of the 24 candidate variables, the only variable that was not included was structural heart disease, which was not statistically significant in the model.

The following variables were selected as the final risk-adjustment variables:
1. Age
2. High risk cardiovascular factors (CC 81, 82, 89, 104)
3. Low risk cardiovascular factors (CC 83, 84, 94, 105, 106)
4. Arrhythmia (CC 92, 93)
5. Advanced cancer (CC 7, 8, 9, 11)
6. Dementia (CC 49, 50)
7. Heart failure (CC 80)
8. Dialysis (CC 130)
9. Disability/Frailty (CC 21, 67, 68, 100, 116, 148, 149, 157, 177, 178, 69)
10. Gastrointestinal and genitourinary disorders (CC 29, 30, 31, 33, 34, 133, 176)
11. Hematological disorders (CC 44, 46)
12. Infectious and immune disorders (CC 1, 3, 4, 5, 45, 85)
13. Kidney disease (CC 128, 131, 132)
14. Liver disease (CC 25, 26, 27, 28)
15. Neurological disorders (CC 48, 61, 65, 70, 72, 73, 74,75, 95, 96, 97, 98, 99, 101, 102, 103, 155)
16. Psychiatric Illness/Substance abuse (CC 51, 52, 53, 54, 55, 56, 57, 58, 59, 60)
17. Pulmonary disease (CC 107, 108, 109, 110, 114, 115)
18. Other advanced organ failure (CC 77, 79)
19. Diabetes severity index (number of complications associated with diabetes based on ICD-9 codes; see Excel attachment, sheet "S.15 Diabetes Severity Index")
20. Iron deficiency anemia (CC 47)
21. Major organ transplant (CC 174)
22. Other organ transplant (CC 175)
23. Hip fracture/major fracture (CC 158, 159)

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)

We performed multiple analyses to assess the effect of sex and SES on model performance. These analyses are informative for future measure use, but the decision not to adjust for sex or SES in this measure was based on conceptual/clinical factors and not on the results of these statistical analyses (see 2b4.3.).

To assess the effect of sex and SES on model performance, we compared deviance R-squared values with and without the variables for sex and SES included as patient-level variables in the

model. We compared the correlation between measure scores with and without sex and SES included in the models, using the Spearman correlation.

For the SES analyses, we also assessed ACO performance among groups of ACOs caring for similar proportions of low SES patients. To do this, we categorized ACOs into quartiles (Q1 indicating ACOs with few low SES patients, Q4 indicating ACOs with many low SES patients). We used boxplots to compare the distribution of RSAARs across ACOs by low SES quartiles.

The SES analyses were performed using both the AHRQ SES index (i.e., low SES, binary variable described above) and Medicaid dual-eligibility status as a proxy for patients' SES status. Statistical analyses were performed using SAS version 9.3 (SAS Institute Inc., Cary, NC).

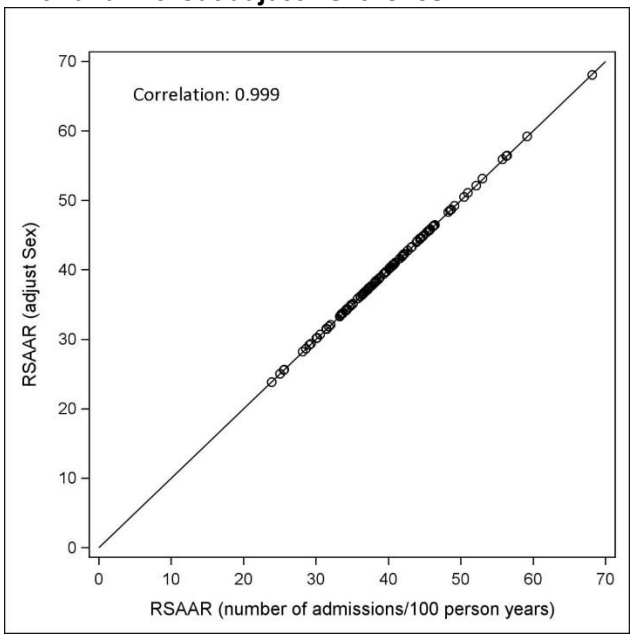
The results of the patient-level analyses indicate that adjustment for sex and for low-SES status as a patient variable in the models did not affect measure performance.

Specifically, related to SES, performance scores did not change appreciably after adjusting the models for patients' SES. As demonstrated in the Testing Form, Section 2b4.11, the Spearman correlation comparing the ACO measure scores estimated with and without risk adjustment for the AHRQ SES Index was 0.981. Similarly, the Spearman correlation for the scores estimated with and without patients' Medicaid dual eligibility was 0.976. These results demonstrate that adjusting for SES at the patient level has little effect on the measure score.

Sex

The deviance R-squared values for the two models, one adjusted for the 23 clinical variables *and* sex, and one adjusted for the 23 clinical variables *with*out sex, were 0.218 and 0.217, respectively, meaning adjustment for sex explained the same amount of variation and did not result in incremental benefit. Comparing the RSAAR with and without sex included in the model resulted in a high degree of correlation (Spearman correlation = 0.999), meaning ACOs performed the same with and without risk adjustment for sex. (Figure 2)

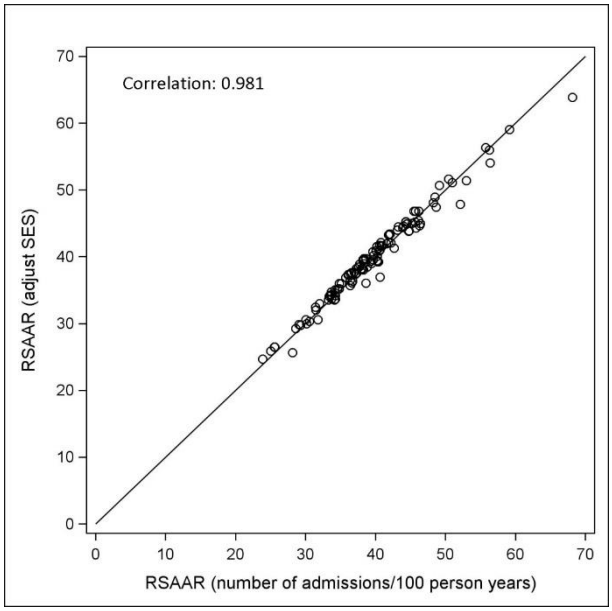
Figure 2. Plot of acute, unplanned admission rates with and without adjustment for sex



AHRQ SES Index

The deviance R-squared values for the two models – one adjusted for the 23 clinical variables *and* low SES, and one adjusted for the 23 clinical variables *without* adjusting for low SES – were 0.218 and 0.217, respectively, meaning adjustment for low SES explained the same variation and did not provide incremental benefit. Comparing the RSAAR with and without low SES included in the model resulted in a high degree of correlation (Spearman correlation = 0.981). The graph demonstrates that, compared with *not* adjusting for low SES, adjusting for low SES results in some ACOs having slightly lower RSAAR scores (below the line) and other ACOs having higher RSAAR scores (above the line). (Figure 3)

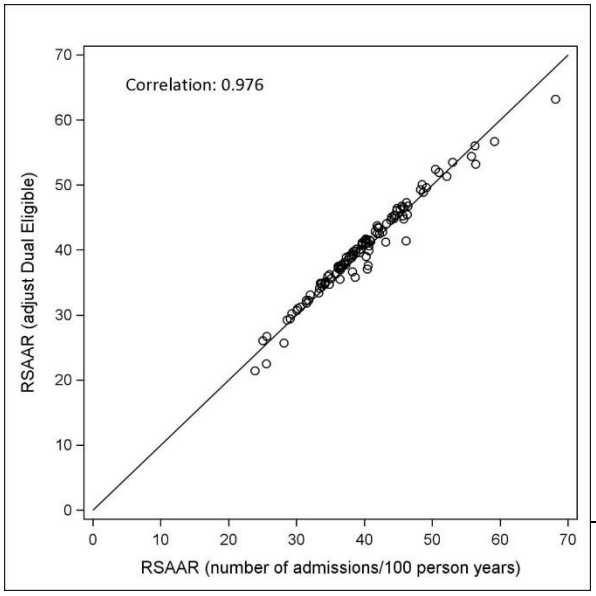
Figure 3. Plot of acute, unplanned admission rates with and without adjustment for AHRQ SES index



Medicaid Dual-Eligibility Status

The deviance R-squared values for the two models, one adjusted for the 23 clinical variables and Medicaid dual-eligibility status, and one adjusted for the 23 clinical variables *without* Medicaid dual-eligibility status, were 0.220 and 0.217, respectively, meaning adjustment for dual-eligibility status explained similar amount of variation and did not appreciably improve model fit. Comparing the RSAAR with and without Medicaid dual-eligibility status included in the model resulted in a high degree of correlation (Spearman correlation = 0.976). The graph demonstrates that, compared with *not* adjusting for Medicaid dual-eligibility status, adjusting for Medicaid dual-eligibility status results in some ACOs having slightly lower RSAAR scores (below the line) and other ACOs having higher RSAAR scores (above

Figure 4. Plot of acute, unplanned admission rates with and without adjustment for dual-eligibility status



In assessing the relationship between the proportions of low SES patients enrolled in an ACO and ACO measure performance, we found that ACOs serving many low SES patients more often perform worse than the national rate compared with ACOs serving few low SES patients. This was true using either the AHRQ SES index (32.1% vs. 3.4%, respectively) or Medicaid dual-eligibility status (28.6% vs. 3.4 %, respectively) as an indicator of patients' SES. However, among ACOs serving many low SES patients, using the AHRQ SES index, 8 ACOs (28.6%) performed 'better than the national rate;' using Medicaid dual-eligibility status as an indicator, 7 ACOs (25.0%) performed 'better than the national rate.'

Figure 5. Boxplots of risk-standardized acute admission rates (RSAARs), comparing ACOs with varying proportions of low SES patients with diabetes (based on AHRQ SES Index; Quartile 1 [Q1]: ACOs with few low SES patients; Quartile 4 [Q4]: ACOs with many low SES patients)

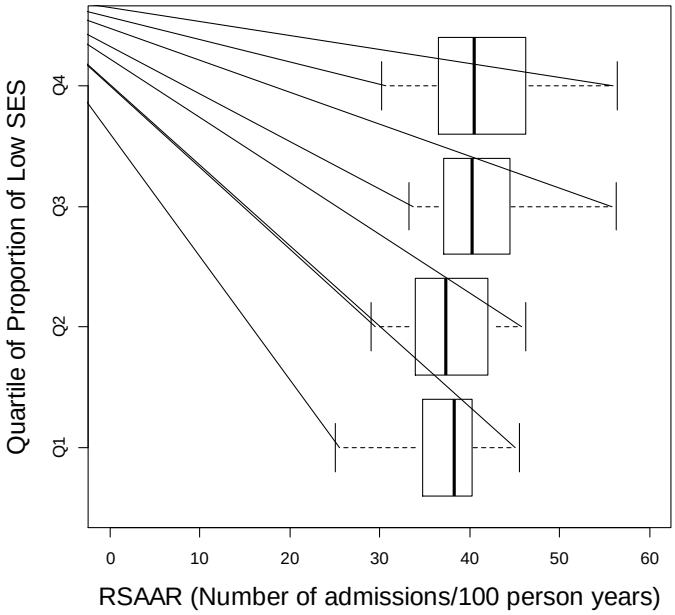
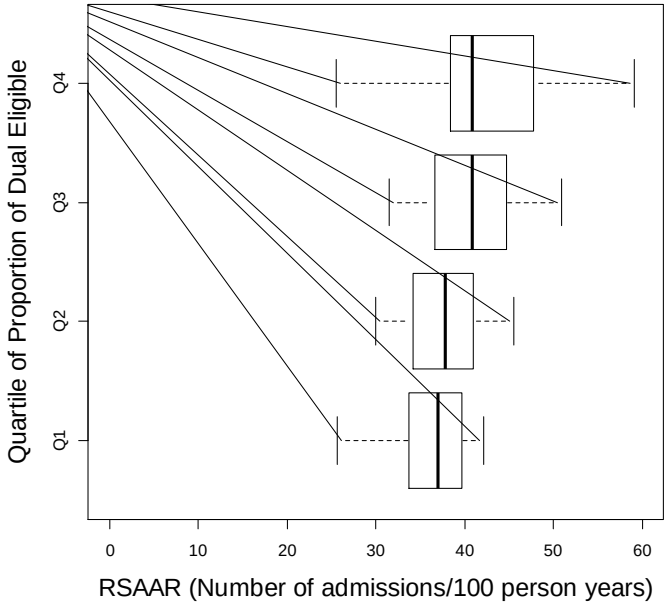


Figure 6. Boxplots of risk-standardized acute admission rates (RSAARs), comparing ACOs with varying proportions of Medicaid dual-eligible patients with diabetes (Quartile 1 [Q1]: ACOs with few Medicaid dual-eligible patients; Quartile 4 [Q4]: ACOs with many Medicaid dual- eligible patients)



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*)

We assessed adequacy of the patient-level risk-adjustment model (described above). We evaluated model performance first in the 2012 Development Sample. We then validated the model performance in the 2012 Validation Sample.

The measure uses the number of acute unplanned hospital admissions per person-year at risk for admission. Because the outcome is a count of hospital admissions – rather than a binary outcome, such as whether or not a patient has been admitted – several routinely used metrics of model performance cannot be applied (for example, we cannot use a c-statistic).

Using the 2012 Development Sample, we computed two summary statistics for assessing the risk-adjustment model performance: goodness-of-fit statistics (deviance R squared) and overfitting indices. We then compared the model performance in the development sample with its performance in the validation sample.

Deviance R squared

Our measure uses a negative binomial function because the outcome is a count of hospital admissions with over-dispersion. We calculated deviance R squared using the deviance residual defined by Cameron [1]. The deviance R squared evaluates how successful the fit is in explaining the variation of the data. Deviance R squared can take on any value between 0 and 1, with a value closer to 1 indicating that a greater proportion of deviance is accounted for by the model. For example, a deviance R squared value of 0.21 means that the fit explains 21% of the total deviance.

Overfitting indices
Overfitting refers to the phenomenon in which a model accurately describes the relationship between the predictive variables and the outcome in the development dataset, but fails to provide valid predictions in new patients.

Model performance among patients at different risk of admission
In order to determine whether the model performs well across groups of patients at different risk of admission, the sample was divided into quartiles of predicted admission rate (highest, second highest, lowest, and second lowest). We then assessed the model probability of the number of admissions compared with the observed probability of the number of admissions.

Generally, residuals measure the departure of fitted values from actual values of the dependent variable, but they cannot be applied to count data. For linear models, a residual is easily defined as the difference between actual and fitted values. For nonlinear models, the definition of a residual is not unique. Specifically, for count data, the raw residual (the observed value minus the fitted value) is heteroskedastic and asymmetric. Therefore, there is no residual that has zero mean, constant variance, and symmetric distribution. For fully parametric models such as negative binomial models, we can compare *predicted* probabilities with *observed* probabilities of each count of admissions. For each patient, we can calculate the *predicted* probability of being admitted to the hospital *n* times (0, 1, 2, …*n*) given this patient's risk factors for hospitalization. For example, a patient has a single predicted admission rate of 2.5 admissions per person-years of exposure; however, given the assumed negative binomial distribution of the risk of admissions, we can also express the patient's risk of admission as the probabilities of observing 0, 1, 2,…10 hospital admissions. Therefore, for each patient, we can calculate a set of predicted probabilities of observing different counts of admissions. The *predicted* probability for a group of patients is the average probability of observing 0, 1, 2, …*n* hospital admissions, given these patients' risk factors for admission. The *observed* probability of each count of admissions for a group of patients is the proportion of these patients admitted to the hospital 0, 1, 2, …*n* times.

Citations
1. Cameron AC, Windmeijer FAG. R-Squared Measures for Count Data Regression Models with Applications to Health-Care Utilization. Journal of Business & Economic Statistics. 1996;14(2):209-220.

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*)**:**

2012 Development Sample results (deviance R squared): 0.217
2012 Validation Sample results (deviance R squared): 0.218

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

2012 Development Sample calibration results (overfitting index): (0.0000, 1.0000)
2012 Validation Sample calibration results (overfitting index): (0.0017, 1.0031)

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

Below are plots of observed vs. predicted probabilities for the number of hospital admissions among four groups of patients: lowest (A) and second lowest (B) predicted admissions; and second highest (C) and highest (D) predicted admissions in the 2012 Development Sample.

Figure 7. Observed vs. predicted probabilities for the number of hospital admissions among lowest predicted admission group. Lowest predicted admission group (8 to 17 admissions per 100 person-years, median: 13 and interquartile range (IQR): 11 to 15.

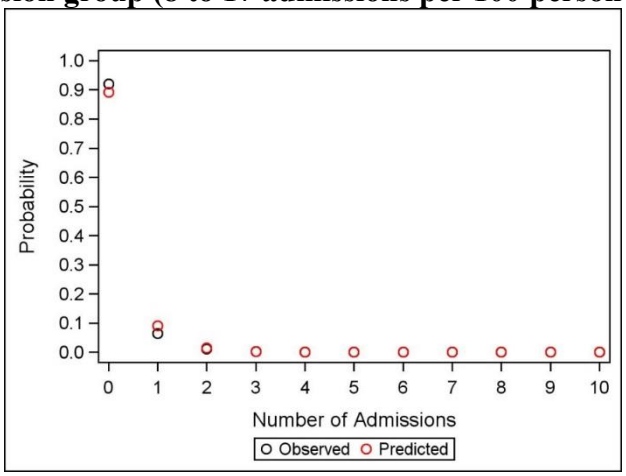


Figure 8. Observed vs. predicted probabilities for the number of hospital admissions among second lowest admission group (17 to 27 admissions per 100 person-years, median: 21, IQR: 19 to 24.

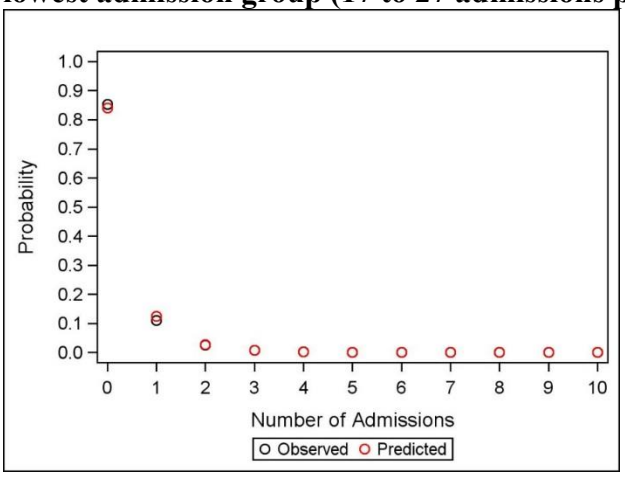


Figure 9. Observed vs. predicted probabilities for the number of hospital admissions among second highest admission group (27 to 53 admissions per 100 person-years, median: 36 IQR: 31 to 43.

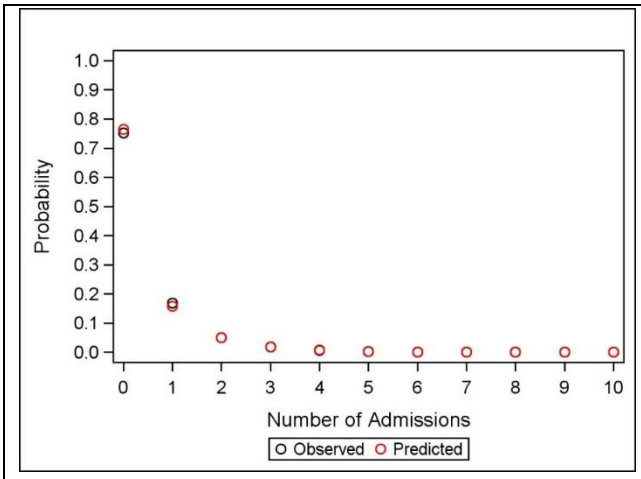
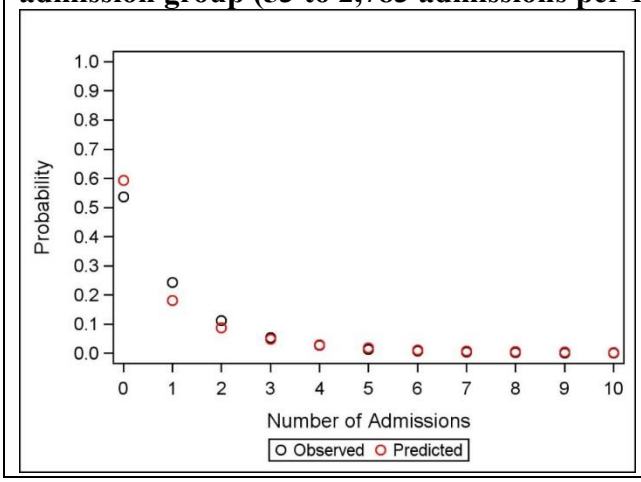


Figure 10. Observed vs. predicted probabilities for the number of hospital admissions among highest admission group (53 to 2,783 admissions per 100 person-years, median: 96, IQR: 69 to 158.



2b4.9. Results of Risk Stratification Analysis:

This measure is not risk stratified.

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*)

Model performance was similar in the development and validation datasets, with strong model discrimination and fit. The over-fitting index of γ_0 close to 0 and γ_1 close to 1 indicates good calibration of the model [1]. Additionally, the risk plots of all four risk groups show that the model performs well across a broad range of risk. In the highest risk group (2b4.8, Figure D), we observed that the model somewhat over-predicts the probability of 0 admissions and somewhat under-predicts the probability of 1 or 2 admissions. In the highest risk group, we observed that the observed and predicted probabilities of the number of 0, 1, or 2 admissions differed slightly. However, these differences were small and somewhat expected among the highest risk group of patients.

Citations

1. Cameron AC, Windmeijer FA. R-squared measures for count data regression models with applications to health-care utilization. *Journal of Business & Economic Statistics.* 1996;14(2):209-220.

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*)

Not applicable.

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)*

The method for discriminating facility-level performance for public reporting has not been determined. For publicly reported readmission measures of hospital outcomes developed with similar methodology, CMS currently estimates an interval estimate for each risk-standardized rate to characterize the amount of uncertainty associated with the rate, compares the interval estimate to the national crude rate for the outcome, and categorizes hospitals as 'better than the national rate,' 'worse than the national rate,' or 'no different than the national rate.' We used that approach here. However, the approach to discriminating performance that would be used for this measure in public reporting has not been determined.

In order to determine interval estimates (IEs), we used bootstrapping methods. In brief, we randomly sampled 114 ACOs with replacement. This is done by randomly selecting an ACO from the 114 ACOs, then placing the selected ACO back into the pool, until we got 114 ACOs, with some ACOs being selected more than once. Performance scores were calculated for each random sample of 114 ACOs. If some ACOs were selected more than once in a bootstrapped sample, we treated them as distinct so that we had random effects to estimate the variance components. This process was repeated many times until 3,000 results were obtained for each ACO.

Using the 95% IEs, we assigned each ACO to one of three performance categories: 'better than the national rate,' 'no different than the national rate,' and 'worse than the national rate.' Each ACO was compared to all Medicare FFS beneficiaries who met our diabetes cohort criteria, so that each ACO was evaluated against the US national admission rate among Medicare FFS patients with diabetes. The ACO was 'better than the national rate' if the 95% IE was completely below the US national Medicare FFS rate among patients with diabetes; 'no different than the national rate' if the 95% IE included the US national Medicare FFS rate among patients with diabetes; and 'worse than the national rate' if the 95% IE was above the US national Medicare FFS rate among patients with diabetes.

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*)

51 (44.7%) ACOs were 'no different than the national rate,' 45 (39.5%) were 'better than the national rate,' and 18 (15.8%) were 'worse than the national rate' of admissions per 100 person-years at risk for hospitalization among the US national Medicare FFS diabetes patient population.

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?*)

These results suggest there are meaningful differences in the quality of care received for patients in the 114 ACOs in the ambulatory setting.

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.***

Items 2b6.1-2b6.3 skipped, as this measure has only one set of specifications.

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*)

Not applicable.

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*)

Not applicable.

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*)

Not applicable.

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*)

Not applicable.

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)

Not applicable.

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*)

Not applicable.

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.**

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)*.

Not applicable. There are no fees, licensing, or other requirements to use any aspect of the measure as specified.

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)
Payment Program	

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

Measure is currently not in use.

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?*)

This measure is not currently publicly reported or used in an accountability application because it only recently completed development. However, in the November 13, 2014 Physician Fee Schedule final rule, CMS finalized adding the measure to the Medicare Shared Savings Program quality measure set (see 79 FR 67912; https://www.gpo.gov/fdsys/pkg/FR-2014-11-13/pdf/2014-26183.pdf).

The measure is planned for pay-for-reporting in the Medicare Shared Savings Program for 2015 and 2016 reporting periods (79 FR 67912, 67916) and for pay-for-performance in the Medicare Shared Savings Program beginning 2017 reporting period (79 FR 67912, 67916).

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 will be used in one or more CMS programs as noted above in 4a.2. The measure has been finalized for use in the Medicare Shared Savings Program. The measure will be pay-for-reporting initially for the 2015 and 2016 reporting periods and then as pay-for-performance beginning in the 2017 reporting period.

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**

Not applicable

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.

The measure is not currently used in a quality improvement program, but the primary goal of the measure is to provide ACOs with information necessary to implement focused quality improvement.

This measure was evaluated by a group of clinical experts and a technical expert panel (TEP) throughout the measure development process. We received input and feedback on key methodological, clinical, and other measure decisions as well as on its utility in

guiding focused quality improvement within ACOs.

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.**

In designing the measure, we sought to minimize the potential of this measure to result in the denial of future care to high-risk individuals. We developed the patient cohort exclusions and risk-adjustment model to ensure providers who care for patients at higher risk of admission will not be disadvantaged in the measure. CMS is committed to monitoring this measure's use and assessing potential unintended consequences over time.

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)**

0018 : Controlling High Blood Pressure
0059 : Comprehensive Diabetes Care: Hemoglobin A1c (HbA1c) Poor Control (>9.0%)
0063 : Comprehensive Diabetes Care: LDL-C Screening
0272 : Diabetes Short-Term Complications Admission Rate (PQI 01)
0274 : Diabetes Long-Term Complications Admission Rate (PQI 03)
0285 : Lower-Extremity Amputation among Patients with Diabetes Rate (PQI 16)
0575 : Comprehensive Diabetes Care: Hemoglobin A1c (HbA1c) Control (<8.0%)
0638 : Uncontrolled Diabetes Admission Rate (PQI 14)
0709 : Proportion of patients with a chronic condition that have a potentially avoidable complication during a calendar year.

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?

No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**

The measures listed above differ in several important ways from the proposed measure: 1. The measure differs in the outcome. The NQF# 0018, 0059, 0063, and 0575 are measures of surrogate outcomes and focus on risk factor control; in contrast, the proposed measure directly evaluates the results of care and assesses an outcome experienced by patients. The NQF # 0709, 0272, 0274, 0638, and 0285 are measures of specific types of hospital admissions; in contrast, the proposed measure includes all-cause acute

admissions to capture broad vulnerabilities of older patients with diabetes to acute exacerbations of their underlying condition as well as co-existing comorbidities. 2. The measure differs in risk adjustment. The existing measures are either not adjusted or adjusted for age and sex. In contrast, the proposed measure is fully adjusted for a broad range of clinical factors that contribute to the risk for admission, allowing for fair comparisons of ACO performance. 3. The measure differs in the target population. Existing measures include adults with ages 18 to 75 or 18 to 65 years of age. In contrast, the target population for the proposed measure are all Medicare FFS beneficiaries with a diagnosis of diabetes, who are 65 years or older. Thus, the focus is focus is on older, complex adults with diabetes.

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.)
Not applicable.

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:** Diabetes_ACO_Admission_Measure_NQF_Appendix_01-29-16_v1.0.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services (CMS)
Co.2 **Point of Contact:** Vinitha, Meyyur, Vinitha.meyyur@cms.hhs.gov, 410-786-8819-
Co.3 **Measure Developer if different from Measure Steward:** Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (CORE)
Co.4 **Point of Contact:** Elizabeth, Drye, Elizabeth.drye@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.
TEP Members:

CORE convened a TEP of clinicians, patients, purchasers, and experts in quality improvement to provide input on key methodological decisions.

Lawrence M. Becker, BS – Xerox Corporation (Director, Strategic Partnerships, Alliances and Analytics)
Alex Blum, MD, MPH – Evergreen Health Cooperative (Chief Medical Officer)
Sanjay Doddamani, MD – Geisinger Health System (System-wide Chief of Advanced Cardiac Disease HF)
Kevin Fiscella, MD, MPH – University of Rochester Medical Center (Professor of Family Medicine)
Elbert Huang, MD, MPH – University of Chicago (Associate Professor of Medicine, Director of the Center for Translational and Policy Research of Chronic Diseases, and Associate Director of the Chicago Center for Diabetes Translation Research)
Bruce Leff, MD – Johns Hopkins University School of Medicine (Professor of Medicine, Division of Geriatric Medicine); The Johns Hopkins University Bloomberg School of Public Health (Faculty, Health Services Research Development Center and Lipitz Center for Integrated Health Care)
Andy Miller, MD, MPH – Healthcare Quality Strategies, Inc. (Medical Director); Colorado Foundation for Medical Care (CMO, Integrating Care for Populations & Communities National Coordinating Center)

Ami Parekh, MD, JD – University of California, San Francisco (Medical Director for Health System Innovation)
Christine Ritchie, MD – University of California, San Francisco (Professor of Medicine, Division of Geriatrics)
Two patients with chronic conditions (anonymous)

CORE Measure Development Team:
Faseeha Altaf, MPH – Research Project Coordinator
Haikun Bao, PhD – Lead Analyst, diabetes measure
Susannah Bernheim, MD, MHP – Director of CMS Projects; Clinical Investigator
Kanchana Bhat, MPH – Senior Project Manager
Ying Dai, PhD – Lead Analyst
Weiwei Zhang, MPH – Supporting Analyst
Elizabeth Drye, MD, SM – Project Director; Project Lead, MCCs measure
Elizabeth Eddy, BA – Research Project Coordinator
Leora Horwitz, MD, MHS – Clinical Investigator
Erin Joyce, BA – Research Assistant
Zhenqiu Lin, PhD – Managing Analyst
Harlan Krumholz, MD, SM – Director, CORE
Kasia Lipska, MD, MHS – Project Lead
Julia Montague, MPH – Research Project Coordinator II/Project Manager
Craig Parzynski, MS – Supporting Analyst
Joseph Ross, MD, MHS – Clinical Investigator, CORE
Erica Spatz, MD, MHS – Project Lead
La'Mont Sutton, MPH – Research Associate
Vera Zhang, MPH – Supporting Analyst

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:**
Ad.3 **Month and Year of most recent revision:**
Ad.4 **What is your frequency for review/update of this measure?** Not applicable.
Ad.5 **When is the next scheduled review/update for this measure?**

Ad.6 **Copyright statement:** Not applicable.
Ad.7 **Disclaimers:** Not applicable.

Ad.8 **Additional Information/Comments:** Not applicable.

NATIONAL QUALITY FORUM

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 #: 2888
De.2. **Measure Title:** Risk-Standardized Acute Admission Rates for Patients with Multiple Chronic Conditions
Co.1.1. **Measure Steward:** Centers for Medicare & Medicaid Services (CMS)
De.3. **Brief Description of Measure:** Rate of risk-standardized acute, unplanned hospital admissions among Medicare fee-for-service (FFS) patients 65 years and older with multiple chronic conditions (MCCs)
1b.1. **Developer Rationale:** As of 2010, more than two-thirds of Medicare beneficiaries had been diagnosed with or treated for two or more chronic conditions [1]. People with MCCs are more likely to be admitted to the hospital than those without chronic conditions or with a single chronic condition. Additionally, they are more likely to visit the emergency department, use post-acute care (such as skilled nursing facilities), and require home health assistance [1]. No quality measures specifically designed for this population exist to assess quality of care or to enable the evaluation of whether current efforts to improve care are successful; this measure is designed to help fill that gap as called for in NQF's "Multiple Chronic Conditions Measurement Framework." [2]

The measure is focused on ACOs because better, coordinated care should lower the risk of hospitalization for this vulnerable population. The measure is designed to illuminate variation in hospital admission rates and incentivize ACOs to develop efficient and coordinated chronic disease management strategies that anticipate and respond to patients' needs and preferences. The measure is also consistent with ACOs' commitment to deliver patient-centered care that fulfills the goals of the Department of Health and Human Services' National Quality Strategy – improving population health, providing better care, and lowering health care costs [3].

The rationale for measuring all-cause acute admissions is to assess the quality of care as experienced by the patient and to drive overall improvements in care quality, coordination, and efficiency that are not specific to certain diseases. Ambulatory care providers can act together to lower patients' risk for a wide range of acute illness requiring admission in several ways:

1. Provide optimal and accessible chronic disease management to reduce catastrophic sequelae of chronic disease. For example:
a. Support healthy lifestyle behaviors and optimize medical management to minimize the risk for cardiovascular events such as stroke and heart attacks; and
b. Carefully monitor and act early to address chronic problems that require major interventions if allowed to progress (for example, assessment and treatment of peripheral artery disease in unresolving infections in order to prevent amputation).

2. Anticipate and manage the interactions between chronic conditions. For example:
a. Closely monitor renal function in patients on diuretic therapy for heart failure and chronic kidney disease;
b. Minimize polypharmacy to reduce drug-drug and drug-disease interactions; and
c. Assess and treat depression to improve self-efficacy and self-management of chronic disease.

3. Provide optimal primary prevention of acute illnesses, such as recommended immunizations and screening.

4. Facilitate rapid, effective ambulatory intervention when acute illness does occur, whether related or unrelated to the chronic conditions. For example:
a. Promptly prescribe antibiotics for presumed bacterial pneumonia and diuretic treatment for fluid overload in heart failure;
b. Empower patients to recognize symptoms and to seek timely care; and
c. Create accessible care options for patients (e.g., weekend or evening hours; capacity to deliver intravenous medications).

5. Partner with the government, local businesses, and community organizations to improve support for patients with chronic illness. For example:
a. Collaborate with home nursing programs;
b. Partner with local businesses to increase opportunities to engage in healthy lifestyle behaviors; and
c. Provide outreach and services at senior centers.

Finally, a number of studies have shown that improvements in the delivery of healthcare services for ambulatory patients with MCCs can lower the risk of admission [4-9]. Demonstrated strategies include improving access to care; supporting self-care in the home; better coordinating care across providers; and integrating social work, nursing, and medical services.

Citations:
1. Centers for Medicare and Medicaid Services. Chronic Conditions Among Medicare Beneficiaries, Chartbook: 2012 Edition. 2012; http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/Downloads/2012Chartbook.pdf. Accessed March 18, 2014.

2. National Quality Forum (NQF). Multiple Chronic Conditions Measurement Framework. 2012; http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=71227

3. U.S. Department of Health and Human Services. Multiple chronic conditions—A strategic framework: Optimum health and quality of life for individuals with multiple chronic conditions. December 2010; http://www.hhs.gov/ash/initiatives/mcc/mcc_framework.pdf. Accessed March 20, 2014.

4. Chan CL, You HJ, Huang HT, Ting HW. Using an integrated COC index and multilevel measurements to verify the care outcome of patients with multiple chronic conditions. BMC health services research. 2012 2012;12:405.

5. Dorr DA, Wilcox AB, Brunker CP, Burdon RE, Donnelly SM. The effect of technology-supported, multidisease care management on the mortality and hospitalization of seniors. Journal of the American Geriatrics Society. Dec 2008;56(12):2195-2202.

6. Levine S, Steinman BA, Attaway K, Jung T, Enguidanos S. Home care program for patients at high risk of hospitalization. American Journal of Managed Care. 2012 Aug 2012;18(8):e269-276.

7. Centers for Medicare & Medicaid Services (CMS). Medicare Health Support. 2012; https://www.cms.gov/Medicare/Medicare-General-Information/CCIP/. Accessed March 27, 2014.

8. Littleford A, Kralik D. Making a difference through integrated community care for older people. Journal of Nursing and Healthcare of Chronic Illness. 2010;2(3):178-186.

9. Sommers LS, Marton KI, Barbaccia JC, Randolph J. Physician, nurse, and social worker collaboration in primary care for chronically ill seniors. Arch Intern Med. Jun 26 2000;160(12):1825-1833.

10. Zhang NJ, Wan TT, Rossiter LF, Murawski MM, Patel UB. Evaluation of chronic disease management on outcomes and cost of care for Medicaid beneficiaries. Health policy (Amsterdam, Netherlands). May 2008;86(2-3):345-354.Brown RS, Peikes D, Peterson G, Schore J, Razafindrakoto CM. Six features of Medicare coordinated care demonstration programs that cut hospital admissions of high-risk patients. Health Affairs. 2012 Jun 2012;31(6):1156-1166.

S.4. **Numerator Statement:** The outcome measured for each patient is the number of acute, unplanned admissions per 100 person-years at risk for admission. Persons are considered at risk for admission if they are alive, enrolled in FFS Medicare, and not currently admitted. (See S.6, Numerator Details, for more information.)
S.7. **Denominator Statement:** Our target population is Medicare FFS patients aged 65 years and older whose combinations of chronic conditions put them at high risk of admission and whose admission rates could be lowered through better care. The National Quality Forum's (NQF's) "Multiple Chronic Conditions Measurement Framework," which defines patients with multiple chronic conditions as people "having two or more concurrent chronic conditions that…. act together to significantly increase the complexity of management, and affect functional roles and health outcomes, compromise life expectancy, or hinder self-management [1]."

Operationally, the measure cohort includes patients with diagnoses in two or more of eight chronic disease groups:

1. Acute myocardial infarction (AMI)
2. Alzheimer's disease and related disorders or senile dementia
3. Atrial fibrillation
4. Chronic kidney disease (CKD)
5. Chronic obstructive pulmonary disease (COPD) and asthma
6. Depression
7. Heart failure
8. Stroke and transient ischemic attack (TIA)

This approach captures approximately 25% of Medicare FFS beneficiaries aged 65 years and older with at least one chronic condition (about 5 million patients in 2012).

Citations:
1. National Quality Forum (NQF). Multiple Chronic Conditions Measurement Framework. 2012; http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=71227
S.10. **Denominator Exclusions:** The measure excludes:
1. Patients without continuous enrollment in Medicare Part A for the duration of the measurement period (or until death).
Rationale: We exclude these patients to ensure full data availability for outcome assessment (Part A during the measurement year).

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Administrative claims
S.26. **Level of Analysis:** Integrated Delivery System

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date: n/a

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?** Not applicable.

New Measure -- 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. The emphasis for maintaining endorsement is noted for each criterion.

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.

- The developer notes improvements in access to care, supporting self-care in the home, better coordinating care across providers, and integrating social work, nursing, and medical services all have the potential to improve admission rates for patients with multiple chronic conditions.

Question for the Committee:

- *Did the developer provide at least one health care structure or process that an ACO can undertake to improve this outcome?*

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 developer provides the following information:

- Using data from the 2012 Medicare Full Sample with 4,937,344 patients, that was composed of 239,551 patients in 114 ACOs, and compared with the 71.9 admissions (per 100 person-years) - the US national Medicare FFS rate of acute, unplanned admissions among patients with MCCs, they found that:
  - The mean risk-standardized acute admission rate (RSAAR) among ACOs for year 2012 was 69.3, median was 68.5.
  - They observed that 45 ACOs (39.5%) had RSAARs that were 'no different than the national rate' and 22 ACOs (19.3%) had RSAAR scores 'worse than the national rate,' and 47 ACOs (41.2%) were 'better than the national rate.'

- The developer provided data from ACO performance score using the 2012 Medicare Full Sample which showed the crude US national Medicare FFS rate of acute, unplanned admissions among patients with MCCs, they found that:
  - The mean risk-standardized acute admission rate (RSAAR) among ACOs for year 2012 was 69.3, median was 68.5.
  - They observed that 45 ACOs (39.5%) had RSAARs that were 'no different than the national rate' and 22 ACOs (19.3%) had RSAAR scores 'worse than the national rate,' and 47 ACOs (41.2%) were 'better than the national rate.'

Disparities

- The developer reports that they examined disparities in ACO performance based on the proportion of patients of low socioeconomic status (SES) being cared for by each ACO.
- The developer found that performance scores did not change appreciably after adjusting the models for patients' SES. The Spearman correlation comparing the ACO measure scores estimated with and without risk adjustment for the AHRQ SES Index was 0.992. Similarly, the Spearman correlation for the scores estimated with and without patients' Medicaid dual eligibility was 0.994. These results demonstrate that adjusting for SES at the patient level has little effect on the measure score.
- Overall, results indicate that SES status plays little role at the patient level, thus measure was not adjusted for patient-level SES. According to the developer, ACOs should and do influence a broad range of patient and community-level factors that can mitigate the risk of admission associated with low SES, and do not want to adjust for modifiable factors.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Given the developer disparities testing results, does the Committee agree that SDS adjustment is not warranted?*

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)

1. Importance to Measure and Report

1a. Evidence to Support Measure Focus

Comments: **Improvements in access to care, supporting self-care in the home, better coordinating care across providers, and integrating social work, nursing, and medical services all have the potential to improve admission rates for patients with multiple chronic conditions.

1b. Performance Gap

Comments: **• The developer provided data from ACO performance score using the 2012 Medicare Full Sample which showed the crude US national Medicare FFS rate of acute, unplanned admissions among patients with MCCs, they found that:

o The mean risk-standardized acute admission rate (RSAAR) among ACOs for year 2012 was 69.3, median was 68.5.

They observed that 45 ACOs (39.5%) had RSAARs that were 'no different than the national rate' and 22 ACOs (19.3%) had RSAAR scores 'worse than the national rate,' and 47 ACOs (41.2%) were 'better than the national rate.' SES status plays little role at the patient level, thus measure was not adjusted for patient-level SES. According to the developer, ACOs should and do influence a broad range of patient and community-level factors that can mitigate the risk of admission associated with low SES, and do not want to adjust for modifiable factors.

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

Comments: **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: Administrative claims

Specifications:

- This measure calculates the Rate of risk-standardized acute, unplanned hospital admissions among Medicare fee-for-service (FFS) patients 65 years and older with multiple chronic conditions (MCCs).
- This is a health outcome measure and the level of analysis is Integrated Delivery System.
- The Numerator is the outcome measured for each patient is the number of acute, unplanned admissions per 100 person-years at risk for admission. Persons are considered at risk for admission if they are alive, enrolled in FFS Medicare, and not currently admitted.
- The Denominator is the Medicare FFS patients aged 65 years and older whose combinations of chronic conditions put them at high risk of admission and whose admission rates could be lowered through better care.

Questions for the Committee :

- *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.

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

- Datasets used for testing included Medicare Parts A and B claims, the denominator file, the Medicare Provider Analysis and Review (MedPAR) file, and the American Community Survey to derive the AHRQ SES index.
- **Data element reliability**:
  - With regard to data element reliability, the developer notes that the measure has been developed to avoid the use of claims data elements that are thought to be coded inconsistently across hospitals or providers, instead using fields that are consequential for payment and which are audited by CMS.
  - In addition, the developer compared frequencies and odds ratios of variables from their risk model to assess the consistency of those variables across samples.
- **Performance score reliability:**
  - The developer defines performance score reliability as the degree to which repeated measurements of the same entity agree with each other.
  - In line with this thinking, the developer's approach to assessing score-level reliability was to consider the extent to which assessments of a hospital using different but randomly-selected subsets of patients produce similar measures of hospital performance. The developers refer to this as a "test-retest" approach; it may also be called a "split-half" method. This is generally considered an appropriate method of testing reliability.

Results of reliability testing

- **Data element reliability**:

- Summarizing the results of this analysis, the developer notes that the mean age and frequency of risk-adjustment variables was similar among the two samples of 2012 data suggesting that the data elements are reliable across the samples.

- **Performance score reliability:**
  - The 2012 full Medicare sample was divided into two subsets of patients randomly. The developer calculated the measure score of all ACOs for each of the two subsets of patients. Each ACO was measured twice, but each measurement was make using distinct sets of measures. The interclass correlation coefficient (ICC) for the two subsets of patients was 0.84, which can be interpreted as excellent correlation, and thus reliable.

Guidance from the Reliability Algorithm

- Question 1. Submitted specifications are precise, unambiguous, and complete. Measure can be consistently implemented.
- Question 2. Empirical reliability testing was conducted using statistical tests with the measure as specified.
- Question 3. Empirical validity testing of patient-level data was conducted.
- Question 4. Reliability testing was conducted with computed performance measure scores for each measured entity.
- Question 5. Random split-half correlation was used to assess the proportion of variability due to real differences among the measured entities.
- Question 6. The ICC was 0.84 which is considered an excellent level of agreement.

Questions for the Committee:

- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*
- *Does the measure testing match the measure specifications?*

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.

- This measure estimates the predicted number of admissions given the Accountable Care Organization's (ACO's) case mix, sample size, and actual admission rate. The outcome for this measure is the number of acute, unplanned admissions per 100 person-years at risk for admission. The outcome includes inpatient admissions to an acute care hospital for any cause during the measurement year, unless an admission is identified as "planned."

Question for the Committee:

- *Are the specifications clear?*

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

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.

- The developer tested the validity of the measure using three different methods:
    - Validity of the claims-based measures. The developer argues that other NQF endorsed mortality and readmission measures have been validated by comparing the claims to the medical records data elements. It is unclear if the risk adjustment validation approach that the developer cites is sufficiently similar to this measure and for this level of analysis and ambulatory patients.
    - The developer also notes that this measure has been validated by using established measure development guidelines. While an important step for measure development, this method of validity testing has generally not be considered sufficient for demonstrating measure validity.
    - Finally, the measure developer completed a systemic face validity assessment of this measure with 9 experts and two patients agreeing that this measure was a valid indicator of health care quality.

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

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:

- Out of the total Medicare FFS patients with multiple chronic conditions (N=5,070,533), the developer excluded 133,189 due to non-continuous enrollment in part A in 2012.

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 developers provided a conceptual framework that was used to develop the risk adjustment model for this measure. This conceptual framework included 4 contextual domains that influence ACO performance including, physical environment, community resources, patient resources, and patient behavioral/personal preferences.
- The measure included demographic factors, and clinical risk factors present at the start of the measurement period.
    - The measure developers reviewed 189 diagnosis groups included in the hierarchical condition category (HCC), and calculated the prevalence of each CC in the year preceding the measurement period. After examining the bi-variate analysis, the developers reduced the list to 46 candidate variables including age.
    - The measure developers did not adjust for contextual factors that impact admissions; however, they did provide data demonstrating that including SDS adjustment did not make a meaningful difference to the measure score of the ACOs. The spearman correlation coefficient that estimated the difference in performance with and without SDS adjustment was 0.992. Thus, the results demonstrate that adjustment had little effect on the measure score.
- **Risk Model Diagnostics:**
    - To assess the overall performance of their risk-adjustment model, the developers computed two summary statistics, including:

- *Risk model discrimination statistics* (the model's ability to explain how successful the fit is in explaining the variation of the data. In this case, the r-sq value was 0.123. In other words, the model was able to explain 12.3% of the total deviance.
- Overfitting indices (model calibration) [presented as (γ0, γ1)]:
  - The developer states that if the γ0 in the validation samples are substantially far from zero and the γ1 is substantially far from one, there is potential evidence of over-fitting. The calibration value of close to 0 at one end and close to 1 to the other end indicates good calibration of the model.
    - 2012 Development Sample (Index): (**0,1**)
    - 2012 Validation Sample: **(-0.0015, 1.0011)**

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 note that the methodology to publicly report this measure has not been determined yet
- For other publically reported measures with the same methodology, CMS categories hospitals at "better than the national rate", "worse than the national rate" and "no different than the national rate".
  - For this measure, 45 ACOs (39.5%) performed no different than the national rate, 47 (41.2%) performed better then the national rate, and 22 (19.3%) performed worse than the national rate. The developers suggest that this demonstrates that there is a meaningful different in performance on this measure.

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

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)

2. Scientific Acceptability of Measure Properties

2a1. & 2b1. Specifications

Comments: **Specification consistent with evidence

2a2. Reliability Testing

Comments: **Reliability testing performed with the data source and level of analysis indicated for the measure.

The developers refer to this as a "test-retest" approach; it may also be called a "split-half" method. This is generally considered an appropriate method of testing reliability.

2b2. Validity Testing

Comments: **Measure developer completed a systemic face validity assessment of this measure with 9 experts and two patients agreeing that this measure was a valid indicator of health care quality.

Face validity testing performed.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **The measure developers did not adjust for contextual factors that impact admissions; however, they did provide data demonstrating that including SDS adjustment did not make a meaningful difference to the measure score of the ACOs. The spearman correlation coefficient that estimated the difference in performance with and without SDS adjustment was 0.992. Thus, the results demonstrate that adjustment had little effect on the measure score.

Good calibration of the model.

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.

The developer states:

- ALL measure data elements are in defined fields in electronic claims and routinely generated or collected by and used by healthcare personnel during the provision of care, coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims).
- There is no cost associated with data collection.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*

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

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3. Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **ALL measure data elements are in defined fields in electronic claims and routinely generated or collected by and used by healthcare personnel during the provision of care, coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims).

There is no cost associated with data collection.

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**
OR
Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

- The developer states:
  - On the November 13, 2014 Physician Fee Schedule final rule, CMS finalized adding the measure to the Medicare Shared Savings Program quality measure set (see 79 FR 67912; https://www.gpo.gov/fdsys/pkg/FR-2014-11-13/pdf/2014-26183.pdf).
  - The measure is planned for pay-for-reporting in the Medicare Shared Savings Program for 2015 and 2016 reporting periods (79 FR 67912, 67916) and for pay-for-performance in the Medicare Shared Savings Program beginning 2017 reporting period (79 FR 67912, 67916). The measure will be pay-for-reporting initially for the 2015 and 2016 reporting periods and then as pay-for-performance beginning in the 2017 reporting period.

***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

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **Currently not publicly reported or use din accountability program.

Criterion 5: Related and Competing Measures

Related or competing measures

- No related or competing measures listed.

Harmonization

- Not applicable.

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: Risk-Standardized Acute Admission Rates for Patients with Multiple Chronic Conditions

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: 1/29/2016

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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: Click here to name the health outcome

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

The quality of care for patients with multiple chronic conditions (MCCs) is generally best assessed by examining outcomes rather than care processes [1]. Patients with MCCs vary in the objectives and goals set for their care; for example, people with the same conditions may place different values on alleviating symptoms, improving function, reducing the risk of acute events, or minimizing drug side effects. Therefore, disease-specific processes or intermediate measures addressed by traditional care measures may not be aligned with patient core preferences. In addition, disease-specific treatments may often be contraindicated in the context of co-existing comorbidities [2]. Moreover, surrogate or intermediate markers of outcomes, such as cholesterol levels, may not have the same relationship to outcomes of importance in patients with MCCs as they do in patients with the single condition [3]. In contrast, outcome measures can focus on endpoints of importance to patients that reflect how the combined care people receive affects their health. Hence, experts have recommended measuring several "universal" outcomes that include health status, functional status, symptom burden, and death in order to evaluate care for

patients with MCCs. Researchers have used additional outcomes [2-4], including admission rates, to assess the success of interventions to improve care.

This measure uses the outcome of acute unplanned admissions to assess care quality. We target this adverse event for several reasons. Patients with MCCs are typically frailer and at higher risk for hospitalizations due to, for example, potentially life-threatening exacerbations of their conditions and complications of complex treatment regimens [1, 3, 5]. They may be persistently physiologically stressed due to challenges maintaining adequate circulation, renal function, and respiration. Moreover, depression, dementia, and/or fatigue may contribute to the challenges they face implementing potentially complex care plans designed to maintain their health status, and their disease burden and treatment regimens in turn can affect their mental well-being. As a result, patients with MCCs may experience an increased vulnerability to common causes of admission including pneumonia and other infections, admissions due to exacerbations of their chronic conditions, and admissions related to frailty (for example, due to falls) [5]. Providers can potentially lower the risk of acute admissions in this high-risk population through better coordinated, more timely, and more effective health care. Hence, efforts to redesign care for patients with MCCs have used admission rates as one outcome to evaluate the success of interventions.

Citations:

1. National Quality Forum (NQF). Multiple Chronic Conditions Measurement Framework. 2012; http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=71227

2. Guiding principles for the care of older adults with multimorbidity: an approach for clinicians: American Geriatrics Society Expert Panel on the Care of Older Adults with Multimorbidity. Journal of the American Geriatrics Society. Oct 2012;60(10):E1-E25.

3. Uhlig K, Leff B, Kent D, et al. A Framework for Crafting Clinical Practice Guidelines that are Relevant to the Care and Management of People with Multimorbidity. J GEN INTERN MED. 2014/04/01 2014;29(4):670-679.

4. Tinetti ME, Fried TR, Boyd CM. Designing health care for the most common chronic condition—multimorbidity. JAMA. 2012;307(23):2493-2494.

5. U.S. Department of Health and Human Services. Multiple chronic conditions—A strategic framework: Optimum health and quality of life for individuals with multiple chronic conditions. December 2010; http://www.hhs.gov/ash/initiatives/mcc/mcc_framework.pdf. Accessed March 20, 2014.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

The rationale for measuring acute unplanned admissions for Accountable Care Organization (ACO) chronic disease patients is that ACOs are established precisely to improve patient-centered care and outcomes for these patients. Providers within an ACO share responsibility for delivering primary preventive services, chronic disease management, and acute care to patients with MCCs. Further, ACOs accept accountability for patient outcomes; providers form ACOs voluntarily and commit to the goals of the ACO program, which include providing better coordinated care and chronic disease management while lowering costs [1]. These program goals are fully aligned with the objective of lowering patients' risk of admission incentivized by the measure [1]. ACOs should be able to lower the risk of acute, unplanned admissions more feasibly than less integrated Medicare fee-for-service providers through strengthening preventive care, delivering better coordinated and more effective chronic disease management, and providing timely ambulatory care for acute exacerbations of chronic disease. ACOs may also need to engage with community organizations and health-related community services to facilitate effective chronic disease management. We provide a more detailed list of potential interventions in the Measure Submission Form under Section 1b.1.

Finally, a number of studies have shown that improvements in the delivery of health care services for ambulatory patients with MCCs can lower the risk of admission [2-7]. Demonstrated strategies include improving access to care, supporting self-care in the home, better coordinating care across providers, and integrating social work, nursing, and medical services. It is our vision that this measure will illuminate variation among ACOs in hospital admission rates for people with MCCs and incentivize ACOs to expand efforts to develop and implement efficient and coordinated chronic disease management strategies that anticipate and respond to patients' needs and preferences.

Citations:

1. Centers for Medicare & Medicaid Services (CMS). Accountable Care Organizations (ACOs): General Information. http://innovation.cms.gov/initiatives/aco/. Accessed September 25, 2014.

2. Dorr DA, Wilcox AB, Brunker CP, Burdon RE, Donnelly SM. The effect of technology-supported, multidisease care management on the mortality and hospitalization of seniors. Journal of the American Geriatrics Society. Dec 2008;56(12):2195-2202.

3. Levine S, Steinman BA, Attaway K, Jung T, Enguidanos S. Home care program for patients at high risk of hospitalization. American Journal of Managed Care. 2012 Aug 2012;18(8):e269-276.

4. Littleford A, Kralik D. Making a difference through integrated community care for older people. Journal of Nursing and Healthcare of Chronic Illness. 2010;2(3):178-186.

5. Chan CL, You HJ, Huang HT, Ting HW. Using an integrated COC index and multilevel measurements to verify the care outcome of patients with multiple chronic conditions. *BMC health services research.* 2012 2012;12:405.

6. Sommers LS, Marton KI, Barbaccia JC, Randolph J. Physician, nurse, and social worker collaboration in primary care for chronically ill seniors. Arch Intern Med. Jun 26 2000;160(12):1825-1833.

7. Zhang NJ, Wan TT, Rossiter LF, Murawski MM, Patel UB. Evaluation of chronic disease management on outcomes and cost of care for Medicaid beneficiaries. Health policy (Amsterdam, Netherlands). May 2008;86(2-3):345-354.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

Not applicable. This is an outcome measure.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Not applicable. This is an outcome measure.

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Not applicable. This is an outcome measure.

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Not applicable. This is an outcome measure.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

Not applicable. This is an outcome measure.

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

Not applicable. This is an outcome measure.

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

Not applicable. This is an outcome measure.

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

Not applicable. This is an outcome measure.

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

Not applicable. This is an outcome measure.

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

Not applicable. This is an outcome measure.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

Not applicable. This is an outcome measure.

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

Not applicable. This is an outcome measure.

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Not applicable. This is an outcome measure.

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

Not applicable. This is an outcome measure.

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Not applicable. This is an outcome measure.

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

Not applicable. This is an outcome measure.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

Not applicable. This is an outcome measure.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

Not applicable. This is an outcome measure.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: Click here to enter date range

Not applicable. This is an outcome measure.

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

Not applicable. This is an outcome measure.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Not applicable. This is an outcome measure.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

Not applicable. This is an outcome measure.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

Not applicable. This is an outcome measure.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

Not applicable. This is an outcome measure.

1a.8 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.

Not applicable. This is an outcome measure.

1a.8.1 What process was used to identify the evidence?

Not applicable. This is an outcome measure.

1a.8.2. Provide the citation and summary for each piece of evidence.

Not applicable. This is an outcome measure.

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
MCC_ACO_Admission_Measure_NQF_Evidence_Form__01-29-16_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)*

As of 2010, more than two-thirds of Medicare beneficiaries had been diagnosed with or treated for two or more chronic conditions [1]. People with MCCs are more likely to be admitted to the hospital than those without chronic conditions or with a single chronic condition. Additionally, they are more likely to visit the emergency department, use post-acute care (such as skilled nursing facilities), and require home health assistance [1]. No quality measures specifically designed for this population exist to assess quality of care or to enable the evaluation of whether current efforts to improve care are successful; this measure is designed to help fill that gap as called for in NQF's "Multiple Chronic Conditions Measurement Framework." [2]

The measure is focused on ACOs because better, coordinated care should lower the risk of hospitalization for this vulnerable population. The measure is designed to illuminate variation in hospital admission rates and incentivize ACOs to develop efficient and coordinated chronic disease management strategies that anticipate and respond to patients' needs and preferences. The measure is also consistent with ACOs' commitment to deliver patient-centered care that fulfills the goals of the Department of Health and Human Services' National Quality Strategy – improving population health, providing better care, and lowering health care costs [3].

The rationale for measuring all-cause acute admissions is to assess the quality of care as experienced by the patient and to drive overall improvements in care quality, coordination, and efficiency that are not specific to certain diseases. Ambulatory care providers can act together to lower patients' risk for a wide range of acute illness requiring admission in several ways:

1. Provide optimal and accessible chronic disease management to reduce catastrophic sequelae of chronic disease. For example:
a. Support healthy lifestyle behaviors and optimize medical management to minimize the risk for cardiovascular events such as stroke and heart attacks; and
b. Carefully monitor and act early to address chronic problems that require major interventions if allowed to progress (for example, assessment and treatment of peripheral artery disease in unresolving infections in order to prevent amputation).

2. Anticipate and manage the interactions between chronic conditions. For example:
a. Closely monitor renal function in patients on diuretic therapy for heart failure and chronic kidney disease;
b. Minimize polypharmacy to reduce drug-drug and drug-disease interactions; and
c. Assess and treat depression to improve self-efficacy and self-management of chronic disease.

3. Provide optimal primary prevention of acute illnesses, such as recommended immunizations and screening.

4. Facilitate rapid, effective ambulatory intervention when acute illness does occur, whether related or unrelated to the chronic conditions. For example:
a. Promptly prescribe antibiotics for presumed bacterial pneumonia and diuretic treatment for fluid overload in heart failure;
b. Empower patients to recognize symptoms and to seek timely care; and
c. Create accessible care options for patients (e.g., weekend or evening hours; capacity to deliver intravenous medications).

5. Partner with the government, local businesses, and community organizations to improve support for patients with chronic illness. For example:
a. Collaborate with home nursing programs;
b. Partner with local businesses to increase opportunities to engage in healthy lifestyle behaviors; and
c. Provide outreach and services at senior centers.

Finally, a number of studies have shown that improvements in the delivery of healthcare services for ambulatory patients with MCCs can lower the risk of admission [4-9]. Demonstrated strategies include improving access to care; supporting self-care in the home; better coordinating care across providers; and integrating social work, nursing, and medical services.

Citations:
1. Centers for Medicare and Medicaid Services. Chronic Conditions Among Medicare Beneficiaries, Chartbook: 2012 Edition. 2012; http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/Downloads/2012Chartbook.pdf. Accessed March 18, 2014.

2. National Quality Forum (NQF). Multiple Chronic Conditions Measurement Framework. 2012; http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=71227

3. U.S. Department of Health and Human Services. Multiple chronic conditions—A strategic framework: Optimum health and quality of life for individuals with multiple chronic conditions. December 2010; http://www.hhs.gov/ash/initiatives/mcc/mcc_framework.pdf. Accessed March 20, 2014.

4. Chan CL, You HJ, Huang HT, Ting HW. Using an integrated COC index and multilevel measurements to verify the care outcome of patients with multiple chronic conditions. BMC health services research. 2012 2012;12:405.

5. Dorr DA, Wilcox AB, Brunker CP, Burdon RE, Donnelly SM. The effect of technology-supported, multidisease care management on the mortality and hospitalization of seniors. Journal of the American Geriatrics Society. Dec 2008;56(12):2195-2202.

6. Levine S, Steinman BA, Attaway K, Jung T, Enguidanos S. Home care program for patients at high risk of hospitalization. American Journal of Managed Care. 2012 Aug 2012;18(8):e269-276.

7. Centers for Medicare & Medicaid Services (CMS). Medicare Health Support. 2012; https://www.cms.gov/Medicare/Medicare-General-Information/CCIP/. Accessed March 27, 2014.

8. Littleford A, Kralik D. Making a difference through integrated community care for older people. Journal of Nursing and Healthcare of Chronic Illness. 2010;2(3):178-186.

9. Sommers LS, Marton KI, Barbaccia JC, Randolph J. Physician, nurse, and social worker collaboration in primary care for chronically ill seniors. Arch Intern Med. Jun 26 2000;160(12):1825-1833.

10. Zhang NJ, Wan TT, Rossiter LF, Murawski MM, Patel UB. Evaluation of chronic disease management on outcomes and cost of care for Medicaid beneficiaries. Health policy (Amsterdam, Netherlands). May 2008;86(2-3):345-354.Brown RS, Peikes D, Peterson G, Schore J, Razafindrakoto CM. Six features of Medicare coordinated care demonstration programs that cut hospital admissions of high-risk patients. Health Affairs. 2012 Jun 2012;31(6):1156-1166.

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.*
We report the variation in ACO performance score using the 2012 Medicare Full Sample.

There were 4,937,344 patients in the 2012 Medicare Full Sample who met our inclusion and exclusion criteria for the measure cohort. Among these, there were 239,551 patients in 114 ACOs.

The crude U.S. national Medicare FFS rate of acute, unplanned admissions among patients with MCCs was 71.9 per 100 person-years.

Among ACOs, the mean RSAAR for calendar year 2012 was 69.3 per 100 person-years (standard deviation = 10.8). The median RSAAR was 68.5 admissions per 100 person-years (interquartile range [IQR] 62.0 to 76.0). The minimum RSAAR score was 48.0; the 5th percentile was 52.7; the 95th percentile was 86.8; and maximum score was 106.5.

There was a substantial amount of ACOs in each performance category. 45 ACOs (39.5%) performed 'no different' from the U.S. national Medicare FFS admission rate of patients with MCCs. An additional 22 ACOs (19.3%) performed 'worse than the national rate,' and 47 ACOs (41.2%) performed 'better than the national rate.'

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.
Not applicable.

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 examined disparities in ACO performance based on the proportion of patients of low socioeconomic status (SES) being cared for by each ACO.

Identification of ACOs caring for few and many 'low SES' patients
We identified low SES patients using two variables: the Agency for Healthcare Research and Quality (AHRQ) SES Index and patient Medicare and Medicaid dual-eligibility status.

Using the AHRQ SES Index (described in the Section 2b4.3 of the NQF Testing form and Appendix E of the heart failure and diabetes ACO admission measure technical report), which is a continuous variable, we created a dichotomous low-SES variable by assessing the distribution of SES scores across a broad sample of Medicare FFS beneficiaries, labeling patients with the lowest 20% of scores as "low SES" (see Testing Form, Section 1.8, for further details). We then categorized ACOs into quartiles, based on the proportion of low SES patients in their cohort (first quartile (Q1) = 'few' low SES patients, fourth quartile (Q4) = 'many' low SES patients).

Similarly, we categorized ACOs by the proportion of Medicaid dual-eligible patients in their cohort into ACOs caring for 'few' (Q1) and 'many' (Q4) Medicaid dual-eligible patients.

Results: AHRQ SES and Medicaid Dual-Eligibility Analyses
Using the AHRQ SES Index, for the 29 ACOs in Q1, the proportion of low-SES patients ranged from 0% to 4.7%; for the 29 ACOs in Q4, the proportion of low-SES patients ranged from 25.3% to 95.5%.

Among the 29 ACOs caring for few low SES patients (Q1), 2 (6.9%) performed 'worse than the national rate,' 16 (55.2%) performed 'no different than the national rate,' and 11 (37.9%) performed 'better than the national rate.' Among the 29 ACOs caring for many low-SES patients (Q4), 11 (37.9%) performed 'worse than the national rate,' 10 (34.5%) performed 'no different than the national rate,' and 8 (27.6%) performed 'better than the national rate.'

Using Medicaid dual eligibility as an indicator of low SES, among the 28 ACOs caring for few Medicaid dual-eligible patients (Q1), the proportion of Medicaid dual-eligible patients ranged from 3.3 to 10.3%; among the 28 ACOs caring for the most Medicaid dual-eligible patients (Q4) the proportion of Medicaid dual-eligible patients ranged from 22.8 to 77.6%.

Among the 28 ACOs with few Medicaid dual-eligible patients (Q1), 1 (3.6%) performed 'worse than the national rate,' 12 (42.9%) performed 'no different than the national rate,' and 15 (53.6%) performed 'better than the national rate.' Among the 28 ACOs with many Medicaid dual-eligible patients (Q4), 11 (39.3%) performed 'worse than the national rate,' 11 (39.3%) performed 'no different than the national rate,' and 6 (21.4%) performed 'better than the national rate.'

The distribution of RSAARs across ACOs caring for increasing proportions of low SES patients reveals two patterns: (1) ACOs in Q1 (few low SES patients) tend to have lower RSAARs than ACOs in Q4 (many low SES patients); (2) there is more variation in RSAARs among ACOs in Q4 as compared with ACOs in Q1-Q3. There are small differences in these patterns when analyses are performed using Medicaid dual eligibility as an indicator of SES status (see Figure 6 of the attached technical report).

Socioeconomic Status Interpretation
Among a group of 114 ACOs, there is substantial variation in performance among ACOs caring for many (Q4) and few (Q1) low SES patients. ACOs serving many low SES patients more often perform worse than the national rate compared with ACOs serving few low SES patients. This was true using either the AHRQ SES index (37.9% vs. 6.9%, respectively) or Medicaid dual-eligibility status (39.3% vs. 3.6%, respectively) as an indicator of patients' SES. However, among ACOs serving many low SES patients, using the AHRQ SES index, 8 ACOs (27.6%) performed 'better than the national rate;' using Medicaid dual-eligibility status as an indicator, 6 ACOs (21.4%) performed 'better than the national rate.'

We also found that performance scores did not change appreciably after adjusting the models for patients' SES. As demonstrated in the Testing Form, Section 2b4.11, the Spearman correlation comparing the ACO measure scores estimated with and without risk adjustment for the AHRQ SES Index was 0.992. Similarly, the Spearman correlation for the scores estimated with and without patients' Medicaid dual eligibility was 0.994. These results demonstrate that adjusting for SES at the patient level has little effect on the measure score.

We did not adjust the measure for patient-level SES. Conceptually, ACOs should and do influence a broad range of patient and community-level factors that can mitigate the risk of admission associated with low SES, and we do not want to adjust for modifiable factors. Empirically, our results indicate that SES status plays little role at the patient level.

References:
1. Wynn B. Analysis of the Joint Distribution of Disproportionate Share Hospital Payments. 2002.

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

3. Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian SV, Carson R. Choosing area based socioeconomic measures to monitor social inequalities in low birth weight and childhood lead poisoning: The Public Health Disparities Geocoding Project (US). J Epidemiol Community Health. 2003a Mar;57(3):186-99

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.**
Not applicable. Data on disparities presented above.

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, 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.**
Individuals with MCCs are a growing part of the U.S. population and account for a disproportionately high percentage of healthcare costs [1-2]. In 2008, over 20 million Medicare FFS beneficiaries (two-thirds) had two or more of 15 chronic conditions; patients with MCCs had disproportionately high admission rates and care costs [3].

Patients with MCCs are typically frailer and at higher risk for hospitalizations due to, for example, potentially life-threatening exacerbations of their conditions and complications of complex treatment regimens. Care is fragmented among multiple providers and provider settings, limiting coordination and clear provider ownership of responsibility for outcomes. Hence, measuring and incentivizing improved ambulatory care for these patients is a national priority [1-2, 4].

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
1. National Quality Forum (NQF). Multiple Chronic Conditions Measurement Framework. 2012; http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=71227

2. U.S. Department of Health and Human Services. Multiple Chronic Conditions—A Strategic Framework: Optimum Health and Quality of Life for Individuals with Multiple Chronic Conditions. December 2010; http://www.hhs.gov/ash/initiatives/mcc/mcc_framework.pdf. Accessed March 20, 2014.
Patient Protection and Affordable Care Act, 42 U.S.C., §3022 (2010).

3. Centers for Medicare & Medicaid Services. Chronic Conditions among Medicare Beneficiaries, Chart Book. Baltimore, MD. 2011; http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/Downloads/2011Chartbook.pdf

4. Patient Protection and Affordable Care Act, 42 U.S.C., §3022 (2010).

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.)*
Not applicable.

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 : Acute Myocardial Infarction, Cardiovascular : Atrial Fibrillation, Cardiovascular : Congestive Heart Failure, Cardiovascular : Ischemic Heart Disease, Coronary Artery Disease, Mental Health : Depression, Neurology : Cognitive Impairment/Dementia, Neurology : Stroke/Transient Ischemic Attack (TIA), Pulmonary/Critical Care : Asthma, Pulmonary/Critical Care : Chronic Obstructive Pulmonary Disease (COPD), Renal : Chronic Kidney Disease (CKD)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Care Coordination : Readmissions, Health and Functional Status, Health and Functional Status : Development/Wellness, Health and Functional Status : Functional Status, Overuse, Prevention, Prevention : Immunization, Prevention : Nutrition, Safety, Safety : Complications, Safety : Medication Safety

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)
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:** MCC_Measure_NQF_Data_Dictionary_01-29-16_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.
Not applicable.

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 measured for each patient is the number of acute, unplanned admissions per 100 person-years at risk for admission. Persons are considered at risk for admission if they are alive, enrolled in FFS Medicare, and not

currently admitted. (See S.6, Numerator Details, for more information.)

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.)*
The measure requires three years of data.

Outcome time window: We observe for the outcome of admission for one full calendar year.

Time period for cohort identification: The cohort is identified using claims in a one- to three-year period prior to the measurement year (varying across the eight chronic disease groups that qualify patients for the cohort).

Risk-adjustment look-back period: Risk-adjustment variables are identified using one year of claims data prior to the measurement year.

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.
Note: The numerator of the measure score is the predicted number of admissions given the Accountable Care Organization's (ACO's) case mix, sample size, and actual admission rate. We use this field to define the outcome.

Outcome Definition:
The outcome for this measure is the number of acute, unplanned admissions per 100 person-years at risk for admission. The outcome includes inpatient admissions to an acute care hospital for any cause during the measurement year, unless an admission is identified as "planned."

Identification of Planned Admissions:
The measure outcome includes only unplanned admissions. Although clinical experts agree that proper care in the ambulatory setting should reduce hospital admissions, variation in planned admissions (such as for elective surgery) does not typically reflect quality differences. We based the planned admission algorithm on the Centers for Medicare & Medicaid Services (CMS) Planned Readmission Algorithm Version 3.0, which CMS originally created to identify planned readmissions for the hospital-wide readmission measure. In brief, the algorithm identifies a short list of always planned admissions (i.e., those where the principal discharge diagnosis is major organ transplant, obstetrical delivery, or maintenance chemotherapy) as well as those admissions with a potentially planned procedure (e.g., total hip replacement or cholecystectomy) AND a non-acute principal discharge diagnosis code. Admissions that include potentially planned procedures that might represent complications of ambulatory care, such as cardiac catheterization, are not considered planned. To adapt the algorithm for this measures, we removed from the potentially planned procedure list two procedures, cardiac catheterization and amputation, because the need for these procedures might reflect progression of clinical conditions that potentially could have been managed in the ambulatory setting to avoid admissions for these procedures. For full details of the planned admission algorithm as adapted, please see Appendix C of the attached technical report. Please see Data Dictionary, sheet "S.6 ICD9-ICD10 Planned Algorithm," for the ICD-9 to ICD-10 crosswalk for the planned admission algorithm.

Outcome Attribution:
The outcome is attributed to the ACO to which the patient is assigned. Patients are assigned to ACOs according to the specific ACO program assignment algorithm. For example, for the Medicare Shared Savings Program, patients are retrospectively assigned to an ACO if they obtained the plurality of their primary care through the ACO's providers during the measurement year. Information on ACO patient assignment can be found here:

https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Downloads/Shared-Savings-Losses-Assignment-Spec-v2.pdf.

Citations:
Brown RS, Peikes D, Peterson G, Schore J, Razafindrakoto CM. Six features of Medicare coordinated care demonstration programs that cut hospital admissions of high-risk patients. Health Affairs. 2012 Jun 2012;31(6):1156-1166.

Center for Medicare and Medicaid Services. Medicare Shared Savings Program Shared Savings and Losses and Assignment Methodology Specifications. 2013; https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Downloads/Shared-Savings-Losses-Assignment-Spec-v2.pdf. Accessed July 30, 2014.

Centers for Medicare & Medicaid Services (CMS). Medicare Health Support. 2012; https://www.cms.gov/Medicare/Medicare-General-Information/CCIP/. Accessed March 27, 2014.

Horwitz LI, Grady JN, Cohen DB, Lin Z, Volpe M, Ngo CK, Masica AL, Long T, Wang J, Keenan M, Montague J, Suter LG, Ross JS, Drye EE, Krumholz HM, Bernheim SM. Development and validation of an algorithm to identify planned readmissions from claims data. J Hosp Med 2015 Oct; 10(10):670-7.

McCarthy D, Cohen A, Johnson MB. Gaining Ground: Care Management programs to reduce hospital admissions and readmissions among chronically ill and vulnerable patients. The Commonwealth Fund, New York. 2013.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Our target population is Medicare FFS patients aged 65 years and older whose combinations of chronic conditions put them at high risk of admission and whose admission rates could be lowered through better care. The National Quality Forum's (NQF's) "Multiple Chronic Conditions Measurement Framework," which defines patients with multiple chronic conditions as people "having two or more concurrent chronic conditions that…. act together to significantly increase the complexity of management, and affect functional roles and health outcomes, compromise life expectancy, or hinder self-management [1]."

Operationally, the measure cohort includes patients with diagnoses in two or more of eight chronic disease groups:

1. Acute myocardial infarction (AMI)
2. Alzheimer's disease and related disorders or senile dementia
3. Atrial fibrillation
4. Chronic kidney disease (CKD)
5. Chronic obstructive pulmonary disease (COPD) and asthma
6. Depression
7. Heart failure
8. Stroke and transient ischemic attack (TIA)

This approach captures approximately 25% of Medicare FFS beneficiaries aged 65 years and older with at least one chronic condition (about 5 million patients in 2012).

Citations:
1. National Quality Forum (NQF). Multiple Chronic Conditions Measurement Framework. 2012; http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=71227

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

Populations at Risk : Individuals with multiple chronic conditions, Senior Care

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)*
Note: The denominator of the measure score is the expected number of admissions for the ACO given its case mix; we use this box to describe the measure cohort.

The cohort is Medicare FFS patients aged 65 years and older receiving ambulatory care during the measurement period with diagnoses that fall into two or more of eight chronic disease groups:

1. AMI
2. Alzheimer's disease and related disorders or senile dementia
3. Atrial fibrillation
4. CKD
5. COPD and asthma
6. Depression
7. Heart failure
8. Stroke and TIA

The disease groups are defined using nine chronic condition categories in CMS's Chronic Condition Data Warehouse (CCW) [1]. We combined two CCW categories into a single chronic disease group – COPD and asthma.

Sheet "S.9 Denominator Details-Cohort" in the attached Data Dictionary Excel file identifies the claim algorithms and the specific International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes for each of the eight chronic disease groups. These are fully aligned with the CCW chronic condition categories. In the CCW, the chronic condition categories are defined using ICD-9-CM diagnoses codes and are assigned to patients using validated claims algorithms for Medicare beneficiaries (based on one to three years of claims data). The measure uses these CCW definitions.

To be included in the cohort, patients must also be enrolled full-time in both Medicare Parts A and B during the year prior to the measurement period.

Citations:
1. Buccaneer. CCW Chronic Conditions: Combined Medicare and Medicaid Data. 2012; https://www.ccwdata.org/cs/groups/.../chron_cond_algo_req_proc.pd. Accessed July 30, 2014.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
The measure excludes:
1. Patients without continuous enrollment in Medicare Part A for the duration of the measurement period (or until death).
Rationale: We exclude these patients to ensure full data availability for outcome assessment (Part A during the measurement year).

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)*
1. Patients without continuous enrollment in Medicare Part A for the duration of the measurement period (or until death).
Rationale: We exclude these patients to ensure full data availability for outcome assessment (Part A during the measurement year).

Lack of continuous enrollment in Medicare Part A is determined by patient enrollment status in FFS Part A using the Medicare Denominator File. The enrollment indicators must be appropriately marked during the measurement period (Part A).

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)*
Not applicable. This measure is not stratified.

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)*
We use a two-level hierarchical negative binomial model to estimate risk-standardized acute, unplanned admissions per 100 person-years at risk for admission. This approach accounts for the clustering of patients within ACOs and variation in sample size.

The model adjusts for clinical risk factors present at the start of the measurement year, age, and the chronic disease categories that qualify the patient for the measure cohort.

Our approach to risk adjustment is tailored to and appropriate for a publicly reported outcome measure, as articulated in the American Heart Association Scientific Statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" [1-2].

The risk-standardization model has 45 variables: age, each of the eight chronic disease groups, and 36 comorbidity variables. We define clinical variables primarily using CMS's Condition Categories (CCs), which are clinically meaningful groupings of more than 15,000 ICD-9-CM diagnosis codes [3]. A map showing the assignment of ICD-9-CM codes to CCs can be found in the attached Data Dictionary Excel file, sheet "S.14 CC to ICD-9." Where ICD-9-CM codes in CCs overlap with those used in the variables that define the eight chronic disease groups, we removed those ICD-9-CM codes from the CCs to eliminate the overlap. Some variables are also defined by subsets of ICD-9-CM codes within CCs. A map showing the assignment of ICD-9-CM codes to CCs can be found in the attached Data Dictionary Excel file, sheet "S.14 Risk model CC to ICD-9." In the Data Dictionary, sheet "S.15 Risk Variable Definitions" provides the detailed CC and ICD-9-CM definitions for the clinical comorbidities, and sheet "S.15 Risk model ICD9-ICD10" contains the crosswalk of ICD-9-CM to ICD-10-CM codes for the risk model variables defined with ICD-9-CM codes.

The risk-adjustment variables are:
Demographic
1. Age (continuous variable)

Eight chronic disease groups:
1. AMI
2. Alzheimer's disease and related disorders or senile dementia
3. Atrial fibrillation
4. CKD
5. COPD and asthma
6. Depression

7. Heart failure
8. Stroke and TIA

Clinical comorbidities defined using CCs or ICD-9-CM codes:
1. Dialysis status (CC 130)
2. Respiratory failure (CC 77, 78, 79)
3. Advanced liver disease (CC 25 [remove ICD-9-CM 572.4], 26, 27, 28)
4. Pneumonia (CC 111, 112, 113)
5. Septicemia/shock (CC 2)
6. Marked disability/frailty (CC 21, 67, 68, 148, 149, 177, 178)
7. Pleural effusion/pneumothorax (CC 114)
8. Hematological diseases (CC 44 [remove ICD-8 283.11], 46)
9. Advanced cancer (CC 7, 8, 9, 11)
10. Infectious and immunologic diseases (CC 1, 3, 4 [remove ICD-9-CM 160.0, 160.1, 160.2, 160.3, 160.4, 160.5, 160.6], 5, 45, 85)
11. Severe cognitive impairment (CC 48, 75, 61, 62)
12. Major organ transplant status (CC 174, 128)
13. Pulmonary heart disease (ICD-9-CM 415.0, 416.0, 416.1, 416.8, 416.9, 417.0, 417.1, 417.8, 417.9)
14. Cardiomyopathy (ICD-9-CM 425.2, 425.4, 425.5, 425.7, 425.8, 425.9, 429.0, 429.1, 425.11, 425.18)
15. Gastrointestinal disease (CC 29, 30, 31, 33, 34)
16. Bone/joint/muscle infections/necrosis (CC 37)
17. Iron deficiency anemia (CC 47)
18. Diabetes with complications (CC 16, 17, 18, 19, 119, 120)
19. Ischemic heart disease except AMI (CC 82, 83, 84, 94; ICD-9-CM 429.5, 429.6)
20. Other lung disorders (CC 109, 115)
21. Vascular or circulatory disease (CC 104, 105 [remove ICD-9-CM 440.1, 442.1], 106)
22. Other significant endocrine disorders (CC 22 [remove ICD-9-CM 271.4, 588.81])
23. Other disability and paralysis (CC 69, 100, 101, 116)
24. Substance abuse (CC 51, 52, 53)
25. Pancreatic disease (CC 32)
26. Other neurologic disorders (CC 71, 72, 73, 74, 102, 103)
27. Arrhythmia (except atrial fibrillation) (CC 92, 93 [remove ICD-9-CM 427.31])
28. Hypertension (CC 91)
29. Hip or vertebral fracture (CC 157, 158)
30. Lower-risk cardiovascular disease (CC 86, 87, 88)
31. Cerebrovascular disease (CC 98, 99)
32. Other malignancy (CC 10 [remove ICD-9-CM 189.0 and 189.9])
33. Morbid obesity (ICD-9-CM V853.5, V853.6, V853.7, V853.8, 278.01, V853.9, V854.4, V854.5, V854.3)
34. Urinary disorders (CC 133 [remove ICD-9-CM 753.21, 753.20, 753.29, 753.22, 753.23], 136 [remove ICD-9-CM 587, 588.0, 588.1, 588.9, 588.89, 753.12, 753.13, 753.15, 753.16, 753.19])
35. Hypertensive heart and renal disease or encephalopathy (CC 89)
36. Psychiatric disorders other than depression (CC 51-54, 56, 57, 59, 60)

Citations:
1. 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.

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

3. Pope, G.C., Kautter, J., Ellis, R.P., et al.: Risk Adjustment for Medicare Capitation Payments Using the CMS-HCC Model. Health Care Financing Review. 2004; 25(4):119-141.

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*)

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.)*
The risk-standardized acute admission rate (RSAAR) for each ACO is calculated as the number of "predicted" to the number of "expected" admissions per 100 person-years, multiplied by the national crude number of admissions per person-year among all Medicare FFS patients with MCCs. All eligible Medicare FFS patients with MCCs are used in the measure score calculation, and a score is generated for each ACO.

In brief, the measure uses a hierarchical (two-level) statistical model that accounts for the clustering of patients within ACOs and accommodates the widely varying sizes of different ACOs. The measure uses a negative binomial model since our outcome is a count of the number of admissions. The first level of the model adjusts for patient factors. The relationship between patient risk factors and the outcome of admission is determined based on a national sample of patients with MCCs. Stated another way, since the effects that risk factors exert on the number of admissions are estimated based on data from all ACO and non-ACO patients in the nation, the expected number of admissions for each ACO is based on the performance of a national group of providers.

The second level of the model estimates a random-intercept term that reflects the ACO's contribution to admission risk, based on its actual admission rate, the performance of other providers with similar case mix, and its sample size. The ACO-specific random intercept is used in the numerator calculation to derive ACO-specific number of "predicted" admissions per person-year.

The measure score is the ratio of predicted admissions over the expected admissions multiplied by the crude national rate. The predicted to expected ratio of admissions is analogous to an observed/expected ratio, but the numerator accounts for clustering and sample-size variation.

The expected number of admissions is calculated based on the ACO's case mix and national average intercept.

The predicted number of admissions is calculated based on the ACO's case mix and the estimated ACO-specific intercept term.

We multiply the ratio for each ACO by a constant, the crude national rate of acute, unplanned admissions per person-years at risk for hospitalization, for ease of interpretation.

To place ACOs in performance categories, for each ACO RSAAR, one can calculate a 95% interval estimate (IE), which is similar to a confidence interval, using standard bootstrapping methods (further described in the Testing Form, Section 2b5.1). Using the 95% IE, one can assign ACOs to one of three performance categories: 'better than the national rate,' 'no different than the national rate,' and 'worse than the national rate.' The ACO is 'better than the national rate' if the 95% IE is completely below the U.S. national rate among Medicare FFS patients with MCCs; 'no different than the national rate' if the 95% IE is included in the U.S. national rate among Medicare FFS patients with MCCs; and 'worse than the national rate' if the 95% IE is above the U.S. national rate among Medicare FFS patients with MCCs.

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.
This 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.
This 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.
Not applicable.

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

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.
Medicare administrative claims and enrollment data

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)*
Integrated Delivery System

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic
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.)*
Not applicable.

2a. Reliability – See attached Measure Testing Submission Form

2b. Validity – See attached Measure Testing Submission Form
MCC_ACO_Admission_Measure_NQF_Testing_Form_01-29-16_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: Risk-Standardized Acute Admission Rates for Patients with Multiple Chronic Conditions
Date of Submission: 1/29/2016
Type of Measure:

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

To develop and to test the patient-level model, we used several 2010-2012 Medicare claims datasets as outlined below:

1. Medicare dataset used to identify the multiple chronic condition (MCC) cohort and patient risk factors for admission:

We used the 2010-2011 Chronic Conditions Data Warehouse (CCW) 100% dataset which includes patients with at least one of the 27 CCW chronic conditions. We used the CCW 2010-2011 Medicare Part A and Part B files to define the cohort and CCW 2011 Medicare Part A and Part B files to identify each patient's risk factors for the outcome of acute, unplanned admissions per person-year at risk for admission. Our MCC cohort is fully encompassed within this dataset of patients with at least one CCW chronic condition.

We used the 2011-2012 Denominator File to determine Medicare fee-for-service (FFS) enrollment, demographic, and death information for beneficiaries in our cohort in order to determine inclusion/exclusion criteria for the cohort.

2. Medicare dataset to capture the outcome (acute, unplanned admissions per person-years at risk for hospitalization):

We used the 2012 Medicare Provider Analysis and Review (MedPAR) 100% FFS dataset, containing Medicare Part A claims, to identify the outcome of admissions.

We used the 2012 Denominator File to determine Medicare FFS enrollment, demographic, and death information for beneficiaries in the MCCs to determine person-years at risk for hospitalization.

3. Dataset to identify assignment of patients to Accountable Care Organizations (ACOs):

We used a file provided by a Center for Medicare & Medicaid Services (CMS) contractor to identify which Medicare fee-for-service (FFS) beneficiaries who were assigned to each of 114 Medicare Shared Savings Program ACOs in the year 2012.

4. Datasets to determine socioeconomic status:

We used the 2008-2012 American Community Survey data from the United States (US) Census Bureau to derive the Agency for Healthcare Research and Quality (AHRQ) socioeconomic status (SES) index for each zip code in the US.

5. Dataset to identify Medicaid dual-eligibility status:

We used the 2012 Denominator File to identify dual-eligible Medicare FFS beneficiaries.

The datasets used for testing vary by testing type; see Section 1.7 for details.

1.3. What are the dates of the data used in testing? Click here to enter date range

We used data from 2010-2011. The dates of the data listed above are as follows:

1. CCW 100% Medicare Parts A and B dataset: 2010-2011

2. MedPAR dataset: 2012

3. ACO assignment data: 2012

4. Denominator File: 2011-2012

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

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

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

The number of measured entities (ACOs) varies by testing type; see Section 1.7 for details.

1.6. How many and which patients 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 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.

As set forth in Section 1.2 above, we use Medicare claims and enrollment data to identify the cohort, to define the outcome, and to accumulate risk-adjustment variables. For measure development and testing, we created datasets using 2010-2012 Medicare data, using 2012 as the measurement year. The datasets, dates, number of measured entities, and number of patients used in each type of testing are as follows: **1) 2012 Medicare Full Sample** This sample includes the cohort of all Medicare FFS beneficiaries meeting our MCC definition for the 2012 measurement year. The 2012 Medicare Full Sample includes 4,937,344 patients with MCCs. Patients were mostly female (59.7%) with an average age of 79.9 years. There were 114 ACOs in the 2012 Medicare Full Sample. Among the 4,937,344 patients with MCCs, 239,551 (4.9%) were assigned to one of 114 ACOs.

-Dataset used for: testing measure exclusions (see Section 2b3), meaningful differences in performance (see Section 2b5), risk-adjustment model (Section 2b4.4b), and all ACO measure score calculations

For model development and testing, we randomly split the 2012 Medicare Full Sample into two equal subsets of patients: the 2012 Development Sample and 2012 Validation Sample (described below).

a) 2012 Development Sample
-This sample includes 2,468,672 patients with MCCs. Patients were mostly female (59.7%), with an average age of 79.9 years. There were 114 ACOs; 119,956 (4.9%) of patients in the 2012 Development Sample were assigned to ACOs.
-Dataset used for: data element reliability (see Section 2a2.3), testing risk-adjustment model (see Section 2b4)

b) 2012 Validation Sample
-This sample includes 2,468,672 patients with MCCs. Patients were mostly female (59.7%), with an average age of 79.9 years. There were 114 ACOs; 119,595 (4.8%) of patients in the 2012 Validation Sample were assigned to ACOs.
-Dataset used for: data element reliability (see Section 2a2.3), testing risk-adjustment model (see Section 2b4)

We also split the 2012 Medicare Full Sample into subsets of patients by randomly splitting each ACO's patients in half and then randomly splitting all non-ACO patients in half.

c) 2012 Reliability Sample 1
-2012 Reliability Sample 1 includes 2,468,700 patients with MCCs. Patients were mostly female (59.7%), with an average age of 79.9 years. 119,803 (4.9%) of patients were assigned to ACOs.
-Dataset used for: measure score reliability (see Sections 2a2 and 2b2)

d) 2012 Reliability Sample 2
-2012 Reliability Sample 2 includes 2,468,644 patients with MCCs. Patients were mostly female (59.7%), with an average of 79.9 years; 119,748 (4.9%) of patients were assigned to ACOs.
-Dataset used for: measure score reliability (see Sections 2a2 and 2b2)

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 used two different indicators of Medicare beneficiaries' socioeconomic status (SES): (1) the SES score of the patient's five-digit zip code, adapted from the Agency for Healthcare Research and Quality (AHRQ) SES Index, which was created for the purpose of characterizing the SES of Medicare beneficiaries and (2) the Medicaid dual-eligibility status of beneficiaries [1]. Although race was available (as black or other) in the Medicare data, we chose not to further evaluate it based on our conceptual model and input from our technical expert panel (TEP) and public comment.

The AHRQ SES Index is based on seven neighborhood variables previously shown to contribute to SES and to be associated with outcomes. They are: (1) median household income, (2) percentage of persons living below the federal poverty level, (3) percentage of persons who are aged >16 years and in the labor force but not employed, (4) median value of owner-occupied homes, (5) percentage of persons aged >25 years who completed at least a 12th-grade education, (6) percentage of persons aged >25 years who completed at least four years of college, and (7) percentage of households that average one or more persons per room. The original AHRQ SES Index was derived using data from the 2000 US Census Bureau and was calculated using US Census Block data, which corresponded to Medicare beneficiaries' nine-digit zip code. For this measure, we used data from the US Census Bureau, American Community Survey (2008-2012) and performed a principal component analysis to derive a composite SES index score for each five-digit zip code, which we then assigned to the patient based on their zip code of residence (i.e., the smallest unit by which we could identify Medicare beneficiaries' home address). The AHRQ SES Index is a continuous variable whereby lower scores indicate lower SES zip codes and higher scores indicate higher SES zip codes.

We created a dichotomous variable from the AHRQ SES index, stratifying zip code scores into 'low SES' and 'non-low SES.' Based on the distribution of the AHRQ SES index among the entire FFS Medicare population in the 5% Medicare FFS sample, we selected the lowest quintile to represent low SES. In this lowest quintile, 21.9% of beneficiaries were Medicaid dual-eligible, as compared with 13.7% in the second lowest quintile. We then categorized each patient as low or non-low SES based on the AHRQ score derived from their zip code of residence.

Additionally, we categorized ACOs based on the proportion of low SES patients in their cohort into quartiles (first quartile [Q1] indicating few low SES patients, fourth quartile [Q4] indicating many low SES patients). Similarly, we categorized ACOs by the proportion of Medicaid dual-eligible patients in their cohort into ACOs caring for 'few' (Q1) and 'many' (Q4) Medicaid dual-eligible patients. For more information on the derivation of the AHRQ SES index and the selection of a low SES thresholds for patients and ACOs, see appendix E of the heart failure and diabetes ACO admission measure technical report

We did not use race in our analyses since differences in risk of admission among groups of different race should be captured in our risk-adjustment model (which includes age and comorbidities). Any remaining differences in the risk for hospitalization among patients of different race may represent disparities in care delivery and quality of care.

Citations

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

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 in Medicare FFS patients, we aimed to utilize only those data elements from claims data that have both face validity and reliability. We avoided the use of fields that are thought to be coded inconsistently across facilities. Specifically, we used fields that are consequential for payment and which are audited. We identified such variables through empiric analyses and our understanding of the CMS auditing and billing policies. We sought to avoid variables which do not meet these standards.

In addition, CMS has in place several hospital auditing programs used to assess overall accuracy of claims-based coding, to ensure appropriate billing, and for overpayment recoupment. CMS routinely conducts data analysis to identify potential problem areas and to 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 assessed the reliability of the data elements by comparing model variable frequencies in our *2012 Development Sample* and *2012 Validation Sample*.

Measure Score Reliability
The reliability of a measurement can be defined as the degree to which repeated measurements of the same entity agree with one another. For our measures of facility performance, the measured entity is the ACO, and reliability is the extent to which repeated measurements of the same ACO give similar results [1].

To calculate measure score reliability, we randomly sampled half of the patients from each ACO and half of the patients who were not in ACOs from the 2012 Medicare Full Sample (2012 Reliability Sample 1 and Sample 2). We calculated the measure score for all the ACOs using data from ACO and non-ACO patients and repeated the calculation using the second half of patients. Thus, each ACO was measured twice, but each measurement was 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 ACO, not of the patients. As a metric of agreement, we calculated the intra-class correlation coefficient (ICC) [2], and assessed the values according to conventional standards [3]. The agreement of the two risk-standardized acute admission rates was quantified for ACOs in each sample using the ICC(2,1) by Shrout and Fleiss [2].

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

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

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

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

Table 1. Risk-adjustment variables - prevalence and rate ratios

Variable	Prevalence of risk factors (%)		Rate ratio	
	Development Sample	Validation Sample	Development Sample	Validation Sample
Demographic				
Age	79.9 (8.0)	79.9 (8.0)	1.0	1.0
Eight chronic disease groups				
Acute myocardial Infarction (AMI)	3.1	3.1	1.0	1.0
Alzheimer's and related disorders or senile dementia	37.8	37.8	1.2	1.2
Atrial fibrillation	31.5	31.5	1.1	1.1
Chronic kidney disease (CKD)	47.0	47.1	1.2	1.2
Chronic obstructive pulmonary disease (COPD) and asthma	41.5	41.6	1.3	1.3
Depression	38.1	38.1	1.1	1.1
Heart failure	57.8	57.8	1.4	1.4
Stroke and transient ischemic attack (TIA)	14.8	14.0	1.0	1.0
Clinical comorbidities Defined using Condition Categories (CCs) or International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes				
Dialysis status (CC 130)	2.2	2.2	1.8	1.8
Respiratory failure (CC 77, 78, 79)	17.0	17.0	1.2	1.2
Advanced liver disease (CC 25 [remove ICD-9-CM 572.4], 26, 27,	2.1	2.1	1.2	1.2

Variable	Prevalence of risk factors (%)		Rate ratio	
	Development Sample	Validation Sample	Development Sample	Validation Sample
28)				
Pneumonia (CC 111, 112, 113)	23.6	23.7	1.2	1.2
Septicemia/shock (CC 2)	6.2	6.2	1.1	1.1
Marked disability/frailty (CC 21, 67, 68, 148, 149, 177, 178)	16.3	16.3	1.3	1.3
Pleural effusion/pneumothorax (CC 114)	12.7	12.7	1.1	1.1
Hematological diseases (CC 44 [remove ICD-9-CM 283.11], 46)	14.3	14.4	1.1	1.1
Advanced cancer (CC 7, 8, 9, 11)	7.8	7.8	1.4	1.4
Infectious and Immunologic diseases (CC 1, 3, 4 [remove ICD-9-CM 160.0, 160.1, 160.2, 160.3, 160.4, 160.5, 160.6], 5, 45, 85)	5.7	5.7	1.1	1.1
Severe cognitive impairment (CC 48, 75, 61, 62)	10.5	10.5	1.1	1.1
Major organ transplant status (CC 174, 128)	0.6	0.6	1.3	1.3
Pulmonary heart disease (ICD-9-CM 415.0, 416.0, 416.1, 416.8, 416.9, 417.0, 417.1, 417.8, 417.9)	9.5	9.5	1.2	1.2
Cardiomyopathy (ICD-9-CM 425.2, 425.4, 425.5, 425.7, 425.8, 425.9, 429.0, 429.1, 425.11, 425.18)	11.7	11.6	1.1	1.1
Gastrointestinal disease (CC 29-31, 33, 34)	22.6	22.6	1.1	1.1
Bone/joint/muscle infections/necrosis (CC 37)	2.2	2.2	1.1	1.1
Iron deficiency anemia (CC 47)	52.1	52.1	1.1	1.1
Diabetes with complications (CC 16-19, 119, 120)	46.2	46.1	1.2	1.1
Ischemic heart disease except AMI	60.0	60.0	1.1	1.1

Variable	Prevalence of risk factors (%)		Rate ratio	
	Development Sample	Validation Sample	Development Sample	Validation Sample
(CC 82,83, 84, 94; ICD-9-CM 429.5, 429.6)				
Other lung disorders (CC 109, 115)	38.7	38.8	1.1	1.1
Vascular or circulatory disease (CC 104, 105 [remove ICD-9-CM codes 440.1, 442.1], 106)	54.3	54.3	1.1	1.1
Other significant endocrine disorders (CC 22 [remove ICD-9-CM codes 271.4, 588.81])	8.0	8.0	1.0	1.0
Other disability and paralysis (CC 69, 100, 101, 116)	7.8	7.8	1.1	1.1
Substance abuse (CC 51-53)	11.0	11.1	1.2	1.2
Pancreatic disease (CC 32)	2.9	2.9	1.1	1.1
Other neurologic disorders (CC 71-74, 102, 103)	27.6	27.5	1.1	1.1
Arrhythmia (except atrial fibrillation) (CC 92, 93 [remove ICD-9-CM 427.31])	36.6	36.6	1.1	1.1
Hypertension (CC 91)	89.5	89.5	1.0	1.0
Hip or vertebral fracture (CC 157, 158)	7.8	7.8	1.1	1.1
Lower-risk cardiovascular disease (CC 86-88)	31.0	31.0	1.0	1.0
Cerebrovascular disease (CC 98, 99)	9.9	9.9	1.1	1.1
Other malignancy (CC 10 [remove ICD-9-CM codes 189.0 and 189.9)]	13.7	13.7	1.0	1.0
Morbid obesity (ICD-9-CM V853.5, V853.6, V853.7, V853.8, 278.01, V853.9, V854.4, V854.5, V854.3)	4.8	4.8	1.1	1.1
Urinary disorders (CC 133 [remove ICD-9-CM codes 753.21, 753.20, 753.29, 753.22, 753.23], 136	34.2	34.2	1.1	1.1

Variable	Prevalence of risk factors (%)		Rate ratio	
	Development Sample	Validation Sample	Development Sample	Validation Sample
[remove ICD-9-CM 587, 588.0, 588.1, 588.9, 588.89, 753.12, 753.13, 753.15, 753.16, 753.19])				
Hypertensive heart and renal disease or encephalopathy (CC 89)	18.5	18.6	1.1	1.1
Psychiatric disorders other than depression (CC 54, 56, 57, 59, 60)	28.2	28.2	1.1	1.1

Measure Score Reliability:
The ICC between the two risk-standardized acute admission rates (RSAARs) was 0.84, which according to the conventional interpretation is excellent [1].

Citations
1. 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?*)

Data Element Reliability Results
Compared with the 2012 Development Sample, the mean age of patients and the frequency of risk-adjustment variables were similar in the 2012 Validation Sample. This suggests that the data elements are reliable across these samples.

Measure Score Reliability Results
The ICC demonstrates excellent agreement across samples, indicating 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)*

We demonstrated measure validity through: reliance on relevant prior validity testing conducted for other claims-based measures, use of established measure development guidelines, and assessment by external groups and a TEP.

Validity of Claims-Based Measures

Our team has demonstrated the validity of using claims data for risk adjustment in lieu of medical record data in estimating facility-level measure scores for a number of hospital outcome measures endorsed by the National Quality Forum (NQF). CMS has validated six NQF-endorsed measures currently in public reporting (acute myocardial infarction [AMI], heart failure, and pneumonia mortality and readmission) with models that used medical record-abstracted data for risk adjustment. Specifically, we conducted claims model validation by building comparable models using abstracted medical record 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]. These findings support this measure's validity; however, we acknowledge that the use of claims data for risk adjustment has been validated for hospital outcomes measure and not for outcome measures among ambulatory patients.

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 [5], 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 through: holding regular discussions with our in-house and clinically diverse working group, consultations with leading outside experts on care for patients with MCCs, consulting our national TEP, and holding a 30-day public comment period.

Yale New Haven Health Services Corporation—Center for Outcomes Research and Evaluation clinicians and statistical experts comprised the working group. The working group members have expertise in quality measurement, clinical management of patients with multiple chronic conditions, statistical modeling, healthcare disparities, and healthcare policy. Through regular in-person meetings and teleconferences, the working group discussed all aspects of measure development, including the cohort and outcome definitions and risk adjustment.

In addition to the working group and in alignment with the CMS MMS, we convened a TEP to provide input and feedback during measure development from a group of recognized experts in relevant fields. The TEP advised on this measures and two related ACO measures under development for ambulatory patients with diabetes and heart failure. To convene the TEP, we released a public call for nominations and selected individuals to represent a range of perspectives including clinicians, patients, and individuals with experience in quality improvement, performance measurement, and healthcare disparities. We held four structured TEP conference calls consisting of presentation of key issues, our proposed approach, and relevant data, followed by open discussion among TEP members.

We also reflected input of two leading experts in designing the cohort: Dr. Cynthia Boyd, Associate Professor of Medicine at Johns Hopkins Bloomberg School of Public Health; and Dr. Mary Tinetti, Professor of Medicine and Endocrinology at Yale University School of Medicine.

Finally, we held a public comment period and received supportive comments from several national stakeholder groups and several individuals generally supporting the cohort definition and the use of admission as an outcome.

List of TEP Members
1. Lawrence M. Becker, BS, Xerox Corporation (Director, Strategic Partnerships, Alliances and Analytics); Rochester, NY
2. Alex Blum, MD, MPH, Evergreen Health Cooperative (Chief Medical Officer); Baltimore, MD
3. Sanjay Doddamani, MD, Geisinger Health System (System-wide Chief of Advanced Cardiac Disease – Heart Failure); Danville, PA
4. Kevin Fiscella, MD, MPH, University of Rochester Medical Center (Professor of Family Medicine); Rochester, NY
5. Elbert Huang, MD, MPH, University of Chicago (Associate Professor of Medicine, Director of the Center for Translational and Policy Research of Chronic Diseases, and Associate Director of the Chicago Center for Diabetes Translation Research); Chicago, IL
6. Bruce Leff, MD, Johns Hopkins University School of Medicine (Professor of Medicine, Division of Geriatric Medicine); The Johns Hopkins University Bloomberg School of Public Health (Faculty, Health Services Research Development Center and Lipitz Center for Integrated Health Care); Baltimore, MD
7. Andy Miller, MD, MPH, Healthcare Quality Strategies, Inc. (Medical Director); East Brunswick, NJ; Colorado Foundation for Medical Care (CMO, Integrating Care for Populations & Communities National Coordinating Center); Englewood, CO
8. Ami Parekh, MD, JD, University of California, San Francisco (Medical Director for Health System Innovation); San Francisco, CA
9. Christine Ritchie, MD, University of California, San Francisco (Professor of Medicine, Division of Geriatrics); San Francisco, CA
10. Two patient representatives.

<u>Process Used to Identify International Classification of Diseases, Tenth Revision (ICD-10) Codes</u>
This application includes ICD-10 codes that correspond to all International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes included in the specifications. The goal was to convert this measure into a new code set, fully consistent with the intent of the original measure.

• ICD-10 diagnosis codes used to the cohort were identified using the 2013 ICD-9-CM to ICD-10-CM General Equivalence Mapping (GEM) files made available by CMS. We then internally performed clinician review of this crosswalk.

• ICD-10 diagnosis codes used to define the pacemaker/CRT/ICD risk variable defined with ICD-9-CM codes were identified using the 2013 ICD-9-CM to ICD-10-CM General Equivalence Mapping (GEM) files made available by CMS. We then internally performed clinician review of this crosswalk.

• ICD-10 diagnosis and procedure codes used to define the Planned Admission Algorithm were identified from the 2014 version of the AHRQ Clinical Classification Software (CCS) categories specified for ICD-10, followed by clinician review. The algorithm also includes some individual ICD-9-CM codes. To create the crosswalk for the ICD-9-level codes, we used the 2013 ICD-9-CM to ICD-10-CM GEM files made available by CMS, followed by clinician review.

Citations

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.

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*)

We did not assess the validity of the measures quantitatively. As noted above, however, the TEP and experts we consulted during measure development were supportive of the MCC cohort definition and use of admission as an outcome. During the public comment period, we also received comments on the measure cohort, outcome, and risk model. The feedback on the measure focus and the measure's use for ACO quality reporting, overall, was positive.

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?*)

There was strong support expressed by the members of the TEP and in public comment for the validity of the measure. There were no strong concerns about the measure. See public comment

document for further details: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/CallforPublicComment.html.

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*)

We determined the exclusions to be appropriate based on clinical and methodological considerations. For this measure, we only exclude patients if they do not have continuous enrollment in Medicare Part A during the measurement year because we cannot assess the outcome for these patients. There are no clinical exclusions.

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*)

Table 2 provides the number of patients excluded from the MCC cohort. Out of the total number of Medicare FFS patients with MCCs (n=5,070,533), we excluded 133,189 (2.6%) due to non-continuous enrollment in Medicare Part A in 2012 because we were not able to adequately capture the outcome for these patients. Among these excluded patients, 133,188 (99.99%) were non-ACO patients and 1 was an ACO patient.

Since the number of excluded patients assigned to ACOs was very low, we did not perform a frequency distribution analysis across ACOs.

The final cohort included 4,937,344 patients.

Table 2. Patients excluded from sample for each exclusion criterion

Exclusion	Number excluded from Medicare FFS MCC cohort	Number of patients excluded from ACOs
Non-continuous enrollment in Part A in 2012	133,188	1

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*)

We needed to exclude patients without continuous enrollment because we could not capture the outcome for these patients. We excluded very few patients based on this criterion. As a result, the measure captures the majority of Medicare FFS patients 65 years and older with MCCs (97.4%).

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** Click here to enter number of factors **risk factors**

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

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

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

Not applicable. This measure is risk adjusted.

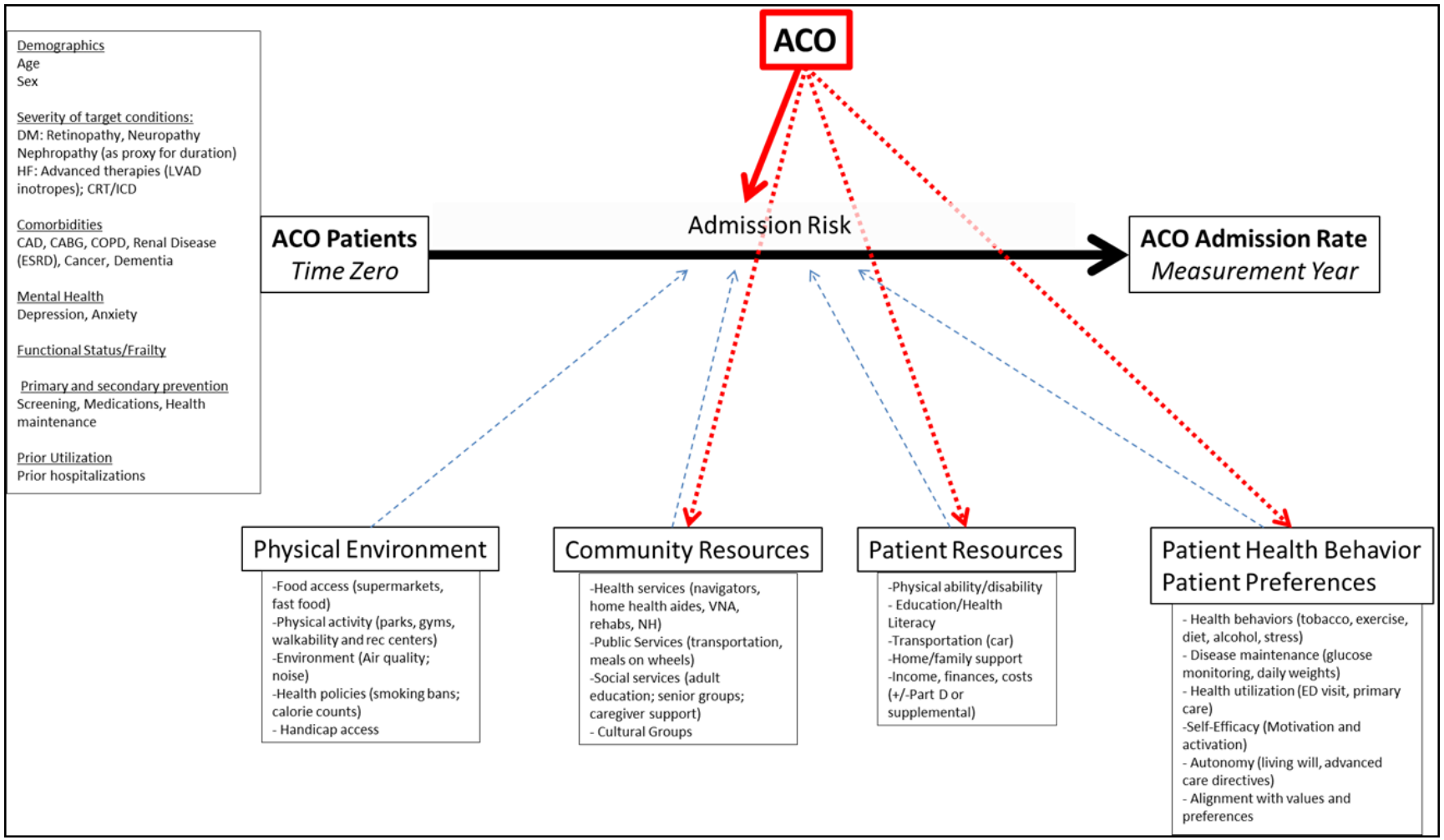
2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression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

We selected the risk-adjustment model variables based on the existing literature, clinical judgment, empirical analyses, and input from our TEP and other experts. We considered factors that may impact the rate of admission, including patient-level factors (e.g., demographics, SES, clinical risk factors on admission); we also considered the impact of other non-clinical factors such as health behaviors and community resources. In this work, we were guided by a conceptual framework that was informed by a literature review and environmental scan, outlining the relationships between potential clinical and contextual factors and rates of admissions among chronic disease populations cared for by ACOs. Importantly, many factors other than traditional medical care delivered in the office or hospital settings will impact health outcomes for patients with chronic disease. For example, ACOs practicing in communities where patients have limited access to transportation, healthy foods, and recreational facilities, may have less success in promoting healthy behaviors among patients with MCCs; this may in turn impact quality outcomes. Recognition of and attention to the health environment may be important for achieving the goals of better care, better health and lower costs and thus, shared savings. The conceptual model (Figure 1) was presented and endorsed by the TEP engaged during the development of this measure. The model recognizes patient-level demographic and clinical factors, along with 4 contextual domains that may influence ACO performance: (1) Physical

environment (e.g., green spaces; safe streets); (2) Community resources (e.g., home health; senior services); (3) Patient resources (e.g., social support; transportation; income); and (4) Patient behavior/personal preferences (e.g., exercise; diet; advanced care directives; preference for intervention).

The model also recognizes the capacity of ACOs to mitigate the effects of many contextual factors on rates of admissions, encompassing both SES and non-SES variables, and supporting our decision not to adjust for contextual factors. Adjusting for contextual factors would obscure important differences in ACO quality and could serve as a disincentive for ACOs to engage with such factors. We did, however, conduct analyses of SES factors to further inform the committee's deliberation.

Figure 1. Conceptual model of factors affecting risk of hospital admission



We describe our approach to risk adjustment for the demographic factors, clinical risk factors, and contextual domains, in turn, below:

1. Demographic factors
We used clinical and conceptual criteria to adjust this measure for age but not sex or race. Age is a clinically recognized risk factor for acute admissions. In contrast, sex or race differences in risk of admission should be captured in our risk-adjustment model (which includes age and comorbidities). Any remaining differences in the risk for hospitalization among patients of different sex or race may represent disparities in care delivery and quality of care. [1,2] We did examine the effects of including sex in the models, since the relationship between sex and acute, unplanned admissions has not been tested in this setting, finding that sex was not significant after adjusting for age and clinical comorbidities.

2. Clinical risk factors
We used clinical, conceptual, and statistical criteria to select clinical risk factors for adjustment. This measure adjusts for clinical risk factors that are present at the start of the measurement period, but not for conditions that arise during the measurement period.

Development of Candidate Clinical Variables
To select candidate variables for risk adjustment, we used Part A and Part B data from one year prior to the measurement year for 100% of the Medicare FFS patients included in the cohort (2012 Medicare Full Sample). We reviewed 189 diagnostic groups included in the Hierarchical Condition Category (HCC) clinical classification system. We defined comorbidities using 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-CM codes to CCs will be prepared for the final draft. To select candidate variables, two clinicians reviewed all 189 CCs and excluded those that were not relevant to the Medicare FFS population (e.g., attention deficit disorder and female infertility) or that documented the use of specific medical equipment or treatment rather than diagnoses, leaving 159 CCs as possible candidate variables.

Using the 159 clinically relevant CCs, our next step was to combine CCs into fewer candidate variables to facilitate building a parsimonious model and to minimize the effect of variation in coding practices on the clinical comorbidities the measure assigns to patients. To inform this step we derived the measure cohort in the 2012 Medicare Full Sample and examined the prevalence of each CC in the year preceding the measurement period (year 2011), the number of hospital admissions per patient-year during the measurement period (year 2012) among patients with and without the CC, and the rate ratio for the number of hospital admissions associated with each CC (the rate of admissions among patients with the disease compared to the rate of admission for patients without the disease). We collapsed clinically similar CCs with similar rate ratios into single clinical variables and eliminated clinically less relevant variables or those with a relatively low rate ratio (<1.3). In addition, to ensure the candidate variables were mutually exclusive, we eliminated diagnosis codes within CCs that overlapped with the ICD-9-CM codes in the CCW-based chronic disease groups. At least two clinical investigators reviewed each of the decisions and reached consensus, and all candidate variables were reviewed by our technical working group. This process resulted in 46 candidate variables – age, sex, 36 CC-based variables, and the eight disease group variables. We recognize that the measure cohort is potentially highly diverse,

since many combinations of conditions can qualify a patient for inclusion in the measures. Therefore, to achieve adequate risk adjustment, we included the eight chronic disease categories that define the cohort as candidate variables.

Candidate model variables are:
Demographic
1. Age (continuous variable)
2. Sex

Eight chronic disease groups
1. AMI
2. Alzheimer's disease and related disorders or senile dementia
3. Atrial fibrillation
4. CKD
5. COPD and asthma
6. Depression
7. Heart failure
8. Stroke and TIA

Clinical comorbidities defined using Condition Categories (CCs) or International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes
1. Dialysis status (CC 130)
2. Respiratory failure (CC 77, 78, 79)
3. Advanced liver disease (CC 25 [remove ICD-9-CM 572.4], 26, 27, 28)
4. Pneumonia (CC 111, 112, 113)
5. Septicemia/shock (CC 2)
6. Marked disability/frailty (CC 21, 67, 68, 148, 149, 177, 178)
7. Pleural effusion/pneumothorax (CC 114)
8. Hematological diseases (CC 44 [remove ICD-8 283.11], 46)
9. Advanced cancer (CC 7, 8, 9, 11)
10. Infectious and Immunologic diseases (CC 1, 3, 4 [remove ICD-9-CM 160.0, 160.1, 160.2, 160.3, 160.4, 160.5, 160.6], 5, 45, 85)
11. Severe cognitive impairment (CC 48, 75, 61, 62)
12. Major organ transplant status (CC 174, 128)
13. Pulmonary heart disease (ICD-9-CM 415.0, 416.0, 416.1, 416.8, 416.9, 417.0, 417.1, 417.8, 417.9)
14. Cardiomyopathy (ICD-9-CM 425.2, 425.4, 425.5, 425.7, 425.8, 425.9, 429.0, 429.1, 425.11, 425.18)
15. Gastrointestinal disease (CC 29, 30, 31, 33, 34)
16. Bone/joint/muscle infections/necrosis (CC 37)
17. Iron deficiency anemia (CC 47)
18. Diabetes with complications (CC 16, 17, 18, 19, 119, 120)
19. Ischemic heart disease except AMI (CC 82, 83, 84, 94; ICD-9-CM 429.5, 429.6)
20. Other lung disorders (CC 109, 115)
21. Vascular or circulatory disease (CC 104, 105 [remove ICD-9-CM 440.1, 442.1], 106)
22. Other significant endocrine disorders (CC 22 [remove ICD-9-CM 271.4, 588.81])
23. Other disability and paralysis (69, 100, 101, 116)

24. Substance abuse (CC 51, 52, 53)
25. Pancreatic disease (CC 32)
26. Other neurologic disorders (CC 71, 72, 73, 74, 102, 103)
27. Arrhythmia (except atrial fibrillation) (CC 92, 93 [remove ICD-9-CM 427.31])
28. Hypertension (CC 91)
29. Hip or vertebral fracture (CC 157, 158)
30. Lower-risk cardiovascular disease (CC 86, 87, 88)
31. Cerebrovascular disease (CC 98, 99)
32. Other malignancy (CC 10 [remove ICD-9-CM 189.0 and 189.9)]
33. Morbid obesity (ICD-9-CM V853.5, V853.6, V853.7, V853.8, 278.01, V853.9, V854.4, V854.5, V854.3)
34. Urinary disorders (CC 133 [remove ICD-9-CM 753.21, 753.20, 753.29, 753.22, 753.23], 136 [remove ICD-9-CM 587, 588.0, 588.1, 588.9, 588.89, 753.12, 753.13, 753.15, 753.16, 753.19])
35. Hypertensive heart and renal disease or encephalopathy (CC 89)
36. Psychiatric disorders other than depression (CC 54, 56, 57, 59, 60)

Final Variable Selection

In order to select the final set of variables, we used the 2012 Development Sample to rank the candidate variables in terms of their importance for the model by comparing the Akaike Information Criterion (AIC) values. The AIC is used to select the best-fitting model using the least number of variables; it is commonly used in variable selection for negative binomial models (which use count data, such as a count of the number admissions) to account for overdispersion (whereby data vary more than expected) [7]. We selected variables starting with the 46 candidate variables. We removed one variable and determined the best combination of 45 variables that resulted in the smallest AIC compared with other combinations of 45 variables. Based on the best 45 variables, we removed one more variable and determined the best 44 variables. We repeated these steps until we reached one variable. Each of the final 46 models, containing combinations of 1 to 46 variables, represents the best model (combination of variables) given the different numbers of variables. We calculated the AIC for these 46 models. Based on the smallest AIC, we selected preliminary final combination of variables.

Given the diversity of patients in the MCC cohort, we investigated further whether these preliminary final variables predicted risk similarly across subgroups of the MCC cohort likely to experience wide variations in admission risk. We stratified the cohort into three age groups: age <75; age 75 to <85; and age >85. Separately, we stratified the cohort into three groups by the number of chronic disease groups each patient was flagged with: patients with exactly two of the eight chronic disease groups; those with three to four; and those with five or more. We compared the direction and magnitude of the rate ratios for each variable for the full group and each of the subgroups to evaluate whether the variables behaved similarly. We evaluated interaction terms between the variables with rate ratios that differed across subgroups by 0.2 or more with (1) the three age categories and/or (2) the three number of disease group categories as relevant. To limit model complexity, we retained interaction terms only if they demonstrated a meaningful improvement in model performance as measured by the deviance R-squared [see Section 2b4.5 for an explanation of this statistic].

3. Socioeconomic status

Based on a conceptual model that was informed by a literature review and environmental scan, we did not adjust for contextual factors which may impact acute admissions, including variables related to SES. ACOs should and do influence a broad range of patient-level and community-level factors that can mitigate the risk of admission associated with the contextual environment.

However, to inform the committee's consideration of the decision not to adjust for SES, we performed focused analyses using SES variables. These analyses are informative for future measure use, but the decision not to adjust for SES in this measure was not based on the results of these statistical analyses.

To assess the potential effect of SES on ACO performance, we first included SES as a patient-level covariate in the models. As there are no standardized methods for assessing a Medicare beneficiary's SES, we used two different indicators of SES: (1) the SES score of the patient's 5-digit zip code, adapted from the AHRQ SES Index [3], which was developed for the purpose of characterizing the SES of Medicare beneficiaries and (2) the Medicaid dual-eligibility status of beneficiaries. We created a dichotomous variable from the AHRQ SES score, defining patients as low SES if they had an AHRQ Score of 0 to 45 and non-low SES if they had an AHRQ score of >45. This cut-point represented the lowest quintile of AHRQ SES scores among the 5% Medicare FFS Sample. In this lowest quintile, 21.9% of patients were Medicaid dual eligible. For further details on how we calculated the AHRQ SES score and developed a dichotomous variable we refer to the attached technical report, Appendix E. Additionally, we performed ACO-level analyses based on the proportion of low SES patients being cared for by an ACO. These methods and results are reported in the NQF Submission form.

4. Contextual Domains
The four contextual domains, which include SES factors, may influence the clinical health status of patients as well as the outcome of acute admissions, impacting ACOs' ability to prevent acute admissions. However, when evaluating provider quality, we do not want to adjust for them, since these affects may be mediated by ACOs, and the measure score should ideally reflect successful efforts to mitigate their impact on admission rates. This approach is consistent with the ACO program design – as part of their mission, ACOs are encouraged to develop strategic partnerships with community-based organizations and businesses in order to improve population health and reduce the risk of admission. It is also supported by growing evidence that integrated health systems can identify and mitigate the degree to which non-health factors impact health outcomes (e.g., by connecting patients with available health-related services).[4]

<u>Citations</u>
1. Glynn LG, Valderas JM, Healy P, et al. The prevalence of multimorbidity in primary care and its effect on health care utilization and cost. Family practice. 2011 Oct 2011;28(5):516-523.

2. Longman JM, M IR, Passey MD, et al. Frequent hospital admission of older people with chronic disease: a cross-sectional survey with telephone follow-up and data linkage. BMC Health Services Research. 2012 2012;12:373.

3. Payne RA, Abel GA, Guthrie B, Mercer SW. The effect of physical multimorbidity, mental health conditions and socioeconomic deprivation on unplanned admissions to hospital: a

retrospective cohort study. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne. 2013 Mar 19 2013;185(5):E221-228.

4. Wolff JL, Starfield B, Anderson G. Prevalence, expenditures, and complications of multiple chronic conditions in the elderly. Archives of Internal Medicine. 2002 Nov 11 2002;162(20):2269-2276.

5. Mazerolle MJ. Making sense out of Akaike's Information Criterion (AIC): its use and interpretation in model selection and inference from ecological data. Accessed 3/14/2014, 2014.

6. Alley DE, Asomugha CN, Conway PH, Sanghavi DM. Addressing Social Needs through Medicare and Medicaid, N Engl J Med 2016; 374:8-11.

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

Based on the smallest AIC among the combination of candidate variables, we retained 45 variables of the 46 candidate variables. The only variable that was not included in the final model was sex, which was not statistically significant in the model.

When we fit the 45-variable model in each of the three age subgroups and separately in the three subgroups stratified by the three categories for the number of chronic disease groups, five variables showed rate ratios that varied across subgroups by more than 0.2 across either the age or number of disease group categories: (1) Alzheimer's/dementia (age); (2) hip or vertebral fracture (age); (3) advanced cancer (age); (4) dialysis status (number of conditions); and (5) liver disease (number of conditions). The interaction terms were all statistically significant ($p < 0.0001$). However, adding the interaction terms to the model only slightly improved the deviance R-squared (it increased from 0.1230 to 0.1246), and it greatly increased model complexity. Therefore, we opted not to add the interaction terms to the model.

The following variables were selected as the final risk-adjustment variables:
Demographic
1. Age (continuous variable)

Eight chronic disease groups
1. AMI
2. Alzheimer's disease and related disorders or senile dementia
3. Atrial fibrillation
4. CKD
5. COPD and asthma
6. Depression
7. Heart failure
8. Stroke and TIA

Clinical comorbidities defined using Condition Categories (CCs) or International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes
1. Dialysis status (CC 130)
2. Respiratory failure (CC 77, 78, 79)
3. Advanced liver disease (CC 25 [remove ICD-9-CM 572.4], 26, 27, 28)
4. Pneumonia (CC 111, 112, 113)
5. Septicemia/shock (CC 2)
6. Marked disability/frailty (CC 21, 67, 68, 148, 149, 177, 178)
7. Pleural effusion/pneumothorax (CC 114)
8. Hematological diseases (CC 44 [remove ICD-8 283.11], 46)
9. Advanced cancer (CC 7, 8, 9, 11)
10. Infectious and Immunologic diseases (CC 1, 3, 4 [remove ICD-9-CM 160.0, 160.1, 160.2, 160.3, 160.4, 160.5, 160.6], 5, 45, 85)
11. Severe cognitive impairment (CC 48, 75, 61, 62)

12. Major organ transplant status (CC 174, 128)
13. Pulmonary heart disease (ICD-9-CM 415.0, 416.0, 416.1, 416.8, 416.9, 417.0, 417.1, 417.8, 417.9)
14. Cardiomyopathy (ICD-9-CM 425.2, 425.4, 425.5, 425.7, 425.8, 425.9, 429.0, 429.1, 425.11, 425.18)
15. Gastrointestinal disease (CC 29, 30, 31, 33, 34)
16. Bone/joint/muscle infections/necrosis (CC 37)
17. Iron deficiency anemia (CC 47)
18. Diabetes with complications (CC 16, 17, 18, 19, 119, 120)
19. Ischemic heart disease except AMI (CC 82, 83, 84, 94; ICD-9-CM 429.5, 429.6)
20. Other lung disorders (CC 109, 115)
21. Vascular or circulatory disease (CC 104, 105 [remove ICD-9-CM 440.1, 442.1], 106)
22. Other significant endocrine disorders (CC 22 [remove ICD-9-CM 271.4, 588.81])
23. Other disability and paralysis
24. Substance abuse (CC 51, 52, 53)
25. Pancreatic disease (CC 32)
26. Other neurologic disorders (CC 71, 72, 73, 74, 102, 103)
27. Arrhythmia (except atrial fibrillation) (CC 92, 93 [remove ICD-9-CM 427.31])
28. Hypertension (CC 91)
29. Hip or vertebral fracture (CC 157, 158)
30. Lower-risk cardiovascular disease (CC 86-88)
31. Cerebrovascular disease (CC 98, 99)
32. Other malignancy (CC 10 [remove ICD-9-CM 189.0 and 189.9])
33. Morbid obesity (ICD-9-CM V853.5, V853.6, V853.7, V853.8, 278.01, V853.9, V854.4, V854.5, V854.3)
34. Urinary disorders (CC 133 [remove ICD-9-CM 753.21, 753.20, 753.29, 753.22, 753.23], 136 [remove ICD-9-CM 587, 588.0, 588.1, 588.9, 588.89, 753.12, 753.13, 753.15, 753.16, 753.19])
35. Hypertensive heart and renal disease or encephalopathy (CC 89)
36. Psychiatric disorders other than depression (CC 54, 56, 57, 59, 60)

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)

We performed multiple analyses to assess the effect of sex and SES on model performance. These analyses are informative for future measure use, but the decision not to adjust for sex or SES in this measure was based on conceptual/clinical factors and not on the results of these statistical analyses (see 2b4.3.).

To assess the effect of sex and SES on model performance, we compared deviance R-squared values with and without the variables for sex and SES included as patient-level variables in the model. We compared the correlation between measure scores with and without sex and SES included in the models, using the Spearman correlation.

For the SES analyses, we also assessed ACO performance among groups of ACOs caring for similar proportions of low SES patients. To do this, we categorized ACOs into quartiles (Q1 indicating ACOs with few low SES patients, Q4 indicating ACOs with many low SES patients). We used boxplots to compare the distribution of RSAARs across ACOs by low SES quartiles.

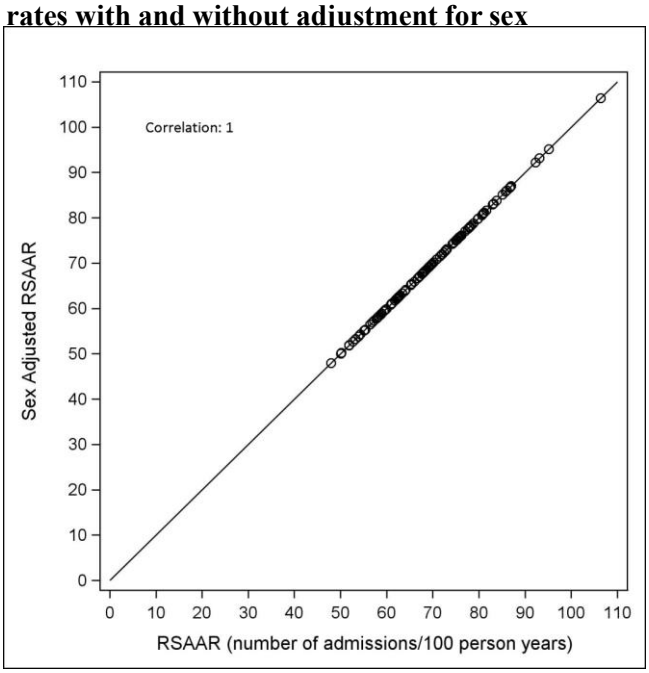
The SES analyses were performed using both the AHRQ SES index (i.e., low SES, binary variable described above) and Medicaid dual-eligibility status as a proxy for patients' SES status. Statistical analyses were performed using SAS version 9.3 (SAS Institute Inc., Cary, NC).
The results of the patient-level analyses indicate that adjustment for sex and for low-SES status as a patient variable in the models did not affect measure performance.

Specifically, related to SES, performance scores did not change appreciably after adjusting the models for patients' SES. As demonstrated in the Testing Form, Section 2b4.11, the Spearman correlation comparing the ACO measure scores estimated with and without risk adjustment for the AHRQ SES Index was 0.992. Similarly, the Spearman correlation for the scores estimated with and without patients' Medicaid dual eligibility was 0.994. These results demonstrate that adjusting for SES at the patient level has little effect on the measure score.

Sex

The deviance R-squared values for the two models, one adjusted for the 45 clinical variables *and* sex, and one adjusted for the 45 clinical variables *without* sex, were 0.123 and 0.123, respectively, meaning adjustment for sex explained the same amount of variation and did not result in incremental benefit. Comparing the RSAAR with and without sex included in the model resulted in a high degree of correlation (Spearman correlation = 1), meaning ACOs performed the same with and without risk adjustment for sex. (Figure 2)

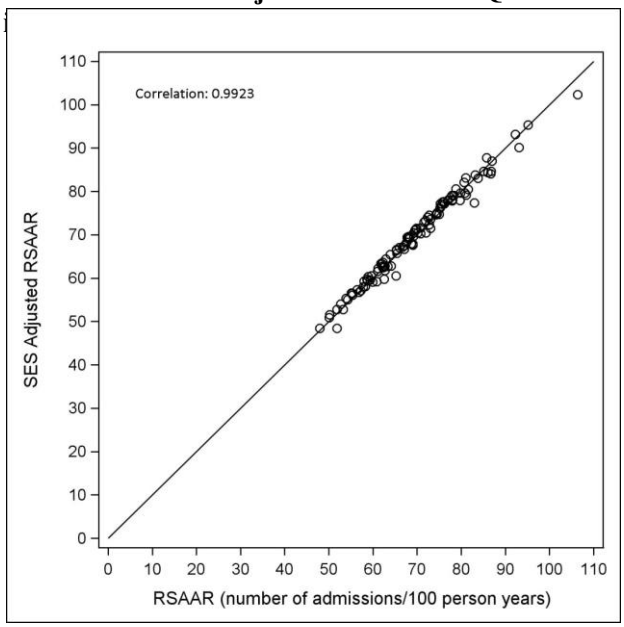
Figure 2. Plot of acute, unplanned admission rates with and without adjustment for sex



AHRQ SES Index

The deviance R-squared values for the two models – one adjusted for the 45 clinical variables *and* low SES, and one adjusted for the 45 clinical variables *without* adjusting for low SES – were 0.124 and 0.123, respectively, meaning adjustment for low SES explained the same variation and did not provide incremental benefit. Comparing the RSAAR with and without low SES included in the model resulted in a high degree of correlation (Spearman correlation = 0.992). The graph demonstrates that, compared with *not* adjusting for low SES, adjusting for low SES results in some ACOs having slightly lower RSAAR scores (below the line) and other ACOs having higher RSAAR scores (above the line). (Figure 3)

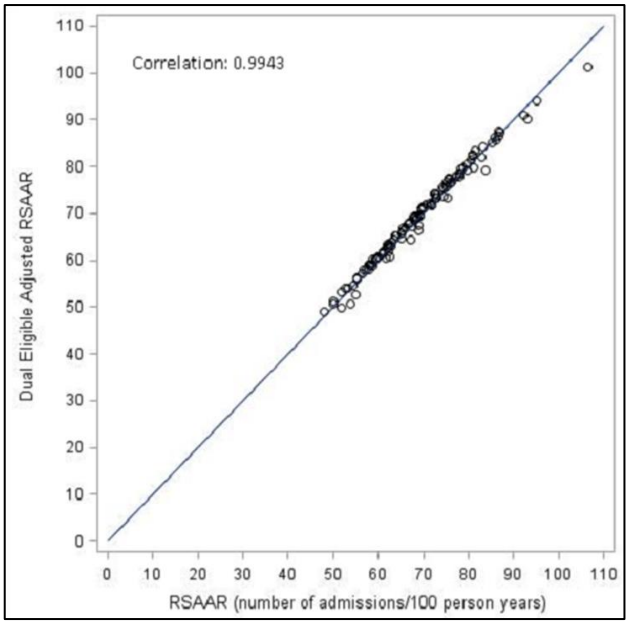
Figure 3. Plot of acute, unplanned admission rates with and without adjustment for AHRQ SES



Medicaid Dual-Eligibility Status

The deviance R-squared values for the two models, one adjusted for the 45 clinical variables *and* Medicaid dual-eligibility status, and one adjusted for the 45 clinical variables *without* Medicaid dual-eligibility status, were 0.124 and 0.123, respectively, meaning adjustment for dual-eligibility status explained similar amount of variation and did not appreciably improve model fit. Comparing the RSAAR with and without Medicaid dual-eligibility status included in the model resulted in a high degree of correlation (Spearman correlation = 0.994). The graph demonstrates that, compared with *not* adjusting for Medicaid dual-eligibility status, adjusting for Medicaid dual-eligibility status results in some ACOs having slightly lower RSAAR scores (below the line) and other

Figure 4. Plot of acute, unplanned admission rates with and without adjustment for dual-eligibility status



In assessing the relationship between the proportions of low SES patients enrolled in an ACO and ACO measure performance, we found that ACOs serving many low SES patients more often perform worse than the national rate compared with ACOs serving few low SES patients. This was true using either the AHRQ SES index (37.9% vs. 6.9%, respectively) or Medicaid dual-eligibility status (39.3% vs. 3.6%, respectively) as an indicator of patients' SES. However, among ACOs serving many low SES patients, using the AHRQ SES index, 8 ACOs (27.6%) performed 'better than the national rate;' using Medicaid dual-eligibility status as an indicator, 6 ACOs (21.4%) performed 'better than the national rate.'

Figure 5. Boxplots of risk-standardized acute admission rates (RSAARs), comparing ACOs with varying proportions of low SES patients with MCC (based on AHRQ SES Index; Quartile 1 [Q1]: ACOs with few low SES patients; Quartile 4 [Q4]: ACOs with many low SES patients)

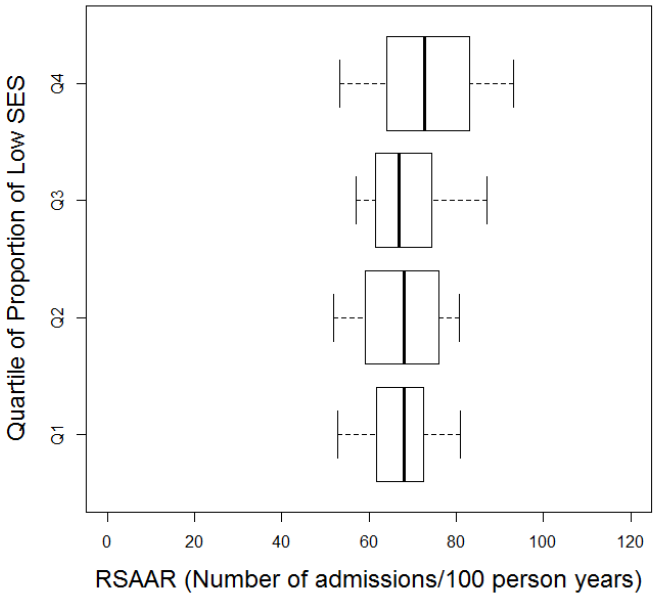
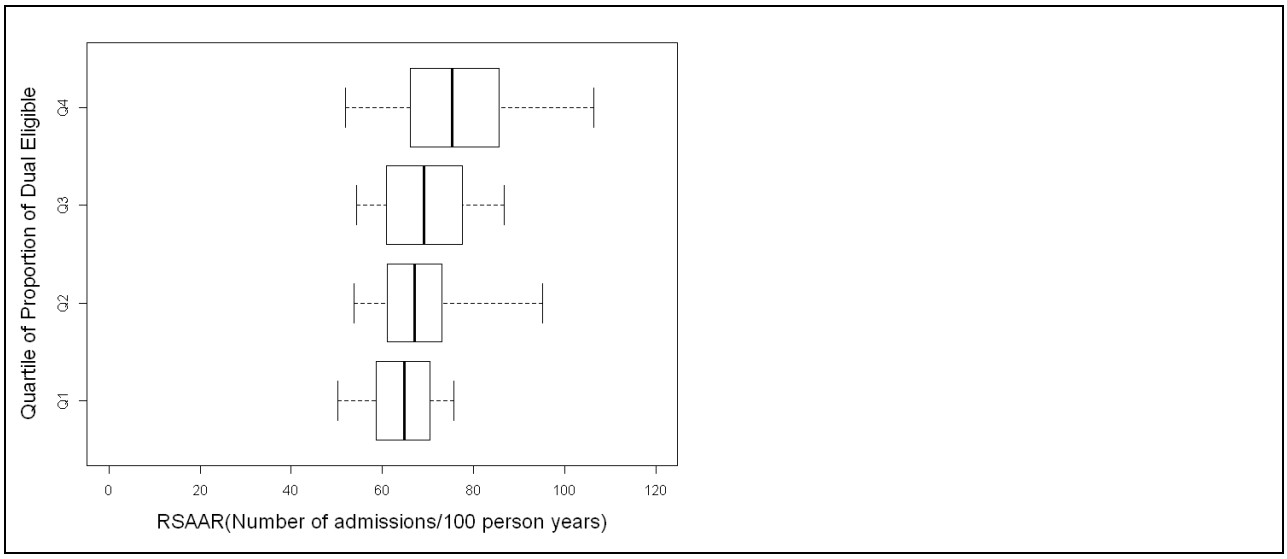


Figure 6. Boxplots of risk-standardized acute admission rates (RSAARs), comparing ACOs with varying proportions of Medicaid dual-eligible patients with MCC (Quartile 1 [Q1]: ACOs with few Medicaid dual-eligible patients; Quartile 4 [Q4]: ACOs with many Medicaid dual- eligible patients)


Quartile of Proportion of Dual Eligible
Q4
Q3
Q2
Q1
0
20
40
60
80
100
120
RSAAR(Number of admissions/100 person years)

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*)

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

We assessed adequacy of the patient-level risk-adjustment model (described above). We evaluated model performance first in the 2012 Development Sample. We then validated the model performance in the 2012 Validation Sample.

The measure uses the number of acute unplanned hospital admissions per person-year at risk for admission. Because the outcome is a count of hospital admissions – rather than a binary outcome, such as whether or not a patient has been admitted – several routinely used metrics of model performance cannot be applied (for example, we cannot use a c-statistic).

Using the 2012 Development Sample, we computed two summary statistics for assessing the risk-adjustment model performance: goodness-of-fit statistics (deviance R squared) and overfitting indices. We then compared the model performance in the development sample with its performance in the validation sample.

Deviance R-squared
Our measure uses a negative binomial function because the outcome is a count of hospital admissions with over-dispersion. We calculated deviance R-squared using the deviance residual defined by Cameron [1]. The deviance R squared evaluates how successful the fit is in explaining the variation of the data. Deviance R squared can take on any value between zero to one, with a value closer to one indicating that a greater proportion of deviance is accounted for by the model. For example, a deviance R-squared value of 0.12 means that the fit explains 12% of the total deviance.

Overfitting indices
Overfitting refers to the phenomenon in which a model accurately describes the relationship between the predictive variables and the outcome in the development dataset, but fails to provide valid predictions in new patients.

Model performance among patients at different risk of admission
In order to determine whether the model performs well across groups of patients at different risk of admission, the sample was divided into quartiles of predicted admission rate (highest, second highest, lowest, and second lowest). We then assessed the model probability of the number of admissions compared with the observed probability of the number of admissions.

Generally, residuals measure the departure of fitted values from actual values of the dependent variable, but they cannot be applied to count data. For linear models, a residual is easily defined as the difference between actual and fitted values. For nonlinear models, the definition of a residual is not unique. Specifically, for count data, the raw residual (the observed value minus

the fitted value) is heteroskedastic and asymmetric. Therefore, there is no residual that has zero mean, constant variance, and symmetric distribution. For fully parametric models such as negative binomial models, we can compare *predicted* probabilities with *observed* probabilities of each count of admissions. For each patient, we can calculate the *predicted* probability of being admitted to the hospital *n* times (0, 1, 2, …*n*) given this patient's risk factors for hospitalization. For example, a patient has a single predicted admission rate of 2.5 admissions per person-years of exposure; however, given the assumed negative binomial distribution of the risk of admissions, we can also express the patient's risk of admission as the probabilities of observing 0, 1, 2,…10 hospital admissions. Therefore, for each patient, we can calculate a set of predicted probabilities of observing different counts of admissions. The *predicted* probability for a group of patients is the average probability of observing 0, 1, 2, …*n* hospital admissions, given these patients' risk factors for admission. The *observed* probability of each count of admissions for a group of patients is the proportion of these patients admitted to the hospital 0, 1, 2, …*n* times.

Citations

1. Cameron AC, Windmeijer FAG. R-Squared Measures for Count Data Regression Models with Applications to Health-Care Utilization. Journal of Business & Economic Statistics. 1996;14(2):209-220.

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

2012 Development Sample results (deviance R-squared): 0.123
2012 Validation Sample results (deviance R-squared): 0.123

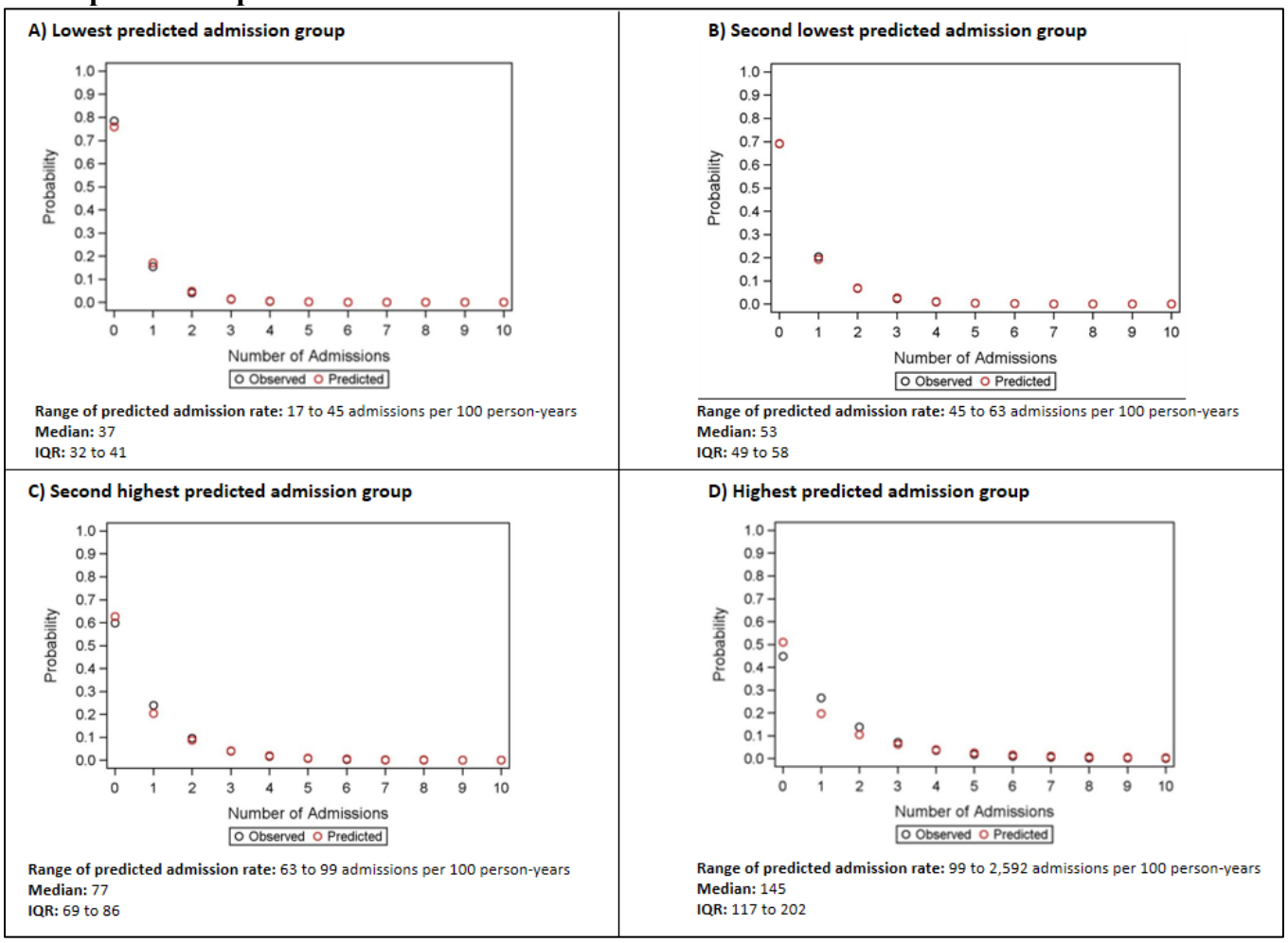
2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2012 Development Sample calibration results (overfitting index): (0.0000, 1.0000)
2012 Validation Sample calibration results (overfitting index): (-0.0015, 1.0011)

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

Below are plots of observed vs. predicted probabilities for the number of hospital admissions among four groups of patients: lowest (A) and second lowest (B) predicted admissions and second highest (C) and highest (D) predicted admissions in the 2012 Development Sample:

Figure 7. Comparison of observed versus predicted probability for the number of hospital admissions among patients with multiple chronic conditions by risk quartile in the 2012 Development Sample



2b4.9. Results of Risk Stratification Analysis:

Not applicable. This measure is not risk-stratified.

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*)

Model performance was similar in the development and validation datasets, with strong model discrimination and fit. The overfitting index of $\gamma 0$ close to 0 and $\gamma 1$ close to 1 indicates good calibration of the model [1]. Additionally, the plots of observed and predicted probabilities for each number of hospital admissions (0, 1, 2, …, 10) across four risk groups showed that the models perform well across a broad range of risk (see Figure 7). In the highest risk group, we observed that the observed and predicted probabilities of the number of zero, one, or two admissions differed slightly. However, these differences were small and somewhat expected among the highest risk group of patients.

Citations

1. Cameron AC, Windmeijer FA. R-squared measures for count data regression models with applications to health-care utilization. *Journal of Business & Economic Statistics.* 1996;14(2):209-220.

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*)

Not applicable.

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)*

The method for discriminating ACO-level performance for public reporting has not been determined. For publicly reported readmission measures of hospital outcomes developed with similar methodology, CMS currently estimates an interval estimate (IE) for each risk-standardized rate to characterize the amount of uncertainty associated with the rate, compares the interval estimate to the national crude rate for the outcome, and categorizes hospitals as 'better than,' 'worse than,' or 'no different than' the US national rate. We used that approach here.

In order to determine IEs, we used bootstrapping methods. In brief, we randomly sampled 114 ACOs with replacement. This is done by randomly selecting an ACO from the 114 ACOs, then placing the selected ACO back into the pool, until we got 114 ACOs, with some ACOs being selected more than once. Performance scores were calculated for each random sample of 114 ACOs. If some ACOs were selected more than once in a bootstrapped sample, we treated them as distinct so that we had random effects to estimate the variance components. This process was repeated many times until 3,000 results were obtained for each ACO.

Using the 95% IEs, we assigned each ACO to one of three performance categories: (1) 'better than,' (2) 'no different than,' and (3) 'worse than' than the US national Medicare FFS admission rate of patients with MCCs. Each ACO was compared to all Medicare FFS beneficiaries who met our MCCs cohort criteria, so that each ACO was evaluated against the US national admission rate among Medicare FFS patients with MCCs. The ACO was 'better than' if the 95% IE was completely below the US national Medicare FFS rate among patients with MCCs; 'no different than' if the 95% IE included the US national Medicare FFS rate among patients with MCCs; and 'worse than' if the 95% IE was above the US national Medicare FFS rate among patients with MCCs.

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*)

45 ACOs (39.5%) had RSAARs that were 'no different' from the US national Medicare FFS admission rate of patients with MCCs. An additional 22 ACOs (19.3%) had RSAAR scores 'worse than the national rate,' and 47 ACOs (41.2%) 'better than the national rate.'

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?*)

These results suggest there are meaningful differences in the quality of care received for patients in the 114 ACOs in the ambulatory setting.

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.***

Items 2b6.1-2b6.3 skipped, as this measure has only one set of specifications.

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*)

Not applicable.

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*)

Not applicable.

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*)

Not applicable.

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*)

Not applicable.

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*)

Not applicable.

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*)

Not applicable.

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.**

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)*.

Not applicable. There are no fees, licensing, or other requirements to use any aspect of the measure as specified.

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)
Payment Program	

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

Measure is currently not in use.

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?*)

This measure is not currently publicly reported or used in an accountability application because it only recently completed development. However, in the November 13, 2014 Physician Fee Schedule final rule, CMS finalized adding the measure to the Medicare Shared Savings Program quality measure set (see 79 FR 67912; https://www.gpo.gov/fdsys/pkg/FR-2014-11-13/pdf/2014-26183.pdf).

The measure is planned for pay-for-reporting in the Medicare Shared Savings Program for 2015 and 2016 reporting periods (79 FR 67912, 67916) and for pay-for-performance in the Medicare Shared Savings Program beginning 2017 reporting period (79 FR 67912, 67916).

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 will be used in one or more CMS programs as noted above in 4a.2. The measure has been finalized for use in the Medicare Shared Savings Program. The measure will be pay-for-reporting initially for the 2015 and 2016 reporting periods and then as pay-for-performance beginning in the 2017 reporting period.

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**

Not applicable.

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.

The measure is not currently used in a quality improvement program, but the primary goal of the measure is to provide ACOs with information necessary to implement focused quality improvement.

This measure was evaluated by a group of clinical experts and a technical expert panel (TEP) throughout the measure development process. We received input and feedback on key methodological, clinical, and other measure decisions as well as on its utility in guiding focused quality improvement within ACOs.

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.**

In designing the measure, we sought to minimize the potential of this measure to result in the denial of future care to high-risk individuals. We developed the patient cohort exclusions and risk-adjustment model to ensure providers who care for patients at higher risk of admission will not be disadvantaged in the measure. CMS is committed to monitoring this measure's use and assessing potential unintended consequences over time.

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.

No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

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?

No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**

Not applicable.

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.)
Not applicable.

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:** MCC_ACO_Admission_Measure_NQF_Appendix_01-29-16_v1.0.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services (CMS)
Co.2 **Point of Contact:** Vinitha, Meyyur, Vinitha.meyyur@cms.hhs.gov, 410-786-8819-
Co.3 **Measure Developer if different from Measure Steward:** Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (CORE)
Co.4 **Point of Contact:** Elizabeth, Drye, Elizabeth.drye@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 (TEP) Members:

CORE convened a TEP of clinicians, patients, purchasers, and experts in quality improvement to provide input on key methodological decisions.

Lawrence M. Becker, BS- Xerox Corporation (Director, Strategic Partnerships, Alliances and Analytics)
Alex Blum, MD, MPH- Evergreen Health Cooperative (Chief Medical Officer)
Sanjay Doddamani, MD- Geisinger Health System (System-wide Chief of Advanced Cardiac Disease Heart Failure)
Kevin Fiscella, MD, MPH- University of Rochester Medical Center (Professor of Family Medicine)
Elbert Huang, MD, MPH- University of Chicago (Associate Professor of Medicine, Director of the Center for Translational and Policy Research of Chronic Diseases, and Associate Director of the Chicago Center for Diabetes Translation Research)
Bruce Leff, MD- Johns Hopkins University School of Medicine (Professor of Medicine, Division of Geriatric Medicine); The Johns Hopkins University Bloomberg School of Public Health (Faculty, Health Services Research Development Center and Lipitz Center for Integrated Health Care)
Andy Miller, MD, MPH- Healthcare Quality Strategies, Inc. (Medical Director);Colorado Foundation for Medical Care (CMO, Integrating Care for Populations & Communities National Coordinating Center)
Ami Parekh, MD, JD- University of California, San Francisco (Medical Director for Health System Innovation)
Christine Ritchie, MD- University of California, San Francisco (Professor of Medicine, Division of Geriatrics)
Two patient representatives

CORE Measure Development Team:
Faseeha Altaf, MPH - Research Project Coordinator
Haikun Bao, PhD – Co-Lead Analyst, diabetes and multiple chronic conditions measure
Susannah Bernheim, MD, MHP - Clinical Investigator
Kanchana Bhat, MPH - Senior Project Manager
Ying Dai, PhD – Co-Lead Analyst, heart failure measure
Elizabeth Drye, MD, SM - Project Director; Project Lead, MCCs measure
Elizabeth Eddy, BA - Research Project Coordinator

Leora Horwitz, MD, MHS - Clinical Investigator
Erin Joyce, BA - Research Assistant
Zhenqiu Lin, PhD - Supporting Analyst
Harlan Krumholz, MD, SM - Director, CORE
Kasia Lipska, MD, MHS - Project Lead, diabetes measure
Julia Montague, MPH - Research Project Coordinator II/Project Manager
Craig Parzynski, MS - Supporting Analyst
Joseph Ross, MD, MHS - Clinical Investigator, CORE
Erica Spatz, MD, MHS - Project Lead, heart failure measure
La'Mont Sutton, MPH – Research Associate
Vera Zhang, MPH – Supporting Analyst

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:**

Ad.3 **Month and Year of most recent revision:**

Ad.4 **What is your frequency for review/update of this measure?** Not applicable.

Ad.5 **When is the next scheduled review/update for this measure?**

Ad.6 **Copyright statement:** Not applicable.

Ad.7 **Disclaimers:** Not applicable.

Ad.8 **Additional Information/Comments:** Not applicable.

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 #: 2860
De.2. **Measure Title:** Thirty-day all-cause unplanned readmission following psychiatric hospitalization in an inpatient psychiatric facility (IPF)
Co.1.1. **Measure Steward:** Centers for Medicare & Medicaid Services
De.3. **Brief Description of Measure:** This facility-level measure estimates an all-cause, unplanned, 30-day, risk-standardized readmission rate for adult Medicare fee-for-service (FFS) patients with a principal discharge diagnosis of a psychiatric disorder or dementia/Alzheimer's disease.

The performance period for the measure is 24 months.
1b.1. **Developer Rationale:** Benefits have been seen in other sectors of care that have a readmission performance measure. The 30-day readmission rate for acute care hospitals held at a constant rate of 19% between 2007 and 2011. After the Hospital Readmissions Reduction Program began in 2012, readmission rates fell to 18.5%, and recent data suggest that these rates continue to decline. This decrease translates to 130,000 fewer hospital readmissions over an eight-month period.[1]

Moreover, because readmission is an outcome measure that is influenced by multiple care processes and structures, as well as the entire healthcare team, it promotes a systems approach to improvement and providing care. A readmission measure promotes shared accountability and collaboration with patients, families, and providers in other settings of care.

Citation for Section 1b.1
1. Centers for Medicare & Medicaid Services. (2013, December 6). New data shows Affordable Care Act reforms are leading to lower hospital readmission rates for Medicare beneficiaries. Retrieved January 15, 2015, from http://blog.cms.gov/2013/12/06/new-data-shows-affordable-care-act-reforms-are-leading-to-lower-hospital-readmission-rates-for-medicare-beneficiaries/

S.4. **Numerator Statement:** The measure estimates the incidence of unplanned, all-cause readmissions to IPFs or short-stay acute care hospitals following discharge from an eligible IPF index admission. We defined readmission as any admission that occurs on or between Days 3 and 30 post-discharge, except those considered planned.
S.7. **Denominator Statement:** The target population for this measure is Medicare FFS beneficiaries aged 18 years and older discharged from an inpatient psychiatric facility with a principal diagnosis of a psychiatric disorder. Eligible index admissions require enrollment in Medicare Parts A and B for 12 months prior to the index admission, the month of admission, and at least 30 days post discharge. Patients must be discharged alive to a non-acute setting (not transferred). A readmission within 30 days is eligible as an index admission, if it meets all other eligibility criteria.
S.10. **Denominator Exclusions:** The measure excludes admissions for patients:
- Discharged against medical advice (AMA)
- With unreliable data (e.g. has a death date but also admissions afterwards)
- With a subsequent admission on day of discharge and following 2 days (transfers/interrupted stay period)

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Administrative claims
S.26. **Level of Analysis:** Facility

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

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?** Not applicable

New Measure -- Preliminary Analysis

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.

Summary of evidence:

- The developer outlines several care processes that can be undertaken by the provider to influence readmissions, such as: connecting patients with severe mental illness to intensive case management (ICM), ensuring stability of condition at discharge, connecting patients to services they will need post-discharge, transitional interventions such as pre- and post-discharge patient education, structured needs assessments, medication reconciliation/education, transition managers, and inpatient/outpatient provider communication, and discharge planning.

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

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.

- In 2012, approximately 43.7 million adults age 18 or older had a mental illness in the past year and 1.9 million adults received psychiatric care in an inpatient setting, and an analysis of Medicare claims data for calendar years 2012 and 2013 showed that more than 20% resulted in readmission to an IPF or a short-stay acute care hospital within 30 days of discharge. In 2012, average payment per discharge was nearly $10,000.
- The 30-day readmission rate for acute care hospitals held at a constant rate of 19% between 2007 and 2011. After the Hospital Readmissions Reduction Program began in 2012, readmission rates fell to 18.5%, and recent data suggest that these rates continue to decline. This decrease translates to 130,000 fewer hospital readmissions over an eight-month period.
- The developer provides the following Risk-Standardized readmission rate distribution across IPFs from January 2012-December 2013 (n=1,696). Rates ranged from 11.0% to 35.4% with an average rate of 21.0%.

Disparities

- The developer provides a detailed document showing SDS variables evaluated with the conceptual framework.
- Results for Race, Age, Gender, Dual Insurance Status and Disability status show results that are less favorable (or worse) than for the reference group.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

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

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

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): Administrative claims

Specifications:

- This measure calculates the all-cause, unplanned, 30-day, risk-standardized readmission rate for adult Medicare fee-for-service (FFS) patients with a principal discharge diagnosis of a psychiatric disorder or dementia/Alzheimer's disease.
- This measure produces a standardized risk ratio (SRR), which is the "predicted" number of readmissions over the "expected" number of readmissions, is calculated for each IPF.
    - The "predicted" number of readmissions is the number of readmissions, given the IPF's performance and its observed case mix, which is calculated by summing the estimated probabilities of readmission for the index admissions contributing to the IPF, based on the IPF-specific intercept and all other risk factors.
    - The "expected" number of readmissions is the number of readmissions given the national performance and its observed case mix, which is calculated by summing the estimated probabilities of readmission for the index admissions contributing to the IPF, based on the average intercept and all other risk factors.
- The denominator includes Medicare FFS beneficiaries aged 18 years and older discharged from an inpatient psychiatric facility with a principal diagnosis of a psychiatric disorder.
- The data sources for this measure include Medicare Part A and B claims, the Medicare Denominator tables, and the Beneficiary cross reference file
- The performance period is 24 months.
- The measure is risk-adjusted using a statistical risk model (see details below).

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *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 [method of reliability testing]

- The developer used data elements from claims data that have been shown to be have face validity in measure development, health services research, and epidemiologic studies.
- The developer also conducted a descriptive analysis of all candidate risk factors and discarded variables with clinically implausible prevalence or incoherent associations with readmissions.
- To test the reliability of facility-level risk-standardized readmission rates (RSRRs), the developer calculated the intra-class correlation coefficient (ICC) using a test-retest approach that examines the agreement between

repeated measures of the same IPF for the same time period.

- The developer used two test-retest approaches to generate independent samples of patients within the same IPF: a split-half sampling design and bootstrapping.
    - For split-half sampling, the developer randomly sampled half of all eligible index admissions in each facility over the two-year period, resulting in two samples that cover the same two-year period but with case volume the size of a measure that would be calculated with one year of data. The ICC in the split-half sampling design was estimated using the RSRRs of the two split-half samples.
    - For bootstrapping, the developer sampled 1,000 pairs of samples from the original measure cohort with replacement (stratified sampling by IPF), resulting in 1,000 pairs of new samples within each IPF with the identical sample size as in the original measure cohort, thus maintaining the sample size of a two-year measure. The ICC in the bootstrap sampling was estimated for each pair of the bootstrap samples. With the 1,000 ICC estimates from the 1,000 pairs of bootstrap samples, the developer determined the distribution of estimated ICC coefficients and thus could calculate the mean and 95% CI of the ICC.

Results of reliability testing

- Split-half sampling:
    - A total of 716,174 admissions over a 2-year period were examined, with 358,087 in each randomly-selected sample. The RSRR was estimated for each sample using a hierarchical logistic regression model.
    - The average RSRR in the two-split-half samples had means of 21.03% and 20.93 percent.
    - The agreement between the two RSRRs for (as measure by an intra-class correlation coefficient (ICC)) was 0.60. The developer notes that this is on the upper limit of "moderate" according to conventional interpretation.
- Bootstrapping:
    - The ICC obtained from the bootstrapping approach, comparing 1,000 pairs of samples of the original measurement cohort, which were sampled with replacement yielding an identical sample size as the original measurement cohort, is 0.78 (95% CI 0.77-0.80).
    - The developer states that this is considered substantial.
    - The developer notes that the bootstrapping approach is considered advantageous because it avoids biased sampling, maintains the original sample size, and allows estimation of ICC confidence intervals.

Guidance from the Reliability Algorithm

Question 1: Submitted specifications are precise, unambiguous, and complete.
Question 2: Empirical reliability testing was conducted using split-half sampling and bootstrapping.
Question 3: Empirical validity testing of patient-level data was conducted.
Question 4: Reliability testing was conducted with computed performance measure scores for each measure
Question 5: The split-half and bootstrapping methods were appropriate for assessing the proportion of variability due to real differences among measured entities.
Question 6: The ICC was .60 in the split-half sampling with is considered moderate and 0.78 in the bootstrapping which is considered a substantial level of agreement.

Questions for the Committee:

- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

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

2b. Validity

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**

- This measure estimates an all-cause, unplanned, 30-day, risk-standardized readmission rate for adult Medicare

fee-for-service (FFS) patients with a principal discharge diagnosis of a psychiatric disorder or dementia/Alzheimer's disease.

- As a rationale for measuring this health outcomes, the developer suggests that readmissions can be decreased with appropriate care received during the index admission and during the discharge process.
- The developer states that actions such as connecting patients with severe mental illness to intensive case management (ICM), ensuring stability of condition at discharge, connecting patients to services they will need post-discharge, transitional interventions such as pre- and post-discharge patient education, structured needs assessments, medication reconciliation/education, transition managers, and inpatient/outpatient provider communication, and discharge planning can reduce rates of readmissions.

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.

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 performed a systematic assessment of face validity of the measure score.
- The developer states that this measure was developed in concordance with national guidelines for publicly reported outcomes measures. The developer states that both definition of the measure and construction of the risk adjustment model are consistent with established standards for outcome measurement defined in the NQF guidance for outcomes measures, the CMS Measures Management System guidance, and the American Heart Association scientific statement on statistical modeling of outcomes measures.
- Input was obtained from an expert workgroup and TEP composed of key stakeholders including experts in psychiatry, psychology, IPF administration, health services research, and epidemiology.
- The developer states that several features of the measure methodology support validity of the measure data and results.
    - Admissions and readmissions are identified through claims data which are used for billing purposes as well as in health services research and epidemiology.
    - Other CMS readmission measures validated their claims data against medical chart abstracted data and found comparable results.
    - The developer followed approaches implemented in previously developed readmission measures that exclude planned readmissions, which would impose noise in the measurement of performance.
    - The workgroup and TEP reviewed the results of additional analyses related to the following measure components: incidence period for readmission, incomplete capture of readmissions related to charge processes (see section 2b3.3 on exclusions related to interrupted stays), cohort exclusions for transfers and discharges against medical advice, and exclusion of planned readmissions from the pool of readmissions that are considered in calculating readmission rates.
    - Sensitivity analyses were performed including separate modeling of psychiatric and non-psychiatric readmission risk in a multinomial model approach and risk model performance in age-and dementia-stratified cohorts.
- Face validity of the measure score was obtained by a TEP vote at the conclusion of measure development. TEP members were asked to in indicate on a scale of 1 to 9 their level of agreement with a face validity statement.

Validity testing results:

- All 17 members of the TEP voted. The median rating was 7, which indicated agreement with the face validity of the measure. Only 1 out of 17 ratings was in the opposite category, disagreement.
- The distribution of the votes was as follows:
  - Agreement (rating 7-9): 10 votes (59%)
  - Neutral (rating 4-6): 6 votes (35%)
  - Disagreement (rating 1-3): 1 vote (6%)

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 goal of the measure is to assess all psychiatric admissions treated by IPFs. Exclusions were considered only for known limitations with claims data.
- To determine the impact of exclusions, the developer examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion in a sample of adult IPF admissions with admission and discharge between January 1, 2012 and December 31, 2013, discharged alive with a psychiatric principal discharge diagnosis, and enrolled in FFS Part A and B in the 12 months prior to admission, the month of admission, and at least 1 month post-discharge (N=781,986).
- The number and percentage of patients excluded for each criterion are as follows:
  1. Unreliable data: **58 (0.0%)**
  2. Transfers and interrupted stats: **56,644 (7.2%)**
  3. Discharged against medical advice (AMA): **9,110 (1.2%)**

Questions for the Committee:

- *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: **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 [Risk adjustment summary

- This measure employs a hierarchical logistic regression model (a form of hierarchical generalized linear model [HGLM]) to create a hospital level 30-day risk-standardized readmission rate (RSRR).
- Variables considered for inclusion in the model 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 as well as other factors from the literature such as a history of discharge against medical advice, aggression, and self-harm.
- To select clinical risk factors, the developers employed a stepwise logistic regression process with backward elimination of variables, using 100 bootstrap samples derived from the entire measure population via random selection with replacement. The developer retained all variables in the stepwise backward elimination that showed an association with readmission at $p<.15$ in 70% of the bootstrap samples.
- The final set of 63 risk-adjustment variables is included in the testing attachment; the odds ratio associated with each variable is also provided.
- The developers also considered a number of variables related to sociodemographic status (SDS) for potential

inclusion in the risk-adjustment model. Candidate SDS variables were selected for examination based on a review of literature and national data sources.

- **Conceptual analysis of the need for SDS adjustment:**
  - The developers note that the key SDS constructs that may affect the risk of readmission of psychiatric patients include income/poverty, disability, race/ethnicity and language barriers, access to care, education, housing stability, and social support..
  - The mechanisms for the effect of sociodemographic factors on health are complex, interrelated, and may result from a lifelong, cumulative effect of social status on health (Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: the Public Health Disparities Geocoding Project. Am J Public Health. 2005;95(2):312-323; Marmot MW, Richard G. (eds.). Social Determinants of Health. 2nd ed. New York: Oxford University Press; 2005; Marmot M. Commentary: mental health and public health. Int J Epidemiol. 2014;43(2):293-296).
    - The impact of SDS factors can be direct or indirect through their effect on health status, the facility selected to obtain care, and the quality of the specific treatments and care received.
    - Additionally, health status can influence SDS factors.
  - External factors, particularly state and local funding for mental healthcare and social support services, can affect a patient's access to services prior to admission and impact the IPF (e.g., public institutions) and can directly affect readmission related to services available after discharge. Risk models typically do not control for differences in such external factors.

- **Empirical analysis of SDS factors:**
  - The developers note that their approach to selecting SDS variables was to identify variables that improve the final clinical risk mode.
  - The developer first evaluated the univariate associations between each candidate variable and readmission alone and when added individually to the clinical model. At this stage, the developer removed variables from further consideration when the association with readmission was in the opposite direction than expected, based on the literature and conceptual framework; in this case, it is likely that the available variables did not fully or accurately represent the identified SDS construct.
    - When each SDS variable was added on its own to the risk model with the clinical risk factors, several SDS variables had much weaker associations with the outcome. These variables include Medicaid status (dual status), original enrollment in Medicare for disability, unemployment, median household income of census tract, low educational attainment in census tract, race/ethnicity, limited English speaking households, and rural-urban community area (RUCA).
    - For the variable median home value in a patient's census tract, we would have expected that patients in neighborhoods with higher home values (higher SDS) would have lower readmission rates, but this was not the case. The odds ratio for this variable was in the opposite direction of other variables within the same construct for income/poverty. Similarly, the association for variables related to access to care providers in the patient's community was in the opposite direction than would have been expected, based on the literature and our conceptual framework that indicated that patients with access to fewer providers would have higher readmission rates. Finally, a similar pattern was observed for the variable percent of people in the patient's census tract with at least a bachelor's degree. The literature and conceptual model indicate that higher educational attainment (higher SDS) is associated with lower risk of readmission, but this was not the case. All of these variables were dropped from further consideration because their associations with readmission could not be explained.
  - Next the developer conducted a cluster analysis to determine if any of the remaining variables are highly correlated. Of highly correlated variable pairs ($r>0.9$), the developer removed the variable with the weaker univariate association with readmission.
  - The developer then addressed three issues related to the interpretation of associations between SDS variables and readmission rates:
    - The relationship between SDS variables and other clinical risk factors that were considered in the final non-SDS risk adjustment model
    - Confounding by IPF performance

    - The differential relationship of relationship of the SDS variable with the readmission risk across IPF RSRR quintiles.
- The analysis to assess the IPF RSRR quintile as a confounder showed that introduction of IPF performance quintile as a covariate did not have much impact on the odds ratios for any of the SDS variables. This indicates that the prevalence of index admissions with a particular SDS risk factor is not appreciably different across IPF RSRR quintiles, and the SDS association with readmission risk cannot be explained with differential representation across RSRR quintiles.
- The analysis to assess the IPF RSRR quintile as a mediator for the association between the SDS variable and readmission risk showed significant interaction terms for two SDS variables, including disability and race. In all instances, the interaction terms indicated that the association between the SDS variable and readmission risk was reduced in IPF quintiles with lower RSRRs. The developer noted that this could indicate that IPFs with lower readmission rates provide higher quality care and interventions to mitigate the effect of the SDS risk factor on readmission, or that IPFs with lower readmission rates serve patients in communities with additional support services for SDS disadvantaged patients.
    - Based on these results and due to concerns about the potential to adjust, at least in part, for IPF quality, the developer dropped the original reason for enrollment in Medicare and race/ethnicity variables from further consideration.
- Among SDS risk factors in the multivariate model, Medicaid enrollment, percent below poverty, percent of crowded households, percent of people with less than high school diploma, and log of percent of limited English households in the census tract were the only variables with statistically significant odds ratios. Model performance was almost identical to the model without any SDS variables included.
- Given the complexity of accurately measuring SDS in current datasets, the developers do not think the empirical evidence is strong enough to warrant inclusion of any of the current SDS variables in the risk model for this measure.

- **Risk Model Diagnostics:**
    - To validate the risk adjustment model, the developer used bootstrapping in which 1,000 bootstrap samples were randomly drawn from the original dataset with replacement. The bootstrap samples were used as the development dataset, and the original cohort was used as the comparison dataset.
    - To assess the overall performance of their risk-adjustment model, the developers computed several summary statistics, including:
        - Calibration: Reflects over-fitting where a developed model with good predictive performance fails to provide valid predictions in a new dataset. Over-fitting is captured with Over-Fitting Indices (γ0, γ1), which are calculated as follows. Let b denote the estimated vector of regression coefficients. Predicted Probabilities are calculated from (p) = 1/(1+exp{-Xb)}, and Z = Xb. A new logistic regression model that includes only an intercept and a slope by regressing the logits on Z is fitted in the validation sample using Logit(P(Y=1|Z)) = γ0 + γ1Z. Estimated values of γ0 far from 0 and estimated values of γ1 far from 1 provide evidence of over-fitting.
        - Discrimination in terms of predictive ability: Reflects the ability to distinguish between high-risk subjects and low-risk subjects as measured by the range between the lowest and highest risk decile.
        - Discrimination in terms of c statistic: Reflects how accurately the model is able to distinguish between an index admission that does or does not have a readmission. A c-statistic of 0.5 represents random prediction and a c-statistic of 1.0 represents perfect prediction.
        - Distribution of residuals: Reflects whether the difference between observed and expected values is normally distributed and suggests similar model performance across various risk levels. The proportion of residuals below -2 and above 2 should be minimal.
        - Model chi‐square: Reflects model goodness of fit in the development dataset but also providing valid predictions in new patients)
    - C-statistic: **0.660**
        - A c-statistic of 0.660 means that for 66% of all possible pairs of patients—one who was readmitted and one who was no—the model correctly assigned a higher probability to those who were readmitted. Generally, a c-statistic of at least 0.70 is considered acceptable.
        - The developers interpret this as “moderate” predictive discrimination.

Indices		Development Model	Validation Using Bootstrapping (95% CI)
Calibration (over-fitting)	γ^0	0	0 (-0.02, 0.01)
	γ^1	1	1 (0.99, 1.01)
Predictive Ability	p10	9%	8.9% (8.8, 9.1)
	p90	42%	41.9% (41.6, 42.9)
Distribution of Residuals			
<-2		0.0	0 (0, 0)
-2 to <0		79.1	79.1 (79.1, 79.1)
0 to <2		13.4	13.4 (13.3, 13.5)
>=2		7.5	7.5 (7.4, 7.6)
Model Wald X^2 (degrees of freedom=61)		37,858	37,917 (37,242, 38,615)

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 standardized risk ratio was calculated as the predicted number of readmissions over the expected number of readmissions (P/E) for each IPF. This is analogous to the observed over expected ratio (O/E) calculated using simple logistic regression.
- The developer estimated the predicted number of readmissions for each IPF using the sum of the estimated probability of readmission for each index admission at that IPF that was calculated from the hospital-specific intercept α (random effect) and all other risk factors. The expected number of readmissions for each hospital was then calculated using the same sum of readmission probabilities for each index admission that was calculated from the average intercept and all other risk factors.
- Because the predicted number of readmissions was calculated based on the hospital's performance and its observed case mix and the expected number was calculated based on the national performance and its observed case mix, an SRR greater than 1 indicates worse quality of care compared to the national average. An SRR less than 1 indicates better quality of care.
- The SRR was then used to calculate RSRR by multiplying SRR by the overall raw readmission rate for all index admissions in the cohort.
- The developer used bootstrapping to calculate 95% confidence intervals for the RSRR to characterize the uncertainty of the estimate.
- The developer calculated the 2.5th and 97.5th percentile of RSRR estimates as the 95% confidence interval of RSRR.
- The developer's interpretation of this data is that the higher proportion of facilities that are categorized as "better than" or "worse than" the national rate relative to some other NQF-endorsed readmission measures (e.g., NQF #1789 Hospital-Wide Readmission Measure) indicates that the measure is able to discriminate between facilities with varying degrees of performance.

	# of IPFs	Percent of IPFs

Better than national rate	140	8.3
No different than national rate	1,257	74.1
Worse than national rate	227	13.4
Fewer than 25 cases during performance period	72	4.2

Question for the Committee:

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

2b6. Comparability of data sources/methods:

Not applicable

2b7. Missing Data

Not applicable

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

Criterion 3. Feasibility

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.

The developer states:

- All measure elements are readily available in electronic sources via administrative claims data, and coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)
- The measure does not present collection burden because data sources needed to implement the measure are readily available, accessible, and timely.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*

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

Criterion 4: Usability and Use

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**

OR

Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

- This measure is planned for use in Public Reporting Quality Improvement with Benchmarking (external benchmarking to multiple organizations).
- The measure has been submitted through the Measures Under Consideration process for the CMS Inpatient Psychiatric Facility Quality Reporting (IPFQR) Program.

Improvement results N/A
Potential harms:

- No unintended negative consequences were identified during testing

Feedback : N/A

***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**

Criterion 5: Related and Competing Measures

Related or competing measures
Not applicable

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: Thirty-day all-cause unplanned readmission following psychiatric hospitalization in an inpatient psychiatric facility (IPF)

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here:

Date of Submission: 1/29/2016

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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: Readmission

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

☐ Process: Click here to name the process

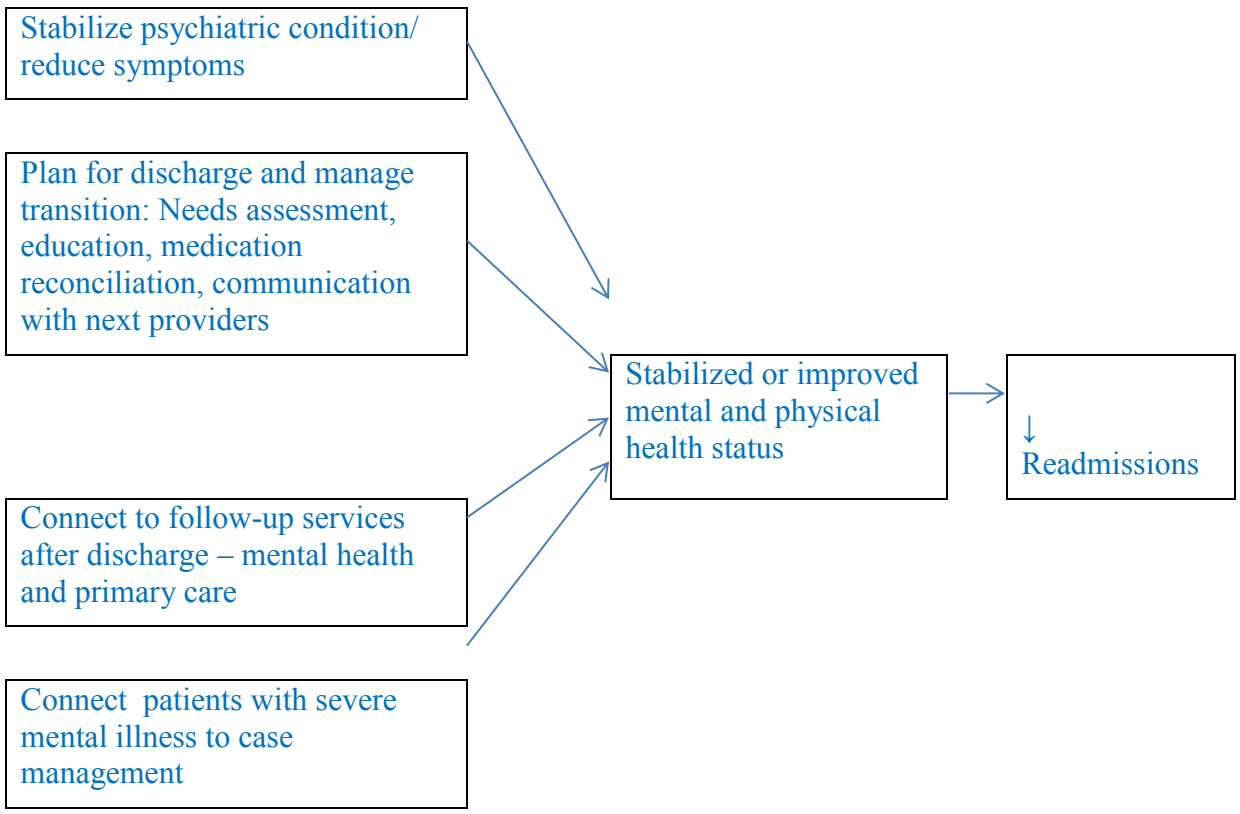
☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

Readmissions can be influenced by the care received during the index admission and during the discharge process.



1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Focused primarily on systematic reviews of the evidence for interventions to prevent readmission, the following information supports the relationship between IPF processes of care and the outcome of readmission. Studies have demonstrated that improvements in the following areas can reduce readmissions:

- Connecting patients with severe mental illness to intensive case management (ICM) may help prevent readmissions. A systematic review of ICM for those with severe mental illness found that compared to standard care, ICM reduced the average number of days in the hospital by 0.86 days per month.[1]
- "Attending to stability of condition" at discharge was found to modestly prevent early readmission by a systematic review of literature on 30-90 day readmissions.[2] Administering effective, evidence-based treatments for psychiatric conditions (e.g., the Veterans Affairs/Department of Defense guideline for management of bipolar disorder)[3] is a pre-requisite to stabilizing patients experiencing an acute episode of a psychiatric disorder and preventing readmissions after discharge.
- Connecting patients to services they will need post-discharge can help prevent readmission. In a study of 30-day behavioral health readmissions using a multistate Medicaid database, a 1% increase in the percent of patients receiving follow-up within seven days of discharge was associated with a 5% reduction in the probability of being readmitted.[4]
- Transitional interventions such as pre- and post-discharge patient education, structured needs assessments, medication reconciliation/education, transition managers, and inpatient/outpatient provider communication have been effective to reduce early psychiatric readmissions. A systematic review of such interventions observed reductions of 13.6% to 37.0%.[5] The time period for counting readmissions varied across studies from 3-24 months post-discharge.
- Similarly, discharge planning in mental health was effective at reducing readmissions. In a systematic review, a meta-analysis of pooled data for 11 studies with a mean follow-up of 3.83 months demonstrated a 34% reduction in risk of readmission.[6]

1a2.1 Citations

1. Dieterich M, Irving CB, Park B, Marshall M. Intensive case management for severe mental illness. *The Cochrane database of systematic reviews.* 2010(10):Cd007906.
2. Durbin J, Lin E, Layne C, Teed M. Is readmission a valid indicator of the quality of inpatient psychiatric care? *J. Behav. Health Serv. Res.* 2007;34(2):137-150.
3. Department of Veterans Affairs/Department of Defense. *Clinical Practice Guideline for Management of Bipolar Disorder in Adults.* Washington, DC: Department of Veterans Affairs, Department of Defense; May 2010.
4. Mark T, Tomic KS, Kowlessar N, Chu BC, Vandivort-Warren R, Smith S. Hospital readmission among medicaid patients with an index hospitalization for mental and/or substance use disorder. *J. Behav. Health Serv. Res.* 2013;40(2):207-221.
5. Vigod SN, Kurdyak PA, Dennis CL, et al. Transitional interventions to reduce early psychiatric readmissions in adults: systematic review. *Br. J. Psychiatry.* 2013;202(3):187-194.
6. Steffen S, Kosters M, Becker T, Puschner B. Discharge planning in mental health care: a systematic review of the recent literature. *Acta Psychiatr. Scand.* 2009;120(1):1-9.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of 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_evidence_attachment-IPF_Readmission.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)*
Benefits have been seen in other sectors of care that have a readmission performance measure. The 30-day readmission rate for acute care hospitals held at a constant rate of 19% between 2007 and 2011. After the Hospital Readmissions Reduction Program began in 2012, readmission rates fell to 18.5%, and recent data suggest that these rates continue to decline. This decrease translates to 130,000 fewer hospital readmissions over an eight-month period.[1]

Moreover, because readmission is an outcome measure that is influenced by multiple care processes and structures, as well as the entire healthcare team, it promotes a systems approach to improvement and providing care. A readmission measure promotes shared accountability and collaboration with patients, families, and providers in other settings of care.

Citation for Section 1b.1
1. Centers for Medicare & Medicaid Services. (2013, December 6). New data shows Affordable Care Act reforms are leading to lower hospital readmission rates for Medicare beneficiaries. Retrieved January 15, 2015, from http://blog.cms.gov/2013/12/06/new-data-shows-affordable-care-act-reforms-are-leading-to-lower-hospital-readmission-rates-for-medicare-beneficiaries/

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.*
Risk-Standardized readmission rate distribution across IPFs
January 2012-December 2013 (n=1,696)

Mean	21.0%
Standard Deviation	3.0%
Min	11.0%
10th percentile	17.3%
20th percentile	18.6%
30th percentile	19.4%
40th percentile	20.2%
50th percentile	20.8%
60th percentile	21.5%
70th percentile	22.3%
80th percentile	23.3%
90th percentile	24.9%
Max	35.4%

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.**
Not applicable. Please see Section 1b.2 for performance data on the measure.

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.*

Disparities are defined using the method from AHRQ National Healthcare Quality and Disparities Report. The difference between two groups must meet the following:

- The difference between the two groups is statistically significant with p <0.05 on a two-tailed test.
- The relative difference between the priority population group and the reference group must have an absolute value of at least 10% when framed positively or negatively ([p1-p2]/p2 >0.1 OR [(1-p1)-(1-p2)]/(1-p2) >0.1).

The results are interpreted as:

- Better = the comparison population estimate is more favorable than reference group estimate by at least 10% and with p-value less than 0.05.
- Worse = the comparison population estimate less favorable than reference group estimate by at least 10% and with p-value less than 0.05.
- Same = comparison population and reference group estimates differ by 10% or less or p-value greater than or equal to 0.05.

Results:

Characteristic: Race // Black // Hispanic // Other // White (reference group)
Index Admissions // 121,783 // 21,174 // 20,604 // 552,613
Readmits // 28,677 // 5,078 // 4,003 // 111,717
Observed Readmit Rate // 23.55 // 23.98 // 19.43 // 20.22
Relative Difference vs Reference // 0.1648 // 0.1863 // -0.039 // 0
p-value // <0.0001 // <0.0001 // 0.0057 //
Disparity compared to Reference // Worse // Worse // Same //

Characteristic: Age // 18-34 // 35-44 // 45-54 // 55-64 // 75-84 // 85+ // 65-74 (reference group)
Index Admissions // 92,281 // 107,682 // 150,626 // 117,317 // 88,310 // 51,404 // 108,554
Readmits // 23,449 // 26,453 // 35,326 // 24,979 // 13,839 // 7,416 // 18,013
Observed Readmit Rate // 25.41 // 24.57 // 23.45 // 21.29 // 15.67 // 14.43 // 16.59
Relative Difference vs Reference // 0.5313 // 0.4804 // 0.4134 // 0.2831 // -0.0556 // -0.1306 // 0
p-value// <0.0001 // <0.0001 // <0.0001 // <0.0001 // <0.0001 // <0.0001 //
Disparity compared to Reference // Worse // Worse // Worse // Worse // Same // Better //

Characteristic: Gender // Male // Female (reference group)
Index Admissions // 348,641 // 367,533
Readmits // 81,514 // 67,961
Observed Readmit Rate // 23.38 // 18.49
Relative Difference vs Reference // 0.2644 //
p-value // <0.0001 //
Disparity compared to Reference // Worse //

Characteristic: Dual Status // Dual // Medicare Only (reference group)
Index Admissions // 420,149 // 296,025
Readmits // 97,431 // 52,044
Observed Readmit Rate // 23.19 // 17.58
Relative Difference vs Reference // 0.319 //
p-value // <0.0001 //
Disparity compared to Reference // Worse //

Characteristic: Disabled // Yes // No (reference group)
Index Admissions // 533,251 // 182,923
Readmits // 122,116 // 27,359
Observed Readmit Rate // 22.9 // 14.96
Relative Difference vs Reference // 0.5311 //
p-value // <0.0001 //
Disparity compared to Reference // Worse //

Characteristic: HRSA MH Shortage Area // Yes // No (reference group)
Index Admissions // 276,062 // 439,593
Readmits // 53,776 // 95,582
Observed Readmit Rate // 19.48 // 21.74
Relative Difference vs Reference // -0.1041 //
p-value // <0.0001 //
Disparity compared to Reference // Better //

Characteristic: HRSA PCP Shortage Area // Yes // No (reference group)
Index Admissions // 269,163 // 446,492
Readmits // 56,299 // 93,059
Observed Readmit Rate // 20.92 // 20.84
Relative Difference vs Reference // 0.0036 //
p-value // 0.4551 //
Disparity compared to Reference // Same //

Characteristic: Rural // Urban // Large Rural // Small Town // Isolated Rural Town // Suburban // Urban (reference group)
Index Admissions // 55,548 // 43,983 // 106,997 // 509,061
Readmits // 9,956 // 7,352 // 19,242 // 112,794
Observed Readmit Rate // 17.92 // 16.72 // 17.98 // 22.16
Relative Difference vs Reference // -0.1911 // -0.2456 // -0.1884 // 0
p-value // <0.0001 // <0.0001 // <0.0001 //
Disparity compared to Reference // Better // Better // Better //

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.**
Not applicable. Please see Section 1b.4 for data on disparities.

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, Patient/societal consequences of poor quality
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.**
Readmission to acute care settings following discharge from inpatient psychiatric facilities (IPF) is both costly to Medicare and undesirable for patients. In 2012, approximately 43.7 million adults age 18 or older had a mental illness in the past year and 1.9 million adults received psychiatric care in an inpatient setting.[1] Our analysis of Medicare claims data for calendar years 2012 and 2013 showed that among the 716,174 IPF admissions for Medicare beneficiaries, more than 20% resulted in readmission to an IPF or a short-stay acute care hospital within 30 days of discharge. Estimates of Medicare payments to IPFs in 2012 indicated that the average payment per discharge was nearly $10,000.[2]

Readmissions for inpatient psychiatric care have the potential to negatively impact millions of individuals suffering from substance abuse and mental health disorders. For those that are avoidable, a readmission causes burden to the patient and the healthcare system.[3] In addition to economic costs, readmissions represent a derailment of recovery and disturbance of relationships.[4] Further, a readmission implies deterioration or exacerbation of a health condition, and this can have implications for patient safety.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
1. Substance Abuse and Mental Health Services Administration. Results from the 2012 National Survey on Drug Use and Health: Mental Health Findings. Retrieved from

http://www.samhsa.gov/data/sites/default/files/2k12MH_Findings/2k12MH_Findings/NSDUHmhfr2012.htm#sec2-1
2. Inpatient Psychiatric Facility Services Payment System. MedPAC. 2014. Retrieved from http://www.medpac.gov/documents/payment-basics/inpatient-psychiatric-facility-services-payment-system-14.pdf
3. Substance Abuse and Mental Health Services Administration. National Mental Health Services Survey (N-MHSS): 2010. Data on Mental Health Treatment Facilities. BHSIS Series S-69, HHS Publication No. (SMA) 14-4837. Retrieved January 9, 2015, from http://www.samhsa.gov/data/sites/default/files/NMHSS2010_Web/NMHSS2010_Web/NMHSS2010_Web.pdf.
4. Maples NJ, Copeland LA, Zeber JE, et al. Can medication management coordinators help improve continuity of care after psychiatric hospitaliztaion? Psychiatr. Serv. 2012;63(6):554-560.

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.)*
Not applicable

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):*
Behavioral Health, Behavioral Health : Alcohol, Substance Use/Abuse, Behavioral Health : Depression, Behavioral Health : Post-Traumatic Stress Disorder (PTSD), Behavioral Health : Serious Mental Illness, Behavioral Health : Suicide, Mental Health, Mental Health : Alcohol, Substance Use/Abuse, Mental Health : Depression, Mental Health : Serious Mental Illness, Mental Health : Suicide, Neurology : Cognitive Impairment/Dementia

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Care Coordination : Readmissions, Patient and Family Engagement, Safety, Safety : Readmissions

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.)*
Not available

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:** S2b_Data_Dictionary-IPF_Readmission-635896801988101932.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Not applicable. This measure is being submitted for initial endorsement.

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 measure estimates the incidence of unplanned, all-cause readmissions to IPFs or short-stay acute care hospitals following discharge from an eligible IPF index admission. We defined readmission as any admission that occurs on or between Days 3 and 30 post-discharge, except those considered planned.

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.)*
The performance period is 24 months.Data 12 months prior to the index admission and 30 days after discharge are needed to identify risk factors and readmissions.

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.
The risk-adjusted outcome measure does not have a traditional numerator and denominator. This section describes the outcome being measured. A readmission is defined as any admission, for any reason, to an IPF or a short-stay acute care hospital (including critical access hospitals) that occurs within 3-30 days after the discharge date from an eligible index admission to an IPF, except those considered planned.

Subsequent admissions on Days 0, 1, and 2 are not counted as readmissions due to transfers/interrupted stay policy. See exclusions for details.

PLANNED READMISSION ALGORITHM
The measure uses the CMS 30-day Hospital-Wide All-Cause Unplanned Readmission (HWR) Measure,Planned Readmission Algorithm version 3.0
Available at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Measure-Methodology.html

The planned readmission algorithm follows two principles to identify planned readmissions:
- Select procedures and diagnoses such as transplant surgery, maintenance chemotherapy/radiotherapy/immunotherapy, rehabilitation, and forceps delivery are considered always planned (summarized in the Data Dictionary, Tables PR1 and PR2).
- Some procedures such as colorectal resection or aortic resection, are considered either planned or unplanned depending on the accompanying principal discharge diagnosis (Data Dictionary, Table PR3). Specifically, a procedure is considered planned if it does not coincide with a principal discharge diagnosis of an acute illness or complication (Data Dictionary, Table PR4).

In the psychiatric patient population, electroconvulsive therapy (ECT) accounted for 41.8% of all potentially planned procedures.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The target population for this measure is Medicare FFS beneficiaries aged 18 years and older discharged from an inpatient psychiatric facility with a principal diagnosis of a psychiatric disorder. Eligible index admissions require enrollment in Medicare Parts A and B for 12 months prior to the index admission, the month of admission, and at least 30 days post discharge. Patients must be discharged alive to a non-acute setting (not transferred). A readmission within 30 days is eligible as an index admission, if it meets all other eligibility criteria.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
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 risk-adjusted outcome measure does not have a traditional numerator and denominator. This section describes the target population for measurement. The target population for this measure is adult Medicare FFS beneficiaries discharged from an IPF. The measure is based on all eligible index admissions from the target population.

An eligible index admission is defined as any IPF admission with the following:
- Admitted to an IPF
- Discharged with a principal diagnosis that indicates psychiatric disorder (AHRQ CCS 650-670)
- Discharged alive
- Age 18 or older at admission

• Enrolled in Medicare FFS Parts A and B during the 12 months before the admission date, month of admission, and at least one month after the month of discharge from the index admission

The measure uses the Clinical Classifications Software (CCS) developed by the Agency for Healthcare Research and Quality (AHRQ), available at https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp, to group ICD9-CM codes into clinically coherent groups.

This measure is limited to admissions for psychiatric causes because IPFs are expected to admit patients who need inpatient care for a psychiatric principal diagnosis (Prospective Payment System for Inpatient Hospital Services. In: Services DoHaH, ed. 42. Vol 412. U.S. Government Publishing Office 2011:535-537). However, a small number of claims (8,658 or 1.1%) had discharge diagnoses that are not in the psychiatric condition categories of CCS 650-670. These admissions could represent coding errors or, more likely, cases where the admission was initiated for psychiatric reasons but during the course of care it became clear that a non-psychiatric illness was the primary diagnosis. Therefore, these admissions are not included in the measure cohort because either they are not typical of inpatient psychiatric facility admissions or they could represent unreliable data.

A readmission to an IPF is counted as another index admission if all denominator criteria are met.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
The measure excludes admissions for patients:
• Discharged against medical advice (AMA)
• With unreliable data (e.g. has a death date but also admissions afterwards)
• With a subsequent admission on day of discharge and following 2 days (transfers/interrupted stay 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)*
DISCHARGE AGAINST MEDICAL ADVICE
Index admissions where there is an indicator in the claims data that patients left against medical advice (AMA) are excluded because the facility may have limited opportunity to complete treatment and prepare for discharge.

UNRELIABLE DATA
Index admissions with unreliable demographic and death information are excluded from the denominator. Unreliable demographic information is defined as age greater than 115 years or missing gender. Unreliable death information is defined as
• An admission with a discharge status of "dead" but the person has subsequent admissions;
• The death date is prior to the admission date; or
• The death date is within the admission and discharge dates for an admission but the discharge status is not "dead".

TRANSFERS/INTERRUPTED STAYS
Index admissions that result in a transfer or interrupted stay are excluded because transfers and interrupted stays cannot always be distinguished from true readmissions in the claims data. This exclusion is defined as an index admission with a readmission on Days 0, 1, or 2 post-discharge.

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)*
Not applicable

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)*
Hierarchical logistic regression is used to estimate a risk standardized readmission rate.

CANDIDATE AND FINAL RISK FACTOR VARIABLES
Four types of risk factors were considered based on empirical analysis, literature review, and clinical judgment:

1. Principal discharge diagnosis of the IPF index admission: Discharge diagnoses were summarized into 13 distinct principal discharge risk variables using a modified version of AHRQ CCS.
2. Comorbidity risk variables: Identified from secondary diagnoses of the index admission and primary or secondary diagnoses of in- and outpatient encounters during the 12-month look-back period using modified CMS condition categories (CC)
3. Other risk factors variables from literature such as history of discharge AMA, aggression and self-harm
4. Age and gender

FINAL SET OF RISK-ADJUSTMENT VARIABLES
Age (7 levels), gender
Principal discharge diagnoses (13)
- CCS 650 Adjustment disorder
- CCS 651 Anxiety
- CCS 652/654/655 ADD/Developmental/Childhood disorders
- CCS 653 Dementia
- CCS 656 Impulse control disorders
- CCS 657.1 Bipolar disorder
- CCS 657.2rc Depressive disorder
- CCS 658 Personality disorder
- CCS 659.1 Schizo-affective disorder
- CCS 659.2 Psychosis
- CCS 660 Alcohol disorder
- CCS 661 Drug Disorder
- CCS 670/663 Other mental disorder

Comorbidities: 26 non-psychiatric CC, 12 psychiatric CC groups
- CC Description (CC or ICD-9-CM)
- AMI (CC 81, 82)
- Anemia (CC 47)
- Arrhythmia (CC 92, 93)
- Asthma (CC 110)
- COPD/Fibrosis (CC 108, 109)
- Delirium (CC 48)
- Diabetes (CC 19, 119, 120)
- Diabetes complications (CC 15-18)
- Dialysis (CC 130)
- Endocrine disease (CC 22, 23)
- Heart disease (CC 83, 84, 89, 90, 104-106)
- Heart failure (CC 80)
- Hematological disorder (CC 44)
- Infection (CC 1, 3-5, 37, 152)
- Injury (CC 150, 151, 155, 156, 160, 162, 163)
- Liver disease (CC 25-29)
- Lung problems (CC 111-115)
- Malnutrition (CC 21)
- Metastasis (CC 7)
- Organ transplant (CC 174, 175)
- Other infection (CC 6)
- Pancreatic disease (CC 32)
- Peptic ulcer (CC 34)
- Seizures (CC 74)
- Uncompleted pregnancy (CC 142, 146, 147)
- Urinary tract disorder (CC 136)
- Adjustment disorder (ICD-9-CM 309.0, 309.22-309.24, 309.28-309.29, 309.3-309.4, 309.82-309.83, 309.89, 309.9, 309.1)
- Anxiety (ICD-9-CM 293.84, 300.01-300.02, 300.00, 300.09, 300.10, 300.20-300.23, 300.29, 300.3, 300.5, 313.0, 313.21, 313.22)
- Bipolar (ICD-9-CM 296.00-296.06, 296.10-296.16, 296.40-296.46, 296.50-296.56, 296.60-296.66, 296.7, 296.80-296.82, 296.89, 296.90, 296.99)
- Depression (ICD-9-CM 296.20-296.26, 296.30-296.36, E950.0-951.1, E951.8, E952.0-952.1, E952.8-953.1, E953.8-953.9, E954, E955.0-955.7, E955.9, E956, E957.0-957.2, E957.9-958.9, E959, 300.4, 311, V62.84)

Developmental disability (CC 66 + ICD-9-CM 758.6-758.7, 758.81, 758.89, 758.9, 759.4, 759.89, 313.1, 313.3, 313.81-313.83, 315.00-315.02, 315.09, 315.1-315.2, 315.31-315.32, 315.34-315.35, 315.39, 315.4-315.5, 315.8-315.9, 313.23, 313.89, 313.9)
Drug/alcohol disorder (CC 51, 52, 53 (except ICD9-CM 305.1) + ICD-9-CM CM 648.31-648.32, 648.34, 655.51, 648.30, 648.33, 655.50, 655.53, 980.0, 965.00-965.02, 965.09, 760.71-760.73, 760.75, 779.5, v654.2)
Intellectual disability (CC 61-64)
Other psych disorders (ICD-9-CM 300.11-300.13, 300.15-300.16, 300.19, 300.6-300.7, 300.81-300.82, 307.1, 307.51, 799.2, 799.21-799.25, 799.29, 300.89, 300.9, 308.0-308.4, 308.9, 312.8, 312.00-312.03, 312.10-312.13, 312.20-312.23, 312.4, 312.81-312.82, 312.89, 312.9, 307.0, 307.9, 307.20-307.23, 307.3, 307.6, 307.7, 309.21, 312.30-312.35, 312.39, 302.0-302.4, 302.50-302.53, 302.6, 302.70-302.76, 302.79, 302.81-302.85, 302.89, 302.9, 306.0-306.4, 306.50-306.53, 306.59, 306.6-306.9, 307.40-307.50, 307.52-307.54, 307.59, 307.80, 307.89, 316)
Personality disorder (CC 57)
Psychosis (CC 56 + ICD-9-CM 295.00-295.05, 295.10-295.15, 295.20-295.25, 295.30-295.35, 295.40-295.45, 295.50-295.55, 295.60-295.65, 295.80-295.85, 295.90-295.95, 297.0-297.3, 297.8-297.9)
PTSD (ICD-9-CM 309.81)
Schizo-affective (ICD-9-CM 295.70-295.75)
Discharged AMA in prior 12 months
Suicide attempt/self-harm — identified by the presence of at least one inpatient or outpatient claim with diagnosis of suicidal attempt or self-harm in the 12-month look-back period.
Aggression — identified by the presence an ICD-9-CM code indicating aggression as a secondary diagnosis on the index admission or on an inpatient or outpatient claim in the 12-month look-back period.

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*)

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.)*
Key Algorithm Steps:
1. Identify all IPF admissions in the performance period.
2. Apply inclusion/exclusion criteria to identify index admissions.
3. Identify readmissions to IPF or short stay acute care hospitals within 30 days of discharge.
4. Apply the planned readmission algorithm to identify unplanned readmissions.
5. Identify risk factors in the 12 months prior to index admission.
6. Run hierarchical logistic regression to compute RSSR for each IPF.

Hierarchical logistic regression is used to model the log-odds of readmission. The two-level specification allows reliable estimates for small-volume hospitals while accepting a certain amount of shrinkage toward the mean. The model includes risk factors as fixed effects and a hospital-specific intercept as random effect. The estimate of hospital-specific intercept reflects the quality of care received at an IPF after adjusting for case mix.

A standardized risk ratio (SRR), which is the "predicted" number of readmissions over the "expected" number of readmissions, is calculated for each IPF. The "predicted" number of readmissions is the number of readmissions, given the IPF's performance and its observed case mix, which is calculated by summing the estimated probabilities of readmission for the index admissions contributing

to the IPF, based on the IPF-specific intercept and all other risk factors. The "expected" number of readmissions is the number of readmissions given the national performance and its observed case mix, which is calculated by summing the estimated probabilities of readmission for the index admissions contributing to the IPF, based on the average intercept and all other risk factors. The confidence interval of the SRR is calculated by bootstrapping. An SRR greater than 1 indicates worse quality of care compared to the national average. An SRR less than 1 indicates better quality of care. The risk-standardized readmission rate (RSRR) is be calculated by multiplying SRR with the overall national readmission rate for better interpretation.

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)*
No diagram provided

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.
Not applicable

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.
Not applicable

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Not applicable

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

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.
For measure calculation, the following Medicare files are required:

- Medicare Denominator tables
- Beneficiary cross reference file
- Institutional claims (Part A)
- Non-institutional claims (Part B)—physician carrier/non-DME

Index admissions and readmissions are identified in the Medicare Part A data. Comorbid conditions for risk adjustment are identified in the Medicare Part A and Part B data in the 12 months prior to and including the index admission. Demographic and fee-for-service (FFS) enrollment information are identified in the Medicare Denominator tables.

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)*
Behavioral Health/Psychiatric : Inpatient
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.)*
Not applicable

2a. Reliability – See attached Measure Testing Submission Form

2b. Validity – See attached Measure Testing Submission Form
NQF_testing_attachment-IPF_Readmission.docx

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

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Thirty-day all-cause unplanned readmission following psychiatric hospitalization in an inpatient psychiatric facility (IPF)
Date of Submission: 1/29/2016
Type of Measure:

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

For measure calculation, the following Medicare files are required:

- Medicare Denominator tables
- Beneficiary cross-reference file
- Institutional claims (Part A)
- Non-institutional claims (Part B)—physician carrier/non-DME

Index admissions and readmissions were identified in the Medicare Part A data. Comorbid conditions for risk adjustment were identified in the Medicare Part A and Part B data in the 12 months prior to and including the index admission. Demographic and fee-for-service (FFS) enrollment information were identified in the Medicare Denominator tables.

1.3. What are the dates of the data used in testing? January 1, 2011 – March 31, 2014. The performance period tested for the measure was January 1, 2012 – December 31, 2013.

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

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

☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

All inpatient psychiatric facilities (IPFs) were included in the analysis. The final measure development cohort included 1,696 IPFs. Among the IPFs, 509 were free-standing facilities, and 1,187 IPF units were within a larger facility. In a two-year measurement period, 72 IPFs had fewer than 25 psychiatric admissions, 1,166 IPFs had 25 to 500 admissions, and 458 IPFs had more than 500 admissions.

1.6. How many and which patients 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*)

This measure was developed for adult admissions to an IPF for Medicare FFS patients who were enrolled in Medicare Parts A & B. The final measure development cohort included 716,174 index admissions. There were 427,273 patients with eligible index admissions during the two-year measurement period. Among them, 49% were male, 77% were white, 59% were enrolled in both Medicare and Medicaid, and 65% were 18 to 64 years of age. Five disorders accounted for over 90% of the 716,174 index admissions: bipolar disorder (22%), depressive disorder (21%), psychosis (18%), schizo-affective disorder (16%), and dementia (14%). The full list of principal discharge diagnoses is shown in Table 1.

Table 1. Index admissions and unadjusted readmission rate by principal discharge diagnosis

Principal Discharge Diagnosis	Denominator	Percent Index Admissions (n=716,174)	Percent Readmitted
CCS 650 Adjustment disorder	6,097	0.9	14.8
CCS 651 Anxiety	8,723	1.2	18.7
CCS 652/654/655 ADD/developmental/childhood disorders	1,854	0.3	17.2
CCS 653 Dementia	99,273	13.9	16.2
CCS 656 Impulse control disorders	2,916	0.4	18.6
CCS 657.1 Bipolar disorder	158,323	22.1	22.5
CCS 657.2/662 Depressive disorder	150,325	21.0	18.0
CCS 658 Personality disorder	1,471	0.2	27.7
CCS 659.1 Schizo-affective disorder	113,218	15.8	26.2
CCS 659.2 Psychosis	131,732	18.4	21.6
CCS 660 Alcohol disorder	19,244	2.7	21.9
CCS 661 Drug disorder	20,560	2.9	19.5
CCS 670/663 Other mental disorder	2,438	0.3	22.7

Note that CCS 657 and CCS 659 were split into two subcategories based on the underlying ICD-9-CM codes of the principal diagnosis to reflect the difference in readmission rates by disorder type and severity.

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.

When we expanded risk adjustment to explore the contribution of sociodemographic variables to risk adjustment and respective effects on risk-standardized readmission rates, we utilized data from the American Community Survey, National Plan and Provider Enumeration System National Provider Identifier (NPPES NPI) Registry, and Health Resources and Services Administration (HRSA) Health Professional Shortage Areas (HPSA) files in addition to the claims data. These data sources allowed us to create additional variables for sociodemographic status (SDS) constructs like access to care and poverty. Variables that required information about the patient's neighborhood required a ZIP code for linkage to these other data sets. Of the total sample of 716,174 index admissions, 1,146 did not have a ZIP code and were therefore excluded from this portion of testing.

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).

To identify potential SDS variables for this measure, we evaluated existing literature on risk factors for readmission following psychiatric discharges. Among the 37 relevant studies, 21 potential variables were identified that represented the SDS constructs for income/poverty, disability, race/ethnicity and language barriers, access to care, education, housing stability, and social supports. We also considered recommendations from the measure workgroup and Technical Expert Panel (TEP). Finally, to ensure consistency in approach with other National Quality Forum (NQF)–endorsed measures, we reviewed risk variables under consideration by other measure developers of admission and readmission measures.

Using the datasets that are currently available (Medicare claims data, American Community Survey, HRSA, HPSA, and NPPES NPI Registry), we identified which constructs could be feasibly evaluated.

Patient-level data such as household income are not available for most of the potential SDS constructs. The only patient-level variables we were able to test were Medicaid enrollment as an indicator of poverty, Medicare Part D enrollment as an indicator of access to prescription drugs, original reason for enrollment in Medicare as an indicator of level of disability, and race.

In the absence of patient-level data, area-based variables provide the best available estimate for the patient, or at a minimum, characterize the patient's exposure to social and economic conditions (Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Race/ethnicity, gender, and monitoring socioeconomic gradients in health: a comparison of area-based socioeconomic measures--the public health disparities geocoding project. *Am J Public Health*. 2003;93(10):1655-1671). The Harvard Public Health Disparities Geocoding Project concluded that the census tract level was best suited to monitor socioeconomic inequalities and specifically recommended percent of persons below poverty level (Krieger, 2003). The researchers found that census tract and block group area-based socioeconomic variables produced similar estimates, while ZIP code measures produced less consistent estimates of expected gradients in health. The authors noted that ascertaining the relative contribution of the individual and the area factors to the association with health was not possible without the patient-level data. However, studies that used both levels of factors had similar results and found that area and individual factors independently and jointly affected some outcomes. Therefore, we created variables of patients' community characteristics based on assignment to census tracts to capture as many SDS constructs reported in the literature as we could.

Note that except for the access variables, which are based on ratios, all of the above-listed variables were ascertained from the American Community Survey information provided on the level of census tracts.

Assignment of index admissions to census tracts was based on the geographic centroid for index admissions with 9-digit ZIP codes, which were available for 80% of all index admissions. For the remainder of index admissions with ZIP codes, we used 5-digit ZIP codes for which we determined the population-weighted centroid, which was then used to assign census tract.

We were not able to create variables for a patient's housing stability, marital status, or availability of social support because that information is not currently collected for all Medicare enrollees.

Table 2 summarizes all considered SDS constructs and whether they were measured on the level of individual patients or neighborhoods, or not available. Supplemental Document #1 to this submission, "Sociodemographic Status Risk Variables – Conceptual Framework and Operationalization," provides more detail on the variables evaluated.

Table 2. SDS constructs with potential variables and level of operationalization

SDS Construct	Variable	Level/Reason Not Used
Income/Wealth	Medicaid enrollment	Patient
	Unemployment	Neighborhood
	Median household income	Neighborhood
	Percentage below poverty level	Neighborhood
	Crowded household	Neighborhood
	Property values	Neighborhood
Disability	Reason for Medicare eligibility	Patient
Race and Ethnicity/ Immigration	Race/ethnicity	Patient
	Percent Hispanic/Latino	Neighborhood
	Limited English language	Neighborhood
Access to Care	HPSA mental health	Neighborhood
	HPSA primary care	Neighborhood
	Psychiatry service access	Neighborhood
	Psychology service access	Neighborhood
	Pharmacy service density	Neighborhood
	Primary care provider density	Neighborhood
	IPF density	Neighborhood
	Rural area	Neighborhood
	Medicare Part D enrollment	Patient
	Uninsured	No variation – all insured
Education	Low education	Neighborhood
	High education	Neighborhood
Socioeconomic Status	Agency for Healthcare Research and Quality (AHRQ) SES categories	Neighborhood
Housing Stability	Housing type, location	Data not available
	Homelessness	Data not available
Social Support	Marital status	Data not available
	Living alone	Data not available
	Level of social support/financial assistance	Data not available

---

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*)

Our measure development process was designed to maximize reliability. We empirically tested reliability of the measure score.

Measure Development Process Designed to Maximize Reliability

To maximize data element reliability, we used data elements from claims data that have been shown to be reliable and have face validity in measure development, health services research, and epidemiologic studies. For example, to optimize sensitivity and specificity of comorbidity risk factors for this measure, we used established algorithms that consider outpatient claims (improved sensitivity) but require at least two claims associated with evaluation and management (E&M) procedure codes to reduce coding errors (improved specificity). We also conducted extensive descriptive analysis of all candidate risk factors and discarded variables with clinically implausible prevalence or incoherent associations with readmissions.

Reliability of Measure Score

To test the reliability of facility-level risk-standardized readmission rates (RSRRs), we calculated the intra-class correlation coefficient (ICC) using a test-retest approach that examines the agreement between repeated measures of the same IPF for the same time period. The randomly sampled sets of admissions from a given hospital are assumed to reflect an independent set of re-measurement of readmission rates for the hospital. Good reliability is assumed if the risk-standardized measure rates calculated from the random datasets for the same IPF are similar. Higher ICC values indicate stronger agreement, and hence, better measure reliability.

We used two test-retest approaches to generate independent samples of patients within the same IPF: a split-half sampling design and bootstrapping. For split-half sampling, we randomly sampled half of all eligible index admissions in each facility over the two-year period, resulting in two samples that cover the same two-year period but with case volume the size of a measure that would be calculated with one year of data. The ICC in the split-half sampling design was estimated using the RSRRs of the two split-half samples.

For bootstrapping, we sampled 1,000 pairs of samples from the original measure cohort with replacement (stratified sampling by IPF), resulting in 1,000 pairs of new samples within each IPF with the identical sample size as in the original measure cohort, thus maintaining the sample size of a two-year measure. The ICC in the bootstrap sampling was estimated for each pair of the bootstrap samples. With the 1,000 ICC estimates from the 1,000 pairs of bootstrap samples, we determined the distribution of estimated ICC coefficients and thus could calculate the mean and 95% CI of the ICC.

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*)

RSRR distributions across IPFs obtained for the two randomly split-half samples that we established for test-retest reliability testing are displayed below. We estimated RSRR for each sample using a hierarchical logistic regression model and RSRR calculations described in 2b5. The average RSRR in the two split-half samples is very similar with means of 21.03 and 20.93 percent (Table 3). The corresponding intra-class correlation coefficient is 0.60.

Table 3. RSRR distributions for IPFs in split-half samples (January 2012–December 2013)

	# Index Admissions	# of IPFs (n≥25)	Mean	SD	Min	10th Percentile	Lower Quartile	Median	Upper Quartile	90th percentile	Max
Sample 1	358,087	1,594	21.03	2.71	12.62	17.73	19.20	20.89	22.72	24.50	31.02
Sample 2	358,087	1,593	20.93	2.56	13.29	17.85	19.14	20.73	22.41	24.36	30.89

The ICC obtained from the bootstrapping approach, comparing 1,000 pairs of samples of the original measurement cohort, which were sampled with replacement yielding an identical sample size as the original measurement cohort, is 0.78 (95% CI 0.77-0.80).

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 ICC of 0.60 obtained from the split-half sample method is on the upper limit of "moderate," according to conventional interpretation (Landis J, Koch G. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159-174). The ICC obtained from the bootstrapping approach is 0.78 (95% CI 0.77-0.80), which is considered "substantial." The chosen bootstrapping approach is considered advantageous because it avoids biased sampling, maintains the original sample size, and allows estimation of ICC confidence intervals (Harrell F. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. New York: Springer; 2001.)

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)*

Validity of this measure was determined by its ability to capture variation in readmission rates across IPFs that are attributable to hospital performance. Our measure development process was designed to maximize validity of the data and the computed risk-adjusted measure score. We assessed validity of the measure score through a stakeholder vote on face validity. We were not able to empirically test validity of the measure score due to the lack of data on IPF quality. However, we conducted empirical validation of the risk model reported in 2b4.

Measure Development Process Designed to Maximize Validity

We developed this measure in concordance with national guidelines for publicly reported outcomes measures. Both definition of the measure and construction of the risk adjustment model are consistent with established standards for outcome measurement defined in the NQF guidance for outcomes measures, the CMS Measures Management System guidance, and the American Heart Association scientific statement on statistical modeling of outcomes measures (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).

We obtained detailed input from an expert workgroup and TEP composed of key stakeholders including experts in psychiatry, psychology, IPF administration, health services research, and epidemiology. The workgroup met frequently to review analyses conducted to support measure specification and risk factor selection. This process enhanced evidence-based decision-making.

Several features of the measure methodology support validity of the measure data and results. First, identifying admissions and discharges for individual patients in claims data is straightforward. Additionally, our measure is based on diagnosis and procedure codes in claims data, which, in addition to being used for billing purposes, are widely used in health services research and epidemiology. Developers of other CMS readmission measures validated their claims data against medical chart abstracted data and found comparable results. Therefore, administrative claims data are widely accepted for use in quality measurement.

Second, for the definition of readmission, we followed approaches implemented in previously developed readmission measures that exclude planned readmissions, which would impose noise in the measurement of performance.

Third, we reviewed with the workgroup and TEP the results of additional analyses related to the following measure components: incidence period for readmission, incomplete capture of readmissions related to charge processes (see section 2b3.3 on exclusions related to interrupted stays), cohort exclusions for transfers and discharges against medical advice, and exclusion of planned readmissions from the pool of readmissions that are considered in calculating readmission rates. We further decided on the exclusion of index admissions without a principal diagnosis related to mental disorders because of small incidence, prohibiting the development of valid risk adjustment models for this population.

Finally, we conducted several sensitivity analyses to ensure optimal model performance. These analyses are listed in Supplemental Document #2, "Draft Technical Report." They include separate modeling of psychiatric and non-psychiatric readmission risk in a multinomial model approach (page 81 of report) and risk model performance in age- and dementia-stratified cohorts (page 86 of report).

For risk adjustment, we conducted a systematic literature review and identified all risk factors that had been used in studies aimed at explaining readmission in psychiatric patients regardless of country, focus on subpopulations, or readmission type. Risk factor selection employed both clinical assessment of risk factor frequencies and plausibility of univariate associations as well as a standard statistical selection process aimed at maximizing the predictive ability of the model.

We paid particular attention to both sensitivity and specificity in risk factor ascertainment by including diagnoses from outpatient billing records, which captured a variety of non-psychiatric comorbidities that were not recorded in the index admission claims. To ensure that the diagnoses assigned to outpatient encounters truly captured the manifestation of a disease as opposed to diagnostic work-up, we restricted outpatient claims to those with E&M procedure codes and required a minimum of two claims with diagnoses within the same Condition Category (CC) grouping.

For risk factor selection, we considered psychiatric and non-psychiatric problems that may necessitate readmission separately to ensure a comprehensive approach to address both etiologies. Because psychiatric etiologies were expected to be dominant, we paid special attention to the sensitivity and specificity of psychiatric risk factors in distinguishing low- and high-risk groups for readmission.

Specifically, we carefully considered the most appropriate way to cluster psychiatric diagnosis codes for risk adjustment. We extracted all ICD-9-CM codes that are included in the AHRQ CCS for principal discharge diagnoses and CMS CC diagnosis classifications that represent mental illness. This mapping exercise resulted in a total of 676 unique ICD-9-CM codes that were grouped into a mental illness category by at least one of the classification algorithms. We then determined differences between the grouping approaches, reviewed frequencies and readmission rates for individual categories as well as individual ICD-9-CM codes, and sought clinical expertise to assess clinical coherence of groupings. This process resulted in regrouping or splitting of several CC and/or CCS categories to optimize the explanatory contribution of each variable in the risk adjustment model.

Face Validity of the Measure Score

Face validity of the measure score was obtained by a TEP vote at the conclusion of measure development. We asked TEP members to indicate on a scale of 1 to 9 their level of agreement with the following face validity statement:

> *The performance score from the readmission measure, as specified (adjusted to account for differences across facilities in the case mix of patients served), represents an accurate reflection of facility-level quality of care related to readmissions.*

Scale: 1-Strongly disagree, 3-Disagree, 5-Neutral, 7- Agree, 9-Strongly agree

We categorized votes as agreement (rating 7-9); neutral (rating 4-6); and disagreement (rating 1-3).
To assess the level of agreement, we identified the category of the median rating and examined the distribution of responses across the three categories to identify the level of disagreement. We identified disagreement if at

least one-third of the ratings were in the agreement category and also one-third in the disagreement category. We reviewed comments to identify any themes related to the ratings.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)
All 17 members of the IPF TEP voted. The median rating was 7, which indicated agreement with the face validity of the measure. Only 1 out of 17 ratings was in the opposite category, disagreement. The distribution of the votes was as follows:
Agreement (rating 7-9): 10 votes (59%)
Neutral (rating 4-6): 6 votes (35%)
Disagreement (rating 1-3): 1 vote (6%)

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?*)
Measure development incorporated all of the aspects discussed above to maximize validity. The risk model was developed and validated as discussed in section 2b4.

The face validity vote indicates that the measure is viewed as valid by the TEP, which is representative of key stakeholders. Only one member disagreed with face validity. Comments for neutral votes reflected either the commenter's inability to assess face validity based on their knowledge and experience or a question about the influence of factors in the post-discharge environment. However, these issues did not cause the TEP members to vote in disagreement with face validity.

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*)
The goal of this measure is to assess all psychiatric admissions treated by IPFs rather than focusing on the outcomes of patients with a specific psychiatric condition. Hence, exclusions were considered only for known limitations with claims data. We analyzed descriptive statistics for the frequency of exclusions.

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*)

Table 4. Selection of the measure population

Exclusion Steps	Total	%
Adult IPF admissions with admission and discharge between January 1, 2012 and December 31, 2013, discharged alive with a psychiatric principal discharge diagnosis, and enrolled in FFS Part A and B in the 12 months prior to admission, the month of admission, and at least 1 month post-discharge	781,986	
• Excluded for unreliable data	58	0.0%
• Excluded for transfers and interrupted stays	56,644	7.2%
• Excluded for discharged AMA	9,110	1.2%
Final cohort (index admissions)	716,174	91.6%

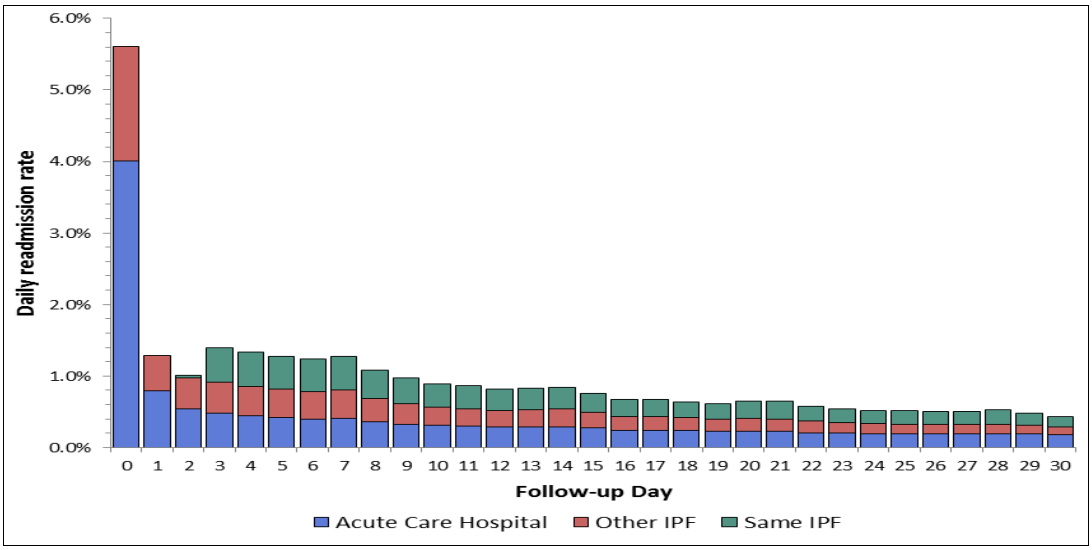
2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must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

- Unreliable data. Patients with unreliable demographic and vital status were not included in the measure because we cannot ensure that those patients meet all of the measure's eligibility criteria.

- Discharged against medical advice. Given that providers have a responsibility to discourage patients with mental illness and potentially impaired decision-making capabilities from leaving AMA and readmission rates for patients who left AMA were higher than those who did not (28.7% versus 20.9%), we were concerned about potentially excluding a particularly vulnerable sub-population of patients from the measure cohort. The workgroup agreed that if admissions that resulted in AMA discharges were to be included in the cohort, the measure would need to be risk adjusted for patients who were admitted involuntarily because these patients leave AMA more frequently and are not evenly distributed across facilities. At the time of measure development, information on involuntary admissions was inadequately captured in claims data. Therefore, index admissions where the patient leaves AMA were excluded from this version of the measure to ensure that results were unbiased with regard to AMA discharges. This exclusion is consistent with the other CMS readmission measures.
- Transfers and interrupted stays. While it would be ideal for the measure to include information on readmissions that occur on Days 0, 1, and 2 post-discharge, these data cannot always be distinguished from transfers and interrupted stays in the claims data. We defined transfers, as in other readmission measures, as a discharge from an IPF (Hospital A) and an admission to another hospital (Hospital B) on the same or next day (Day 0 or Day 1) or a discharge from an IPF (Hospital A) that occurs after admission to another hospital (Hospital B). In these scenarios, the admissions to Hospital A were excluded from the measure cohort, and the admission to Hospital B that met all other eligibility criteria were included as the index admission in the measure cohort.

  An interrupted stay, as defined by CMS reimbursement policy, is a readmission to any IPF before midnight on the third consecutive day following discharge from an IPF. The interrupted stay billing procedure requires one claim if a patient is readmitted to the same IPF within 3 days (Day 0, 1, 2), whereas two claims would be submitted if the patient is readmitted to a different IPF or an acute care facility during this time frame. As a result of this billing policy, very few readmissions to the same IPF appear in the claims data on Days 0, 1, or 2 (Figure 1).

Figure 1. Distribution of readmissions per follow-up day by admitting IPF type (n=716,174)



Admissions with a second admission on Days 0 and 1 post-discharge are already excluded from the measure cohort as transfers. As a result, the interrupted stay policy has implications only for index admissions with readmissions that occur on Day 2 post-discharge. Inclusion of index admissions with readmissions on Day 2 in the measure cohort could create bias because readmissions to different IPFs or acute care hospitals are visible in claims data, while readmissions to the same IPF are not. The readmission locations could be related to the availability of local resources or other parameters related to IPF performance. Therefore, all index admissions with a readmission on Day 2 were excluded from the measure cohort, and readmissions on Days 0 to 2 were not

considered to calculate readmission rates. Like transfers, subsequent admissions to different IPFs on Day 2 that meet all other eligibility criteria were included as an index admission in the measure cohort.

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 63 risk factors**

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

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

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

Not applicable, the measure is risk adjusted.

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*)

Enter the conceptual description (logical rationale or theory informed by literature and content experts) of the causal pathway between the patient sociodemographic factors, patient clinical factors, quality of care, and outcome in Section 2b4.3 of the Measure Testing Attachment

Figure 2 is a simplified representation of the influence of health and SDS factors on the outcome of 30-day readmission. For performance assessment, we only controlled for patient factors that were present prior to the start of care. The risk factors for health status at admission to the IPF included in the risk model are principal diagnosis of the IPF index admission, comorbidities, demographics of age and gender, and prior history of being discharged against medical advice, aggressive behavior, or suicidal ideation, suicide attempt or self-harm. As previously stated when we described the cohort, we used the AHRQ CCS categories for the principal diagnoses in risk adjustment. However, for risk adjustment, we collapsed principal discharge diagnosis ICD-9-CM codes into larger categories, but reviewed crosswalks carefully to ensure optimal capture of differences in readmission rates. This resulted in development of subcategories for schizophrenia/psychosis and bipolar/depressive disorders and further collapsing of developmental/childhood disorders and "other psychiatric disorders."

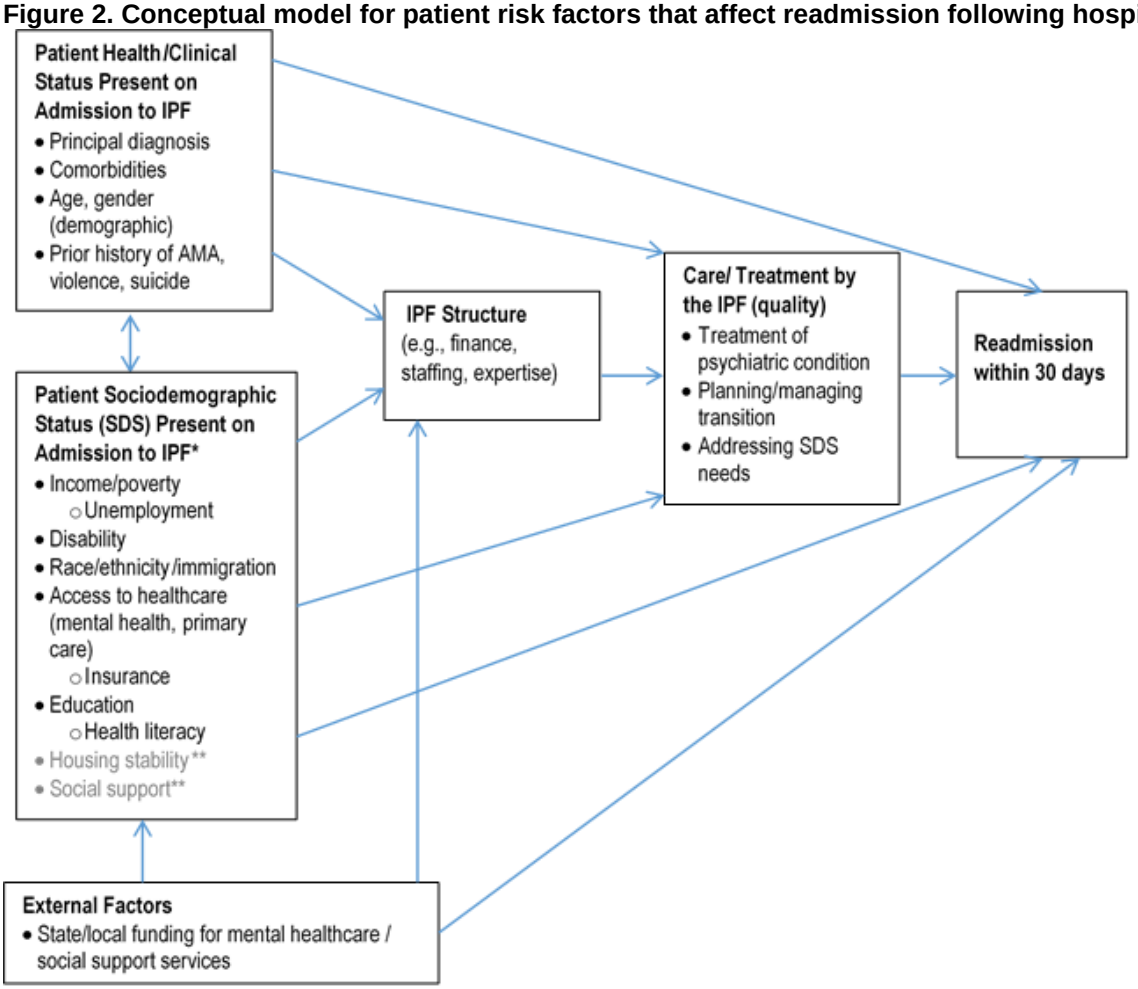
For the comorbidities, we used the CMS CC categories to form clusters on comorbidities, but reviewed crosswalks to optimize the predictive performance of each cluster in capturing ICD-9-CM codes with similar associations with readmissions. This resulted in modification of the ICD-9-CM to CC crosswalk, mostly in following assignments in the comparable CCS category or collapsing certain CC categories based on similar readmission rates. Information on comorbidities was ascertained from the secondary diagnosis of the index admission, after careful review and exclusion of conditions that may represent hospital-acquired complications rather than preexisting comorbidities, principal or secondary diagnoses of hospital admissions during the 12-month look-back period, or presence of at least two outpatient encounter claims with principal or secondary diagnoses of the same CC.

We also identified other variables in the literature that are relevant for the inpatient psychiatric population. These included history of discharge against medical advice, suicide attempts or self-harm, electroconvulsive therapy/transcranial magnetic stimulation (ECT/TMS), or aggression; admission source (as proxy for involuntary admission); and count of psychiatric comorbidities.

The key SDS constructs that may affect the risk of readmission of psychiatric patients include income/poverty, disability, race/ethnicity and language barriers, access to care, education, housing stability, and social support. As depicted in Figure 2, the impact of SDS factors on readmission can be direct or indirect through their effect on health status, the facility selected to obtain care, and the quality of the specific treatments and care received. Additionally, health status can influence SDS factors. The mechanisms for the effect of sociodemographic factors on health are complex, interrelated, and may result from a lifelong, cumulative effect of social status on health (Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: the Public Health Disparities Geocoding Project. *Am J Public Health*. 2005;95(2):312-323; Marmot MW, Richard G. (eds.). *Social Determinants of Health*. 2nd ed. New York: Oxford University Press; 2005; Marmot M. Commentary: mental health and public health. *Int J Epidemiol*. 2014;43(2):293-296). Table 1 of supplemental document #1 to this submission, "Sociodemographic Status Risk Variables – Conceptual Framework and Operationalization," provides additional information on the potential pathways related to each SDS factor.

External factors, particularly state and local funding for mental healthcare and social support services, can affect a patient's access to services prior to admission and impact the IPF (e.g., public institutions) and can directly affect readmission related to services available after discharge. Risk models typically do not control for differences in such external factors.

Figure 2. Conceptual model for patient risk factors that affect readmission following hospitalization



*Operationalized at patient level and/or neighborhood level, as indicated in table under 1.8
**Data not available to operationalize

Selection of Clinical Risk Factors

To select clinical risk factors, we employed a stepwise logistic regression process with backward elimination of variables, using 100 bootstrap samples derived from the entire measure population via random selection with

replacement. For each sample, we ran a logistic regression model including all candidate variables. We retained all variables in the stepwise backward elimination that showed an association with readmission at $p<0.15$ in 70% of the bootstrap samples. Note that use of higher p values is recommended because backward elimination models tend to select models that are smaller than desirable for predictive purposes.

Selection of SDS Risk Factors

Our approach to selecting SDS variables was to identify variables that improve the final clinical risk model. First, we evaluated the univariate associations between each candidate variable and readmission alone and when added individually to the clinical model. At this stage, we removed variables from further consideration when the association with readmission was in the opposite direction than expected, based on the literature and conceptual framework; in this case, it is likely that the available variables did not fully or accurately represent the identified SDS construct.

Next we conducted a cluster analysis to determine if any of the remaining variables were highly correlated. Of highly correlated variable pairs ($r>0.9$), we removed the variable with the weaker univariate association with readmission.

For the remaining variables, we then addressed three issues related to the interpretation of associations between SDS variables and readmission rates. The first issue concerned the relationship between the SDS variable and the other clinical risk factors that were considered in the final non-SDS risk adjustment model. Because SDS risk factors affect health status and cause clinical problems and vice versa, we aimed to examine to which degree our SDS variables were independently associated with readmission rates if all relevant clinical risk factors were considered. We therefore examined the relationship of each individual SDS variable with readmission rates adjusted for all risk factors listed in S.14 in a simple logistic regression framework.

The second issue concerned confounding by IPF performance, which describes a scenario where the association between the SDS variable and readmission rates may actually reflect a correlation between this variable and IPF performance (e.g., index admissions with the SDS variable are more frequently admitted to IPFs with higher readmission rates). Such a finding may suggest that the SDS variable is not independently associated with readmission rates. We expanded the analysis of each individual SDS variable adjusted for clinical risk factors by entering the IPF RSRR quintile estimated from the non-SDS risk model as an additional covariate. A changing association of the SDS variable with readmission risk indicates that index admissions with the SDS variable are clustered in certain RSRR quintiles. The interpretation of such an observation depends on the assumption about the mechanism of an association between the SDS variable and readmission. If the SDS variable is assumed to cause readmission, then its larger presence in certain IPF RSRR quintiles explains some of the observed variation in RSRR, and inclusion of the variable in risk adjustment models can be considered. If the variable is assumed to have no association, then its observed association with readmission rates is based on its differential presence in certain IPF quintiles and inclusion in risk adjustment models should not be considered.

The third issue concerned the differential relationship of the SDS variable with the readmission risk across IPF RSRR quintiles. This analysis was aimed at exploring whether the effect of the SDS variable on readmission risk was consistent across all IPF RSRR quintiles, or if, for example, the SDS variable had a significant association with readmission risk in IPFs with high but not with low RSRR. In order to examine such an effect, we expanded the previous logistic regression analyses that examined the association between readmission risk and each SDS variable adjusted for clinical risk factors and the IPF RSRR quintile by including an interaction term for the SDS variable and the IPF RSRR quintile. While the previous odds ratios average the association between the SDS variable and readmission risk across IPFs, the interaction term allows examination of whether the association differs across quintiles of IPF performance. A significant interaction term and reduction in association between the SDS variable and readmission risk may indicate that IPFs in certain RSRR quintiles may be able to mitigate the effect of the SDS factor (e.g., through higher quality care). Alternatively, it could potentially indicate that IPFs in certain RSRR quintiles may serve patients in communities with support services that mitigate the effect of the SDS factor.

Based on the analyses described above, we decided which SDS variables to evaluate in a multivariate model with the full set of clinical variables listed in S.14. We assessed the impact on the model performance compared to the clinical risk factor–only model in terms of predictive ability, c-statistic, distribution of residuals, model chi square, and distributions of risk-standardized readmission rates.

Finally, considering the contribution of the SDS variables on risk model performance, we evaluated the SDS variables based on their feasibility for use in a national CMS measure.

2b4.4a. What were the statistical results of the analyses used to select risk factors?
This section summarizes analyses involving the clinical risk factor model. The results of the SDS risk factor analyses are in 2b4.4b.

Supplemental Document #2, Draft Technical Report (page 35, Table 12), includes the lists the frequencies and readmission rates of all candidate risk variables. Two variables were eliminated from further model development at this stage:

- Admission legal status: While the expert workgroup noted that the inability to capture involuntary admissions should be considered when interpreting readmission measure rates because patients' cooperation with treatment regimens post-discharge is expected to be lower for patients admitted involuntarily, admission legal status was removed from further model development because of concerns about the reliability of the claims variable. The expert workgroup ultimately agreed that this variable likely does not capture the full spectrum of involuntary admissions and might therefore result in erroneous associations.
- History of ECT/TMS: This variable was removed from further model development because of low frequency and inconsistent associations with the outcome. It showed protective effects, while the literature showed predominantly predictive effects, suggesting its function as proxy for disorder severity.

Table 13 on page 39 of the same document details the output of the selection process, including the number of times a variable was selected, and how many times its beta estimate was positive, indicating a predictive association. The variables that were removed at this stage include: comorbidities of dementia, senility, other cancer, plegia/amputation, sepsis, cardio-respiratory failure, renal failure, coagulation defects, cerebral disease, skin ulcer, cancer, and count of psychiatric comorbidities. The final clinical model is presented in Table 8 in section 2b4.4b.

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)

Enter the analyses and interpretation resulting in decision to include or not include SDS factors in section 2b4.4b of the Measure Testing Attachment. This analysis could include:

- *Variation in prevalence of the factor across measured entities*
- *Empirical association with the outcome (univariate)*
- *Contribution of unique variation in the outcome in a multivariable model*
- *Assessment of between-unit effects versus within-unit effects to evaluate potential clustering of disadvantaged patients in lower quality units*

When each SDS variable was added on its own to the risk model with the clinical risk factors, several SDS variables had much weaker associations with the outcome. These variables include Medicaid status (dual status), original enrollment in Medicare for disability, unemployment, median household income of census tract, low educational attainment in census tract, race/ethnicity, limited English speaking households, and rural-urban community area (RUCA) (Table 5). This is in line with our conceptual framework that SDS and health are interrelated. Some of the effect of SDS on readmission outcomes are captured by health and clinical status.

Table 5. Univariate associations with unplanned all-cause readmission for SDS variables

	Unadjusted		Adjusted for Clinical Risk Factors	
	Odds Ratio	95% CI	Odds ratio	95% CI

Dual status	1.415	1.399	1.432	1.050	1.036	1.064
% Unemployment in Census Tract (CT)	1.013	1.012	1.014	1.004	1.003	1.005
Log Median Household (HH) Income in CT	0.833	0.821	0.846	0.952	0.937	0.967
% Below Poverty in CT	1.007	1.007	1.008	1.002	1.002	1.003
% Crowded HH in CT	1.017	1.016	1.018	1.007	1.006	1.008
Median Owner-occupied HH Value in CT	1.015	1.011	1.018	1.014	1.010	1.018
% Low Education in CT	1.008	1.008	1.009	1.004	1.003	1.004
% High Education in CT	0.999	0.999	0.999	1.000	1.000	1.001
Original Enrollment for Disability	1.689	1.665	1.713	1.048	1.020	1.076
CT is HPSA Mental Health	0.871	0.860	0.881	0.938	0.927	0.950
CT is HPSA Primary Care	1.004	0.993	1.016	1.007	0.995	1.020
PCP Access	1.029	1.028	1.031	1.017	1.015	1.018
IPF Access	2.171	1.621	2.909	3.680	2.707	5.004
Psychiatrist Access	1.098	1.093	1.104	1.054	1.048	1.059
Psychologist Access	1.133	1.122	1.144	1.055	1.044	1.066
Part D Enrollment before IPF Admission	0.989	0.988	0.990	0.998	0.997	0.999
% Hispanic in CT	1.005	1.005	1.005	1.003	1.002	1.003
Log % with Limited English in CT	1.123	1.117	1.130	1.063	1.056	1.069
Race – White	Reference			Reference		
Race – Black	1.216	1.198	1.234	1.095	1.078	1.113
Race – Hispanic	1.245	1.206	1.286	1.064	1.028	1.100
Race – Other	0.952	0.919	0.986	0.920	0.887	0.954
RUCA – Urban	Reference			Reference		
RUCA – Suburban	0.770	0.757	0.784	0.873	0.858	0.889
RUCA – Large Rural	0.767	0.750	0.785	0.873	0.853	0.894
RUCA – Small Town	0.705	0.687	0.724	0.838	0.816	0.861
RUCA – Unknown	0.946	0.524	1.707	1.146	0.626	2.097

In the univariate analyses, several variables had associations in the opposite direction than was expected, based on the literature and the conceptual framework, and were removed from further consideration (Table 5). For the variable median home value in a patient's census tract, we would have expected that patients in neighborhoods with higher home values (higher SDS) would have lower readmission rates, but this was not the case. The odds ratio for this variable was in the opposite direction of other variables within the same construct for income/poverty. Similarly, the association for variables related to access to care providers in the patient's community was in the opposite direction than would have been expected, based on the literature and our conceptual framework that indicated that patients with access to fewer providers would have higher readmission rates. Finally, a similar pattern was observed for the variable percent of people in the patient's census tract with at least a bachelor's degree. The literature and conceptual model indicate that higher educational attainment (higher SDS) is associated with lower risk of readmission, but this was not the case. All of these variables were dropped from further consideration because their associations with readmission could not be explained.

The analysis to assess the IPF RSRR quintile as a confounder showed that introduction of IPF performance quintile as a covariate did not have much impact on the odds ratios for any of the SDS variables (Table 6). This indicates that the prevalence of index admissions with a particular SDS risk factor is not appreciably different across IPF RSRR quintiles, and the SDS association with readmission risk cannot be explained with differential representation across RSRR quintiles.

Table 6. Comparison of SDS associations with readmission rates adjusted for clinical risk factors with and without adjustment for IPF RSRR quintile

	SDS Variable Adjusted For Clinical Risk Factors and IPF RSRR Quintile			SDS Variable Adjusted For Clinical Risk Factors Only		
	Odds Ratio	95% CI		Odds ratio	95% CI	
Quintile 1 Versus 5	0.4907	0.4809	0.5008	---	---	---
Quintile 2 Versus 5	0.6266	0.6150	0.6383	---	---	---
Quintile 3 Versus 5	0.7114	0.6988	0.7243	---	---	---
Quintile 4 Versus 5	0.7979	0.7851	0.8109	---	---	---
Dual Status	1.036	1.022	1.051	1.050	1.036	1.064
% Unemployment in CT	1.003	1.003	1.002	1.004	1.004	1.003
Log Median HH Income in CT	0.940	0.925	0.910	0.939	0.952	0.937
% Below Poverty in CT	1.003	1.003	1.002	1.003	1.002	1.002
% Crowded HH in CT	1.002	1.001	1.003	1.007	1.006	1.008
% Low Education in CT	1.002	1.002	1.003	1.004	1.003	1.004
Part D Enrollment before IPF Admission	0.999	0.998	1.000	0.998	0.997	0.999
Original Enrollment for Disability	1.045	1.017	1.073	1.048	1.020	1.076
% Hispanic	1.001	1.001	1.001	1.003	1.002	1.003
Log % Limited English	1.0201	1.024	1.018	1.030	1.063	1.056
White	Reference			Reference		
Race – Black	1.062	1.045	1.080	1.095	1.078	1.113
Race – Other	0.9112	0.911	0.879	0.945	0.920	0.887
Race – Hispanic	0.980	0.947	1.014	1.064	1.028	1.100
RUCA – Urban	Reference			Reference		
RUCA – Suburban	0.927	0.911	0.943	0.873	0.858	0.889
RUCA – Large rural	0.951	0.928	0.973	0.873	0.853	0.894
RUCA – Small town	0.917	0.893	0.942	0.838	0.816	0.861
RUCA – Unknown	1.207	0.658	2.212	1.146	0.626	2.097

The analysis to assess the IPF RSRR quintile as a mediator for the association between the SDS variable and readmission risk showed significant interaction terms for two SDS variables, including disability and race (Table 7). In all instances, the interaction terms indicated that the association between the SDS variable and readmission risk was reduced in IPF quintiles with lower RSRRs. For example, the row in Table 7 labeled "Disabled*quintile 1 versus 5" has an odds ratio of 0.8403, which indicates that a patient originally enrolled in

Medicare for disability has a 16% lower odds of readmission at a hospital in quintile 1 (higher performance) than he/she does at a hospital in quintile 5 (lower performance). As described above, this could indicate that IPFs with lower readmission rates provide higher quality care and interventions to mitigate the effect of the SDS risk factor on readmission, or that IPFs with lower readmission rates serve patients in communities with additional support services for SDS disadvantaged patients. Based on these results and due to concerns about the potential to adjust, at least in part, for IPF quality, we dropped the original reason for enrollment in Medicare and race/ethnicity variables from further consideration.

Table 7. SDS variables with significant interaction terms for IPF RSRR quintile adjusted for clinical risk factors

	Odds Ratio	95% CI	
Disability	1.1358	1.0943	1.1788
Quintile 1 Versus 5	0.5634	0.5392	0.5886
Quintile 2 Versus 5	0.6975	0.6686	0.7276
Quintile 3 Versus 5	0.7719	0.7413	0.8038
Quintile 4 Versus 5	0.8483	0.8172	0.8805
Disabled * Quintile 1 Versus 5	0.8403	0.7998	0.8829
Disabled * Quintile 2 Versus 5	0.8760	0.8357	0.9182
Disabled * Quintile 3 Versus 5	0.9048	0.8649	0.9466
Disabled * Quintile 4 Versus 5	0.9295	0.8918	0.9687
Black Versus White	1.0914	1.0624	1.1212
Other Versus White	0.8723	0.8176	0.9307
Hispanic Versus White	1.0324	0.9833	1.0840
Quintile 1 Versus 5	0.4999	0.4886	0.5115
Quintile 2 Versus 5	0.6334	0.6200	0.6471
Quintile 3 Versus 5	0.7170	0.7023	0.7319
Quintile 4 Versus 5	0.8035	0.7885	0.8188
Black * Quintile 1 Versus 5	0.9368	0.8857	0.9907
Black * Quintile 2 Versus 5	0.9397	0.8959	0.9856
Black * Quintile 3 Versus 5	0.9714	0.9277	1.0172
Black * Quintile 4 Versus 5	0.9741	0.9353	1.0146
Other * Quintile 1 Versus 5	1.0356	0.9214	1.1640
Other * Quintile 2 Versus 5	1.1334	1.0105	1.2713
Other * Quintile 3 Versus 5	0.9947	0.8858	1.1170
Other * Quintile 4 Versus 5	1.0897	0.9899	1.1995
Hispanic * Quintile 1 Versus 5	0.7994	0.7010	0.9117
Hispanic * Quintile 2 Versus 5	0.9260	0.8184	1.0477
Hispanic * Quintile 3 Versus 5	0.9509	0.8557	1.0566
Hispanic * Quintile 4 Versus 5	0.9152	0.8397	0.9975

Finally, using the remaining SDS candidate risk variables and the clinical risk variables, we compared the multivariate model to one with only clinical risk factors (note that the description of the development and testing of the clinical model are reported in items 2b4.5-2b4.7). Among SDS risk factors in the multivariate model, Medicaid enrollment, percent below poverty, percent of crowded households, percent of people with less than high school diploma, and log of percent of limited English households in the census tract were the only variables with statistically significant odds ratios (Table 8). Model performance was almost identical to the model without any SDS variables included.

Table 8. Risk adjustment model parameters (simple logistic regression)

	Model with Clinical And SDS Risk Factors				Model with Clinical Risk Factors Only			
Risk Variable Name Description	P-Value	Odds Ratio	95% CI		P-Value	Odds Ratio	95% CI	
Intercept	<.0001	0.111	0.097	0.128	<.0001	0.083	0.080	0.086
Demographic Factors								
Gender: Male	<.0001	1.224	1.209	1.240	<.0001	1.225	1.209	1.240
Age								
18-34	<.0001	1.149	1.129	1.169	<.0001	1.304	1.257	1.353
35-44	<.0001	1.092	1.075	1.11	<.0001	1.238	1.194	1.283
45-54	<.0001	1.044	1.029	1.059	<.0001	1.182	1.142	1.223
55-64	0.1056	0.988	0.974	1.003	<.0001	1.110	1.073	1.149
65-74	<.0001	0.899	0.885	0.914	0.9164	0.998	0.967	1.031
75-84	<.0001	0.945	0.927	0.964	0.0115	1.041	1.009	1.074
85+	Reference				Reference			
Principal Discharge Diagnosis on Index Admission								
CCS 650 Adjustment disorder	<.0001	0.763	0.711	0.818	<.0001	0.704	0.653	0.759
CCS 651 Anxiety	0.0907	0.954	0.903	1.008	<.0001	0.878	0.828	0.931
CCS 652/654/655 ADD/developmental/childhood disorders	0.5634	0.967	0.861	1.085	0.056	0.885	0.782	1.003
CCS 653 Dementia	<.0001	1.22	1.186	1.255	<.0001	1.111	1.080	1.144
CCS 656 Impulse control disorders	0.0273	0.902	0.824	0.989	0.0002	0.832	0.754	0.918
CCS 657.1 Bipolar disorder	0.0120	1.029	1.006	1.053	0.0002	0.961	0.942	0.981
CCS 657.2/662 Depressive disorder	0.0015	0.963	0.941	0.986	<.0001	0.894	0.873	0.915
CCS 658 Personality disorder	0.0016	1.196	1.070	1.336	0.1555	1.091	0.968	1.229
CCS 659.1 Schizo-affective disorder	Reference				Reference			
CCS 659.2 Psychosis	<.0001	1.107	1.081	1.134	<.0001	1.048	1.027	1.070
CCS 660 Alcohol disorder	0.0399	1.041	1.002	1.082	0.1069	0.967	0.929	1.007
CCS 661 Drug disorder	<.0001	0.872	0.839	0.906	<.0001	0.810	0.779	0.844
CCS 670/663 Other mental disorder	0.6751	1.02	0.929	1.121	0.2817	0.946	0.855	1.047
Comorbidities								
Psychiatric								
Delirium	<.0001	1.066	1.047	1.086	<.0001	1.064	1.045	1.084
Drug/alcohol disorder	<.0001	1.109	1.093	1.125	<.0001	1.119	1.103	1.135
Schizo-affective disorder	<.0001	1.323	1.302	1.345	<.0001	1.337	1.316	1.359
Psychosis	<.0001	1.154	1.137	1.170	<.0001	1.161	1.145	1.178
Bipolar disorder	<.0001	1.231	1.213	1.248	<.0001	1.235	1.217	1.252
Depression	0.0008	0.971	0.954	0.988	<.0001	0.966	0.949	0.983
Personality disorder	<.0001	1.202	1.183	1.222	<.0001	1.191	1.173	1.211
Anxiety	<.0001	1.096	1.081	1.110	<.0001	1.087	1.073	1.102

Risk Variable Name Description	Model with Clinical And SDS Risk Factors				Model with Clinical Risk Factors Only			
	P-Value	Odds Ratio	95% CI		P-Value	Odds Ratio	95% CI	
Adjustment disorder	<.0001	1.120	1.085	1.155	<.0001	1.111	1.077	1.146
PTSD	<.0001	1.041	1.021	1.061	<.0001	1.039	1.019	1.059
Other psych disorders	<.0001	1.112	1.093	1.131	<.0001	1.111	1.092	1.130
Intellectual disability	0.0857	1.023	0.997	1.050	0.1888	1.018	0.991	1.045
Developmental disability	0.8408	1.003	0.977	1.029	0.9721	1.000	0.975	1.027
Non-psychiatric								
Other infection	<.0001	1.073	1.056	1.091	<.0001	1.081	1.064	1.098
Metastasis	0.0135	1.115	1.023	1.215	0.0105	1.119	1.027	1.220
Diabetes complications	0.0056	1.037	1.011	1.063	0.0013	1.043	1.016	1.069
Diabetes	<.0001	1.032	1.017	1.048	<.0001	1.032	1.016	1.048
Malnutrition	0.372	1.013	0.985	1.041	0.2453	1.016	0.989	1.045
Hematological disorder	0.0012	1.147	1.055	1.247	0.0008	1.153	1.061	1.253
Seizures	<.0001	1.088	1.070	1.107	<.0001	1.091	1.073	1.109
Heart failure	<.0001	1.085	1.061	1.110	<.0001	1.082	1.058	1.107
Arrhythmia	<.0001	1.068	1.048	1.088	<.0001	1.068	1.049	1.089
Asthma	<.0001	1.057	1.039	1.074	<.0001	1.068	1.050	1.086
Dialysis	<.0001	1.357	1.248	1.476	<.0001	1.373	1.263	1.493
Endocrine disease	<.0001	1.074	1.058	1.090	<.0001	1.073	1.057	1.089
Anemia	<.0001	1.094	1.079	1.110	<.0001	1.101	1.086	1.117
AMI	<.0001	1.093	1.049	1.139	<.0001	1.094	1.050	1.140
Pancreatic disease	<.0001	1.104	1.063	1.147	<.0001	1.103	1.062	1.146
Urinary tract disorder	<.0001	1.047	1.026	1.070	<.0001	1.045	1.023	1.067
Peptic ulcer	<.0001	1.088	1.061	1.116	<.0001	1.086	1.059	1.114
Infection	<.0001	1.076	1.056	1.096	<.0001	1.082	1.062	1.102
Liver disease	<.0001	1.135	1.113	1.157	<.0001	1.149	1.127	1.172
Heart disease	<.0001	1.046	1.030	1.062	<.0001	1.047	1.031	1.063
COPD/fibrosis	<.0001	1.091	1.075	1.107	<.0001	1.092	1.076	1.108
Lung problems	0.0003	1.032	1.014	1.050	0.0031	1.026	1.009	1.044
Organ transplant	0.0276	1.119	1.012	1.236	0.0273	1.119	1.013	1.236
Uncompleted pregnancy	0.0475	1.082	1.001	1.170	0.0268	1.092	1.010	1.181
Injury	<.0001	1.044	1.031	1.058	<.0001	1.041	1.028	1.055
Variables from Literature								
Discharged AMA in prior 12 months	<.0001	1.497	1.470	1.523	<.0001	2.239	2.173	2.307
Not discharged AMA in prior 12 months	0.0008	0.983	0.973	0.993	<.0001	1.453	1.429	1.478
No admissions to determine AMA	Reference				Reference			
Suicide attempt/self-harm	<.0001	1.172	1.152	1.192	<.0001	1.181	1.161	1.201
Aggression	<.0001	1.086	1.060	1.112	<.0001	1.090	1.064	1.117
SDS Variables								
Dual eligibility	<.0001	1.043	1.028	1.057	---	---	---	---
% unemployed in CT	0.1100	1.001	1.000	1.002	---	---	---	---
Log median HH income in CT	0.2592	1.017	0.988	1.047	---	---	---	---
% below poverty in CT	0.0137	1.001	1.000	1.002	---	---	---	---
% crowded HH in CT	0.0067	0.998	0.996	0.999	---	---	---	---
% low education in CT	<.0001	1.002	1.001	1.003	---	---	---	---

Risk Variable Name Description	Model with Clinical And SDS Risk Factors				Model with Clinical Risk Factors Only			
	P-Value	Odds Ratio	95% CI		P-Value	Odds Ratio	95% CI	
Part D enrollment before admission	0.6993	1.000	0.999	1.001	---	---	---	---
% Hispanic in CT	0.1064	1.000	1.000	1.001	---	---	---	---
Log % limited English in CT	<.0001	1.031	1.022	1.040	---	---	---	---
RUCA – Urban	Reference				---	---	---	---
RUCA – Suburban	0.2018	0.924	0.818	1.043	---	---	---	---
RUCA – Large Rural	0.1661	0.917	0.811	1.037	---	---	---	---
RUCA – Small town	0.0500	0.884	0.782	1.000	---	---	---	---
RUCA – Unknown	0.2966	1.294	0.797	2.100	---	---	---	---

These analyses show that the remaining SDS variables had only a minimal impact on model discrimination (change in c-statistic of .001) and predictive ability (the difference between observed and predicted readmission rates at the 10th and 90th percentiles changed by 0.2 or less). The model performance parameters are listed in Table 9.

Table 9. Model performance parameters

Indices		Model with Clinical & SDS Factors	Model with Clinical Risk Factors Only
Sample Size		715,655	716,174
Predictive Ability	p10	Observed 10.2%, predicted 8.8%	Observed 10.2%, predicted 9.0%
	p90	Observed 43.7%, predicted 42.1%	Observed 43.4%, predicted 41.9%
Discrimination C-Statistic		0.661	0.660
Distribution of Residuals			
<-2		0.1	0.0
-2 to <0		79.1	79.1
0 to <2		13.3	13.4
>=2		7.5	7.4
Model Wald X^2 (df=74)		38,461 (p<0.001)	37,858

Although we did not think the above analyses indicated that the SDS variables tested should be added to the final risk model, we also analyzed what the impact would be on computed risk-adjusted rates. As seen in Table 10, with SDS adjustment, the median rate was essentially unchanged compared to clinical alone (20.81% vs. 20.80%) and the mean, minimum, maximum and other percentiles changed by less than 1 percentage point.

Table 10. Computed Risk Standardized Readmission Rates for Clinical and Clinical + SDS Models

Readmission Rate	N	mean	SD	Min	p10	Lower Quartile	Median	Upper Quartile	p90	Max
Observed	1696	19.38%	6.49%	0.00%	12.24%	15.46%	19.10%	22.86%	27.33%	46.67%
RSRR Clinical	1696	21.00%	3.01%	10.97%	17.34%	18.99%	20.80%	22.75%	24.95%	35.41%
RSRR Clinical + SDS	1696	20.99%	2.87%	11.09%	17.48%	19.13%	20.81%	22.67%	24.70%	34.49%

Given the complexity of accurately measuring SDS in current datasets, we do not think the empirical evidence is strong enough to warrant inclusion of any of the current SDS variables in the risk model for this measure. For example, Medicaid enrollment (dual status) was the best patient-level indicator of poverty available to us for this testing; however, it is a crude proxy for poverty, given the varying eligibility requirements by state and age. Similarly, for the other variables that require mapping a patient's ZIP code to other information about their

neighborhood, the inclusion of these variables would require the addition of an exclusion criterion that would remove all index admissions without a patient ZIP code. Furthermore, the creation of these variables is time intensive and would add significantly to the computational complexity of the measure with minimal change to model performance or distributions of RSRRs.

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*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

If stratified, skip to 2b4.9

To validate the risk adjustment model, we used bootstrapping in which 1,000 bootstrap samples were randomly drawn from the original dataset with replacement. The bootstrap samples were used as the development dataset, and the original cohort was used as the comparison dataset. This approach allows the use of the entire dataset for model development and a nearly unbiased estimate of predictive accuracy with relatively low variance compared with other validation approaches, such as data splitting and cross-validation (Harrell F. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis.* New York: Springer; 2001). We computed the following summary statistics to assess model performance:

- Calibration: Reflects over-fitting where a developed model with good predictive performance fails to provide valid predictions in a new dataset. Over-fitting is captured with Over-Fitting Indices ($\gamma 0$, $\gamma 1$), which are calculated as follows. Let *b* denote the *estimated vector* of regression coefficients. *Predicted Probabilities* are calculated from (p) = 1/(1+exp{-Xb)}, and *Z* = *Xb*. A new logistic regression model that includes only an intercept and a slope by regressing the logits on Z is fitted in the validation sample using Logit(P(Y=1|Z)) = $\gamma_0 + \gamma_1 Z$. Estimated values of γ_0 far from 0 and estimated values of γ_1 far from 1 provide evidence of over-fitting.
- Discrimination in terms of predictive ability: Reflects the ability to distinguish between high-risk subjects and low-risk subjects as measured by the range between the lowest and highest risk decile.
- Discrimination in terms of c statistic: Reflects how accurately the model is able to distinguish between an index admission that does or does not have a readmission. A c-statistic of 0.5 represents random prediction and a c-statistic of 1.0 represents perfect prediction.
- Distribution of residuals: Reflects whether the difference between observed and expected values is normally distributed and suggests similar model performance across various risk levels. The proportion of residuals below -2 and above 2 should be minimal.
- Model chi□square: Reflects model goodness of fit.

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**
c-statistic = 0.660

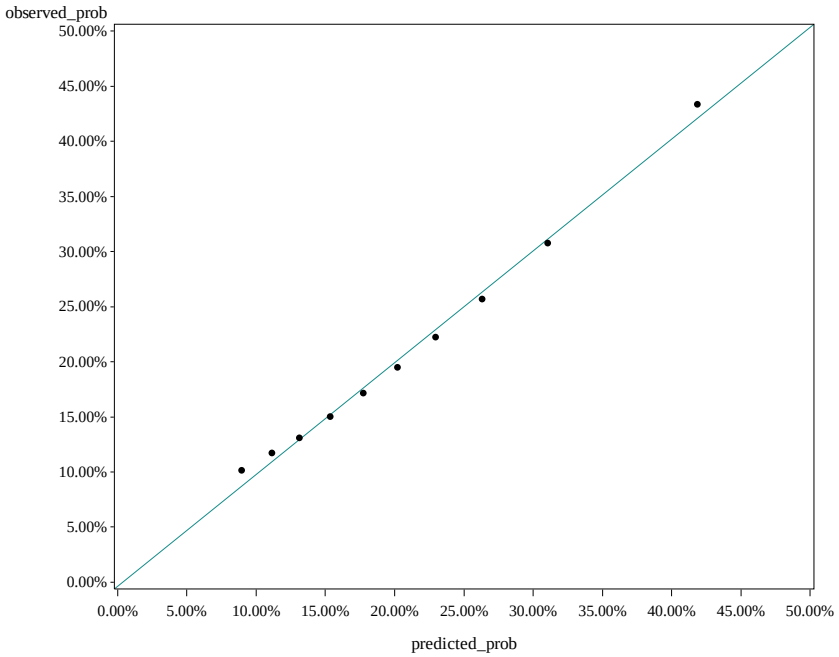
2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Table 11. Risk model performance

Indices		Development Model	Validation Using Bootstrapping (95% CI)
Calibration (over-fitting)	γ^0	0	0 (-0.02, 0.01)
	γ^1	1	1 (0.99, 1.01)
Predictive Ability	p10	9%	8.9% (8.8, 9.1)
	p90	42%	41.9% (41.6, 42.9)
Distribution of Residuals			
<-2		0.0	0 (0, 0)
-2 to <0		79.1	79.1 (79.1, 79.1)
0 to <2		13.4	13.4 (13.3, 13.5)
>=2		7.5	7.5 (7.4, 7.6)
Model Wald X^2 (degrees of freedom=61)		37,858	37,917 (37,242, 38,615)

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

Figure 3. Risk decile calibration plot



2b4.9. Results of Risk Stratification Analysis:

This measure is risk adjusted and is not stratified.

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*)

Risk adjustment model performance parameters showed excellent calibration with no indication of over-fitting. The upper and lower decile of predicted readmission probabilities spans 33%, suggesting good discrimination. The c-statistic of 0.660 suggests moderate predictive discrimination, expressed as the model's ability to distinguish between index admissions that are and are not readmitted.

Estimated model performance parameters are fully confirmed in the validation with near-identical values, owing to the large sample size (716,174 index admissions) within and across IPFs. Statistical findings of excellent

calibration are confirmed when comparing observed to predicted probabilities by risk deciles (see plot in 2b4.8). The results are comparable to other NQF-endorsed readmission measures developed for other settings.

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*)
Referenced in 2b2.2

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)*

Risk-standardized readmission rates (RSRR) for each IPF were estimated from the results of the hierarchical logistic regression model as follows. The standardized risk ratio was calculated as the predicted number of readmissions over the expected number of readmissions (P/E) for each IPF. This is analogous to the observed over expected ratio (O/E) calculated using simple logistic regression. We estimated the *predicted* number of readmissions for each IPF using the sum of the estimated probability of readmission for each index admission at that IPF that was calculated from the hospital-specific intercept α (random effect) and all other risk factors. The expected number of readmissions for each hospital was then calculated using the same sum of readmission probabilities for each index admission that was calculated from the average intercept and all other risk factors.

The standardized risk ratio is then calculated as

$$SRR_j = pred_j/exp_j \qquad (2)$$

where

$$pred_j = \Sigma logit^{-1} (\alpha_j + \beta * Z_{ij}) \qquad (3)$$

$$exp_j = \Sigma logit^{-1} (\mu + \beta * Z_{ij}) \qquad (4)$$

Because the predicted number of readmissions was calculated based on the hospital's performance and its observed case mix and the expected number was calculated based on the national performance and its observed case mix, an SRR greater than 1 indicates worse quality of care compared to the national average. An SRR less than 1 indicates better quality of care.

The SRR was then used to calculate RSRR by multiplying SRR by the overall raw readmission rate for all index admissions in the cohort. We used bootstrapping to calculate 95% confidence intervals for the RSRR to characterize the uncertainty of the estimate. Specifically, we sampled the IPFs with replacement for the bootstrap sample. All index admissions were included in the bootstrap sample if a particular IPF was sampled. IPFs sampled more than once were treated as different hospitals. We ran hierarchical logistic regression on the bootstrap samples. The model results provide the set of hospital-specific intercepts and corresponding variances: $\{\alpha_j , var[\alpha_j]\}$. Since we included the same index admissions for the same IPF in each bootstrap sample, to account for the variability in the hospital random effect, we sampled the hospital-specific intercept from $\alpha_j^* N(\alpha_j , var[\alpha_j])$. We then calculated SRR and RSRR, where SRR is calculated as $SRR_j = \Sigma logit^{-1} (\alpha_j^* + \beta * Z_{ij})/\Sigma logit^{-1} (\mu + \beta * Z_{ij})$. For IPFs sampled more than once in the bootstrap sample, we randomly selected one SRR and RSRR for this sample. Finally, for each IPF, we had 1,000 SRR/RSRR results derived from 1,000 bootstrap samples. We calculated the 2.5^{th} and 97.5^{th} percentile of RSRR estimates as the 95% confidence interval of RSRR.

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*)

Table 12. Distribution of IPF performance categorization

	# of IPFs	Percent of IPFs
Better than national rate	140	8.3
No different than national rate	1,257	74.1
Worse than national rate	227	13.4
Fewer than 25 cases during performance period	72	4.2

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 higher proportion of facilities that are categorized as "better than" or "worse than" the national rate relative to some other NQF-endorsed readmission measures (e.g., NQF #1789 Hospital-Wide Readmission Measure) indicates that the measure is able to discriminate between facilities with varying degrees of performance.

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.***

Enter a comparison of performance scores with and without SDS factors in the risk adjustment model in Section 2b6 of the Measure Testing Attachment.

NOTE: If the measure has more than 1 set of specifications/instructions (e.g., 1 for medical record abstraction and 1 for claims data), then section 2b6 must also be used to demonstrate comparability of the performance scores.
Not applicable – the final measure does not include SDS risk factors.

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*)

- *In Section 2b6.1,enter the method of testing conducted to compare performance scores with and without SDS factors in th*e risk adjustment model for the same entities. Describe the steps and the statistical approach used

Not applicable – only claims data were used.

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*)

- In Section 2b6.2, enter the statistical results from testing the differences in the performance scores with and without SDS factors in the risk adjustment model. (e.g., correlation, rank order)

Not applicable

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*)

- In Section 2b6.3, provide an interpretation of your results in terms of the differences in performance scores with and without SDS factors in the risk adjustment model for the same entities. What do the results mean and what are the norms for the test conducted?

Not applicable

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*)

Missing data were not a problem, given that we used processed claims. As presented in 2b3.2-3, only 58 admissions (0.0%) were excluded due to unreliable data, which included missing gender.

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*)

Not applicable

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*)

Not applicable

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.**
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.
Testing demonstrated that the measure was feasible to specify and calculate using CMS administrative claims data. Data sources needed to implement the measure are readily available, accessible, and timely.

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)*.
The administrative data (collected by CMS primarily for billing purposes) are used as the data source for this measure. Therefore, the cost of data collection is negligible.

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)
Public Reporting Quality Improvement with Benchmarking (external benchmarking to multiple organizations)	

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

Not applicable. The measure is not currently in use.

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?*)

Measure development concluded in Q4 2015. The measure is being submitted for initial endorsement.

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.*)

The measure has been submitted through the Measures Under Consideration process for the CMS Inpatient Psychiatric Facility Quality Reporting (IPFQR) Program.

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**

Not applicable. The measure is being submitted for initial endorsement.

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.

Understanding all-cause readmission rate at the facility-level will facilitate the identification and implementation of innovative care coordination interventions to reduce readmissions. The following studies demonstrate that readmissions can be mitigated by IPFs and that variation in risk-adjusted readmission rates is in part a reflection of the quality of care provided at those facilities. This measure assesses an outcome that reflects the quality of multiple care processes in IPFs and will help focus attention and efforts for improvement.

• Connecting patients with severe mental illness to intensive case management (ICM) may help prevent readmissions. A systematic review of ICM for those with severe mental illness found that compared to standard care, ICM reduced the average number of days in the hospital by 0.86 days per month[1].

• "Attending to stability of condition" at discharge was found to modestly prevent early readmission by a systematic review of

literature on 30-90 day readmissions[2]. Administering effective, evidence-based treatments for psychiatric conditions (e.g., the Veterans Affairs/Department of Defense guideline for management of bipolar disorder)[3] is a pre-requisite to stabilizing patients experiencing an acute episode of a psychiatric disorder and preventing readmissions after discharge.

- Connecting patients to services they will need post-discharge can help prevent readmission. In a study of 30-day behavioral health readmissions using a multistate Medicaid database, a 1% increase in the percent of patients receiving follow-up within seven days of discharge was associated with a 5% reduction in the probability of being readmitted[4].
- Transitional interventions such as pre- and post-discharge patient education, structured needs assessments, medication reconciliation/education, transition managers, and inpatient/outpatient provider communication have been effective to reduce early psychiatric readmissions. A systematic review of such interventions observed reductions of 13.6% to 37.0%[5]. The time period for counting readmissions varied across studies from 3-24 months post-discharge.
- Similarly, discharge planning in mental health was effective at reducing readmissions. In a systematic review, a meta-analysis of pooled data for 11 studies with a mean follow-up of 3.83 months demonstrated a 34% reduction in risk of readmission[6].

Citation for Section 4b.2

1. Dieterich M, Irving CB, Park B, Marshall M. Intensive case management for severe mental illness. Cochrane Database Syst Rev. 2010(10):CD007906.
2. Durbin J, Lin E, Layne C, Teed M. Is readmission a valid indicator of the quality of inpatient psychiatric care? J Behav Health Serv Res. 2007;34(2):137-150.
3. VA/DoD Clinical Practice Guideline for Management of Bipolar Disorder in Adults (BD). Department of Veterans Affairs; 2010. Available at: http://www.healthquality.va.gov/bipolar/bd_306_sum.pdf.
4. Mark TL, Mark T, Tomic KS, et al. Hospital readmission among medicaid patients with an index hospitalization for mental and/or substance use disorder. J Behav Health Serv Res. 2013;40(2):207-221.
5. Vigod SN, Kurdyak PA, Dennis CL, et al. Transitional interventions to reduce early psychiatric readmissions in adults: systematic review. Br J Psychiatry. 2013;202(3):187-194.
6. Steffen S, Kösters M, Becker T, Puschner B. Discharge planning in mental health care: a systematic review of the recent literature. Acta Psychiatr Scand. 2009;120(1):1-9.

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.**

No unintended negative consequences were identified during testing.

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)**

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
1768 : Plan All-Cause Readmissions (PCR)
1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)
1891 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization

2375 : PointRight ® Pro 30™
2380 : Rehospitalization During the First 30 Days of Home Health
2496 : Standardized Readmission Ratio (SRR) for dialysis facilities
2502 : All-Cause Unplanned Readmission Measure for 30 Days Post Discharge from Inpatient Rehabilitation Facilities (IRFs)
2504 : 30-day Rehospitalizations per 1000 Medicare fee-for-service (FFS) Beneficiaries
2510 : Skilled Nursing Facility 30-Day All-Cause Readmission Measure (SNFRM)
2512 : All-Cause Unplanned Readmission Measure for 30 Days Post Discharge from Long-Term Care Hospitals (LTCHs)

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
Hospital, 30-day all-cause risk-standardized readmission rate (RSRR) following acute ischemic stroke hospitalization (Steward: CMS/Yale)
Acute myocardial infarction (AMI): risk-adjusted rate of urgent readmission within 30 days following discharge for AMI (Steward: Canadian Institute for Health Information)
Obstetrics: risk-adjusted rate of urgent readmission for obstetric patients within 30 days of discharge (Steward: Canadian Institute for Health Information)
Surgery: risk-adjusted rate of urgent readmission for adult surgical patients within 30 days of discharge (Steward: Canadian Institute for Health Information)
Medical: risk-adjusted rate of urgent readmission for adult medical patients within 30 days of discharge (Steward: Canadian Institute for Health Information)

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?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
This measure is harmonized to the extent possible with NQF 1789 Hospital-Wide All-Cause Unplanned Readmission Measure (HWR), which is the most closely related measure. Both measures evaluate all-cause, unplanned readmissions following discharge for a broad range of diagnoses. The proposed measure specifically evaluates inpatient psychiatric facilities whereas NQF 1789 evaluates acute-care hospitals. The major differences are: The proposed measure for IPF excludes transfers on Days 0 and 1 and also subsequent admissions on Day 2 because billing procedures related to interrupted stays prevent distinguishing all readmissions during that period; NQF 1789 excludes transfers on Days 0 and 1. The proposed measure has only one risk model; NQF 1789 has 5 risk models for different patient cohorts. Although the proposed measure is also facility-wide, the cohort for this measure is all psychiatric conditions and multiple risk models were not needed. The proposed measure counts readmissions to IPFs and short-stay acute care hospitals (including critical access hospitals): NQF 1789 counts readmissions to short-stay acute care hospitals, not to IPFs. The proposed measure includes patients with psychiatric diagnoses of CCS 650-670: 1789 excludes CCS 650, 651, 652, 654, 655, 656, 657, 658, 659, 662, 670.

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 NQF-endorsed measures that address the same target population. NQF 1789 includes only some patients with psychiatric disorders (substance use and dementia).

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:** NQF_Supplemental_Document-IPF_Readmission.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services
Co.2 **Point of Contact:** Vinitha, Meyyur, Vinitha.Meyyur@cms.hhs.gov, 410-786-8819-
Co.3 **Measure Developer if different from Measure Steward:** Health Services Advisory Group, Inc.
Co.4 **Point of Contact:** Megan, Keenan, mkeenan@hsag.com, 616-425-1997-

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.
MEASURE DEVELOPMENT WORKGROUP
Susannah Bernheim, MD, MHS – Yale New Haven Health Services Corporation Center for Outcomes Research & Evaluation
Alisa Busch, MD, MS – McLean Hospital
Marina Cecchini, MBA – UF Health Shands Psychiatric and UF Health Shands Rehab Hospitals
Betsy Dodd, PharmD, BCPP – University of Florida
Frank Ghinassi, PhD, ABPP – Western Pyschiatric Institute and Clinic Steve Pittman, PhD – Meridian Behavioral Healthcare, Inc.
Andrea Goldenson, PharmD, MS, PhD – Malcom Randall Veterans Affairs Medical Center
Tracy Lenzini, BS – Grand Traverse Health Advocates
Kathleen McCann, RN, PhD – National Association of Psychiatric Health Systems
Gayle Olano-Hurt, MPH, CPHQ, PMC – Sheppard Pratt Health System
Irene Ortiz, MD, MSW – Molina Healthcare of New Mexico
Thomas Penders, MS, MD, DLFAPA – North Carolina Psychiatric Association
Lucille Schacht, PhD – National Association of State Mental Health Program Directors Research Institute, Inc.
Lisa Shea, MD – Butler Hospital
Jeffrey Scott Harman, MS, PhD – University of Florida, Health Service Research
Ben Staley, PharmD – UF Health Pharmacy Services
Rajiv Tandon, MS, MD – University of Florida, Department of Psychiatry; Malcam Randall Veterans Affairs Medical Center
Thomedi Ventura, MS, MSPH – Telligen

The measure workgroup established clinical definitions of the outcome being measured and operationalized the measure specifications. Workgroup members reviewed results from testing and continued to be involved in the iterative process of measure specification revisions.

TECHNICAL EXPERT PANEL (TEP)
Alisa Busch, MD, MS – McLean Hospital
Kathleen Delaney, PhD, PMH-NPRN – Rush College of Nursing
Jonathan Delman, PhD, JD, MPH – Systems and Psychosocial Advanced Research Center, University of Massachusetts Medical School
Frank Ghinassi, PhD, ABPP – Western Pyschiatric Institute and Clinic
Eric Goplerud, PhD – NORC at the University of Chicago
Geetha Jayaram, MD – Schools of Medicine, Health Policy and Management and the Armstrong Institute for Patient Safety, Johns Hopkins University
Charlotte Kauffman, MA, LCPC – State of Illinois-Division of Mental Health
Tracy Lenzini, BS – Grand Traverse Health Advocates
Kathleen McCann, RN, PhD – National Association of Psychiatric Health Systems
Gayle Olano-Hurt, MPH, CPHQ, PMC – Sheppard Pratt Health System
Mark Olfson, MD, MPH – New York State Psychiatric Institute
Irene Ortiz, MD, MSW – Molina Healthcare of New Mexico
Thomas Penders, MS, MD, DLFAPA – North Carolina Psychiatric Association

Lucille Schacht, PhD – National Association of State Mental Health Program Directors Research Institute, Inc.
Lisa Shea, MD – Butler Hospital
Thomedi Ventura, MS, MSPH – Telligen
Elvira Ryan, MBA, BSN, RN – The Joint Commission

The TEP evaluated the proposed measure and discussed the strengths and weaknesses of the proposed measure and made recommendations regarding measure specifications, inclusion and exclusion criteria, and appropriate risk adjustment.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2016

Ad.3 **Month and Year of most recent revision:**

Ad.4 **What is your frequency for review/update of this measure?** Annually

Ad.5 **When is the next scheduled review/update for this measure?**

Ad.6 **Copyright statement:** Not applicable

Ad.7 **Disclaimers:** Not applicable

Ad.8 **Additional Information/Comments:** None

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 #: 2884
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 is a cancer-specific measure. It provides the rate at which all adult cancer patients (= 18 years old), regardless of payer type, have an unplanned re-hospitalization within 30 days of an index admission. The readmission is defined as a subsequent inpatient admission to the reporting facility, which occurs within 30 days of the discharge date of an eligible index admission.
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. In 2014, the Alliance of Dedicated Cancer Centers, or ADCC (an organization of eleven comprehensive cancer centers that are reimbursed differently by Medicare), and the Comprehensive Cancer Center Consortium for Quality Improvement, or C4QI (a group of nineteen academic medical centers that collaborate to measure and improve the quality of cancer in their centers), began developing a cancer-specific unplanned readmissions measure. This measure is designed to reflect the unique clinical aspects of oncology patients and to yield readmission rates that more accurately reflect the quality of care that may be obfuscated by a broader readmission measure, such as the CMS Hospital-Wide All-Cause Readmission measure (HWR #1789).

S.4. **Numerator Statement:** This outcome measure demonstrates the rate at which adult cancer patients (=18 years old at the index admission) are readmitted to a PPS-exempt Cancer Hospital (PCH) within 30 days of discharge from an index admission at the same PCH. The numerator includes all eligible patients with a readmission to a PCH within 30 days of the discharge date from an index admission with an admission status of urgent or emergency
S.7. **Denominator Statement:** All adult inpatient admissions with a diagnosis of malignant cancer at PCHs over the defined measurement period. The outcome measure examines the rate of unplanned readmissions within 30 days of discharge of this population.
S.10. **Denominator Exclusions:** The following patients are excluded from the denominator population: 1) patients transferred to another acute care facility during the index admission; 2) having missing or incomplete data; 3) admitted to an inpatient hospice bed; and, 4) discharged Against Medical Device (AMA).

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Administrative claims
S.26. **Level of Analysis:** Facility

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

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

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

De.4. **IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** N/A

New Measure -- Preliminary Analysis

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:
  - "Cancer is the second leading cause of death in the United States, with nearly 600,000 cancer-related deaths expected this year. It is estimated that more than 1.7 million Americans will be diagnosed with cancer in 2015, and nearly 14.5 million Americans with a history of cancer were alive in 2014."
  - While there are readmission measures for pneumonia, AMI, and HF, cancer has lagged behind in the development of readmission rates. Thus, the ADCC developed this measure.
  - 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).

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

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 developer provides performance data from alpha testing conducted during March 2013 - August 2014, and beta testing conducted between October 2014 – February 2015.
- The developer provides a range of unadjusted readmission rates from 14.5-15.8.

Disparities

- The developer states that "in an effort to account for socioeconomic risks associated with readmission, the risk adjustment approach utilized payer status as a proxy for this risk factor. As administrative claims data are used in this approach, payer status appears to offer a more refined level of adjustment than using a three digit zip code as a proxy for this element. Of note, race and gender were evaluated as potential risk factors for this measure, but were not determined to be statistically-significant for this measure."

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

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)

1. Importance to Measure and Report

1a. Evidence to Support Measure Focus

Comments: **Outcome measure: evidence for multiple interventions/processes that potentially improve readmission rates.

1b. Performance Gap

Comments: **Yes. Developers report unadjusted readmission rates from 14.5-15.4% across the PCHs

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

Comments: **N/A

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

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

- This measure calculates the 30-day unplanned rehospitalization rate for adult cancer patients.
- The Numerator includes all eligible patients with a readmission to a PCH within 30 days of the discharge date from an index admission with an admission status of urgent or emergency.
- The Denominator is all adult inpatient admissions with a diagnosis of malignant cancer at PCHs over the defined measurement period.
- The denominator population is defined using ICD-9 and ICD-10 codes; a list of applicable codes is included in the submission.
- The measure utilizes inpatient administrative claims data submitted on the UB-04 Uniform Bill, developed by CMS.
- The measure's time window is one year.
- The measure is risk-adjusted using a statistical risk model (see details below).

Questions for the Committee:

- *Does the Committee agree that the readmission should be limited to the facility from which the patient was discharged?*
- *Does the Committee agree with the way the numerator population is defined?*
- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *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**

- The developer has assessed reliability at the data element level.
- The reliability of the measure was testing by comparing the level of agreement with the planned/unplanned indicator based on the sample chart review.
- A Kappa score was calculated for the overall agreement of the two measures and the facility-level agreement
- Inter-rater reliability analyses (Kappa) were performed to determine consistency between Planned/Unplanned readmission type and inclusion in the measure numerator for individual participating facilities. Kappa scores ranged from 0.080 to 1.000 with asymptotic standard error ranging from 0.000 to 0.113.
- The developer notes that a moderate level of agreement (0.772) resulted when Kappa scores across the ten participating facilities were averaged. However, while seven out of the ten participating facilities have Kappa scores above 0.800, three centers had scores ranging from 0.080 to 0.690. This large variability requires further investigation to identify the sources of discrepancy. Variation in applied definitions of "planned" and/or "unplanned" readmissions is one explanation for the widespread Kappa scores. A second source of variation may be

the internal facility's guidelines for determining the type of admission. Third, some variation may be due to numerator exclusion criteria (i.e., admissions with a primary diagnosis of chemotherapy or radiation therapy encounter or progression of disease).

- Admission status variables for the 30-Day Unplanned Readmissions for Cancer Patients measure are defined using the Type of Admission/Visit on the CMS UB-04 Uniform Bill. The following definitions were applied:
- Emergency (code 1): The patient required immediate medical intervention as a result of severe, life threatening or potentially disabling conditions. Generally, the patient was admitted through the emergency room.
- Urgent (code 2): The patient required immediate attention for the care and treatment of a physical or mental disorder. Generally, the patient was admitted to the first available, suitable accommodation.
- Elective (code 3): The patient's condition permitted adequate time to schedule the availability of a suitable accommodation.

- Planed readmissions were defined as follows:
  - Planned readmissions are those within 30 days of discharge from an acute care hospital that are a scheduled part of the patient's plan of care. Planned readmissions are not counted as outcomes in this measure.
  - Unplanned readmissions are defined as an acute clinical event experienced by a patient that requires urgent re-hospitalization. Unplanned readmissions are counted as outcomes in this measure. For the purpose of the 30-Day Unplanned Readmissions for Cancer Patients measure, unplanned readmissions include those with an "emergency" or "urgent" Type of Admission/Visit.

Guidance from the Reliability Algorithm

- Question 1. Submitted specifications are precise, unambiguous, and complete. Measure can be consistently implemented.
- Question 2. Empirical reliability testing was conducted using statistical tests with the measure as specified.
- Question 3. Empirical validity testing of patient-level data was conducted.
- Question 4. Reliability testing was not conducted with computed performance measure scores for each measured entity.
- Question 8. Reliability testing was conducted with patient-level data elements that are used to construct the measure.
- Question 9. A Kappa score was an appropriate method for assessing the reliability of critical data elements.
- Question 10. A Kappa score of 0.772 resulted when scores from participating facilities were averaged. This is considered a moderate level of agreement.

Questions for the Committee:.

- *Is the test sample adequate to generalize for widespread implementation?*
- *Is the definition of planned/unplanned readmissions sufficiently previse to support widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*
- *Does the Committee agree the definition of planned vs. unplanned is precise and can be consistently implemented?*

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

2b. Validity

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**

Specification not completely consistent with evidence

- This measure calculates the 30-day unplanned rehospitalization rate for adult cancer patients.
- As a rationale for measuring this health outcome, the developers note the unique clinical aspects of oncology patients and that readmission rates for these patients may be obscured by a broader readmission measure, such as the Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

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.

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**

- The developer conducted validity testing of the critical data elements and assessed face validity of the measure score.
- The developer notes that criterion-related validity demonstrates if the new measure specifications actually measure true 30-day unplanned readmissions for cancer patients. This was demonstrated after alpha testing, using CY2012 data for participating facilities to demonstrate the appropriateness of the denominator population and numerator inclusion and exclusion criteria.
- 90.9% of unplanned readmissions that were reviewed during beta testing were included in the new 30-Day Unplanned Readmission for Cancer Patients measure. Moreover, 86.2% of planned readmissions were not included, indicating that the new measure is accurately capturing 30-day unplanned readmissions in cancer patients
- Cross-tabulation—a sensitivity of 0.879 and a specificity of 0.896 were found when comparing Planned/Unplanned readmission assignment with numerator inclusion criteria, using a sample size of 1,235 patients across all ten participating facilities.
- The developer interprets this as beta results showed that the specified measure is valid, generating global sensitivity and specificity scores of 0.879 and 0.896, respectively.

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 following patients are excluded from the denominator population:
  1) patients transferred to another acute care facility during the index admission;
  2) having missing or incomplete data;
  3) admitted to an inpatient hospice bed; and,
  4) discharged Against Medical Device (AMA).
- The developer notes that exclusions for this measure were determined by the measure development work group as the measure was being considered and developed.
- The developer notes that the inclusion of these patients would substantially skew the results of the measure.

Questions for the Committee:

- *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: **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

- Potential risk adjustors for 30-Day Unplanned Readmissions for Cancer Patients were identified by the following methods:
  - Review of the literature to determine which patient-level risk adjustors were included in risk- adjusted NQF-endorsed and CMS measures
  - Convening a group of cancer-hospital analytics experts with experience in creating readmission prediction models; and,
  - Convening a group of physician subject-matter experts from the cancer hospitals
- The developer notes that since there are only eleven cancer hospitals, it was not practical to include hospital-level adjustors, such as hospital size or teaching status, in the model.
- This measure uses logistic regression to estimate the probably of an unplanned readmission.
- The probability of readmission was then summed over the index admissions for each hospital to calculate the expected readmission rate.

- **Conceptual analysis of the need for SDS adjustment:**
  - The developer notes that since this measure is to be implemented based on claims data, many socioeconomic variables that may impact the likelihood of readmission were not included in this model due to issues of data availablity. 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.
    - In an effort to account for socioeconomic risks associated with readmission, the risk adjustment approach utilized payer status as a proxy for this risk factor. As administrative claims data are used in this approach, payer status appears to offer a more refined level of adjustment than using a three digit zip code as a proxy for this element.
- **Empirical analysis of SDS factors:**
  - The developer used payer status as a proxy for low socioeconomic stats. Low SES was defined as a primary payer of Medicaid, Charity, or Self-Pay Uninsured.
  - Additionally, race and gender were evaluated as potential risk factors for this measure, but were not determined to be statistically-significant for this measure.
- **Risk Model Diagnostics:**
  - To assess the overall performance of their risk-adjustment model, the developers computed several summary statistics, including c-statistic, ROC, Hosmer and Lemeshow Goodness of Fit Test, the likelihood ratio test and Akaike Information Criterion (AIC).
  - The c-statistic for the modeling data set was 0.6572 and 0.6554 for the validation dataset. The Brier Score was 0.1271 for the modeling data set and 0.1253 for the validation dataset.
  - The developer suggests the stability of these fit statistics supports the validity of the model.

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.*
- *Does the Committee agree that there a sufficient conceptual rationale for selecting risk adjustment factors provided by the develoeprs?*

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

- The developer calculated the expected admission rate for each hospital. Those values were then compared to the observed readmission rate. The aggregate rate for all six hospitals was used as the "standard" rate. 95% confidence intervals were then calculated to determine if the risk-adjusted rate for each hospital was different from the standard rate.
- The developers suggest the results indicate this risk adjustment approach will demonstrate statistically meaningful differences among facilities. Additionally, the developer notes this approach creates a standardized benchmark measure for expected unplanned readmissions among cancer patients for comparison across all PCHs.

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

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)

2. Scientific Acceptability of Measure Properties

2a1. & 2b1. Specifications

Comments: **Specs clear

2a2. Reliability Testing

Comments: **Interrater reliability testing (kappa) for determination of planned vs unplanned admissions—the averaged kappa was moderate--0.772 across 10 PCHs. However, there was considerable variability ranging from 0.08 to 1.0. Developers commented that this may be related to facility specific policies re: characterization of admissions as urgent/emergent/elective. Additionally, ?human error/subjectivity in completion of UB04s?

2b2. Validity Testing

Comments: **Face validity: Developers describe criterion based validity testing. Developers also report a sensitivity of 0.879 and a specificity of 0.896 (compared planned/unplanned admissions included in the numerator for a sample of 1,235 patients. There are only 11 PCHs, and not all reported during the development period, and there are only 11 months of claims data.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **Exclusions: do not report on testing. Just as a point of clarification, are pts who are TRANSFERRED to another hospital (and return) excluded or only those who are discharged to another hospital? On face, the attribuion of the readmission would be problemation (as would the imputation of the performance issue) but if a patient were transferred for a service not available at the

PCH and then returned, then perhaps should still be included in the denominator
risk adjustment: Strategy: potential factors appear to have been selected through consensus, prior use in readmissions measures and availability in claims data. In general only patient specific factors were included that were available from claims data with an evaluation of the effect (+/-) on readmission rates compared with unadjusted rates Because of limitations of claims data, decided to use payer source as a marker for low SES--conceptual rationale for SDS adjustments not included. c statistic reported for development and validation sets 0.6572 and 0.6554 respectively. Brier score was provided as a measure of calibration--similar on validation and development sets.
do not see a comparison with/without SDS (actually payer source as a marker for Low SES).
Several of the risk adjustment variables were not present at admission--discharge to hospice, surgical drg, LOS. Rationale for inclusion--these factors had an effect on readmission rate and arguably were patient specific factors. LOS may not be a true pt specific factor: might reflect pt acuity/complexity and/or issues with hospital efficiency/quality of processes of care.
Unclear that their analyses confirm that this measures meaningful differences in quality.
No identified missing data.

Criterion 3. Feasibility

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.

The developer states:

- This measure is based on administrative claims data (e.g., DRG, ICD-9/10), which the developers note are routinely generated and collected as part of hospitals' billing processes.
- The developer indicates that all data elements are in defined fields in electronic claims.
- To minimize differences in data capture across participating centers, data element definitions were aligned through the use of data dictionaries, where available (American College of Surgeons NCDB PUF: http://ncdbpuf.facs.org/?q=print-pdf-all; and, CMS Research Data Distribution Center LDS Inpatient SNF Claim Record Data Dictionary, Version November 2009: https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets/downloads/SAFldsSNFNov2009.pdf).
- Some data fields proposed for collection during beta testing presented a greater challenge for centers to obtain such as routine collection of variables from clinical trials, record of appointments regardless of setting, among others.
- Beta testing highlighted some discrepancies in "emergency," "urgent," and "elective" Type of Admission/Visit reported via administrative claims data. Preliminary data collected by measure developer suggests some improvement, but will require formal investigation and testing. Measure developer plans to test during reporting year 2016.
- The measure developer plans to focus on automating risk adjustment in 2016, to account for facilities with higher proportions of patients with hematologic cancers or lower socioeconomic status.

Questions for the Committee:

- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3. Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **Administrative claims data, so main body of data is electronically available/feasible. Main exception likely to be appropriate assignment of type of admission (urgent/emergent/elective) as this impacts assignment of case to numerator. Developers acknowledge this as a concern that they will further formally test in 2016.

Criterion 4: Usability and Use

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**

OR

Planned use in an accountability program? ☐ **Yes** ☐ **No**

Accountability program details

- Future plans for Public reporting
- Accountability programs:
  - UHC, City of Hope, Sylvester Comprehensive Cancer Care, Seattle Cancer Care Alliance
  - The developer provides links to the websites listed above.
- This measure was on the 2013/2014 MUC list and received conditional support from the MAP pending NQF endorsement. The developer expects the measure to be included in future rule-making; potentially as early as the FY 2017 Hospital Inpatient PPS rule making.

Improvement results N/A

Potential harms N/A

Feedback :

- For the 2014 pre-rulemaking, MAP gave this measure conditional support pending NQF endorsement.

***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

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **Not currently being used for public reporting. Currently being used for accountability programs at several PCHs

Criterion 5: Related and Competing Measures

Related or competing measures

- 1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

Harmonization

Not completely harmonized but the developer provides the following rationale:

- "Where possible, the 30-Day Unplanned Readmissions for Cancer Patients measure was harmonized with the HWR measure. For example, the 30-Day Unplanned Readmissions for Cancer Patients measure excludes from the denominator index admissions where the patient is transferred to another facility or where the patient is discharged AMA. Two important distinctions—
    - The HWR measure captures readmissions at any acute care facility, regardless of where the index admission occurred. Conversely, the 30-Day Unplanned Readmissions for Cancer Patients measure captures readmissions at the reporting facility only. Therefore, the 30-Day Unplanned Readmissions for Cancer Patients measure will not capture patients discharged from a PCH and readmitted within 30 days at another facility.
    - The HWR measure defines unplanned readmissions based on procedure codes and discharge diagnosis categories. However, the 30-Day Unplanned Readmissions for Cancer Patients measure utilizes the Type of Admission/Visit submitted on the UB-04 Uniform Bill, coded based on provider notes in the medical record since readmissions in cancer patients are not always predictable based on procedure code and discharge diagnosis.

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: 1/29/2016

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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: Click here to name the health outcome

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

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 estimated that more than 1.7 million Americans will be diagnosed with cancer in 2015, and nearly 14.5 million Americans with a history of cancer were alive in 2014. Cancer disproportionately affects older Americans, with 78% of all cancers diagnosed in people 55 years of age and older.[2] 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] Accordingly, 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 Alliance of Dedicated Cancer Centers, or ADCC, and the Comprehensive Cancer Center Consortium for Quality Improvement, or C4QI, (together, Measure Developers) began development of a cancer-specific unplanned readmissions measure: *30-Day Unplanned Readmissions for Cancer Patients*. The ADCC is an organization of eleven comprehensive cancer centers that are reimbursed differently by Medicare, and C4QI is a group of eighteen academic medical centers that collaborate to measure and improve the quality of cancer in their centers. 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. 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 care delivery that may be obscured by a broader readmission measure, such as the *Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)*, stewarded by CMS.

REFERENCES

1. Centers for Disease Control (CDC). Detailed Tables for the National Vital Statistics Report (NVSR) "*Deaths: Final Data for 2013.*"; Available at: https:// http://www.cdc.gov/nchs/data/nvsr/nvsr64/nvsr64_02.pdf. Accessed December 3, 2015.

2. American Cancer Society. Cancer Facts & Figures 2015. 2015. Available at: http://www.cancer.org/acs/groups/content/@editorial/documents/document/acspc-044552.pdf.

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.* Jan 19 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.* Apr 24 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.* Apr 2 2009;360(14):1418-1428.

6. Measure Applications Partnership (MAP), National Quality Forum (NQF). Performance Measurement Coordination Strategy for PPS-Exempt Cancer Hospitals. 2012. Available at: http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=71217. Accessed August 15, 2012.

7. Spinks TE, Walters R, Feeley TW, et al. Improving Cancer Care through Public Reporting of Meaningful Quality Measures. *Health Aff (Millwood).* Apr 2011;30(4):664-672.

8. Albright HW, Moreno M, Feeley TW, et al. The implications of the 2010 Patient Protection and Affordable Care Act and the Health Care and Education Reconciliation Act on cancer care delivery. *Cancer.* Apr 15 2011;117(8):1564-1574.

9. Rochefort MM, Tomlinson JS. Unexpected readmissions after major cancer surgery: an evaluation of readmissions as a quality-of-care indicator. *Surgical oncology clinics of North America.* Jul 2012;21(3):397-405, viii.

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

11. Bejanyan N, Bolwell BJ, Lazaryan A, et al. Risk factors for 30-day hospital readmission following myeloablative allogeneic hematopoietic cell transplantation (allo-HCT). *Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation.* Jun 2012;18(6):874-880.

12. Horwitz L, Partovian C, Lin Z, et al. *Hospital-Wide All-Cause Unplanned Readmission Measure Final Technical Report.* New Haven, CT: Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation (YNHHSC/CORE), Prepared for Centers for Medicare & Medicaid Services (CMS); July 2012.

13. McHugh ML. Interrater reliability: the kappa statistic. *Biochemia medica.* 2012;22(3):276-282.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of 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

30-Day_Unplanned_Readmissions_for_Cancer_Patients_Evidence_Form-635896776786386904.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)*

For many cancer patients, readmission following hospitalization may be preventable and should be addressed to potentially lower costs and improve patient outcomes. In 2014, the Alliance of Dedicated Cancer Centers, or ADCC (an organization of eleven comprehensive cancer centers that are reimbursed differently by Medicare), and the Comprehensive Cancer Center Consortium for Quality Improvement, or C4QI (a group of nineteen academic medical centers that collaborate to measure and improve the quality of cancer in their centers), began developing a cancer-specific unplanned readmissions measure. This measure is designed to reflect the unique clinical aspects of oncology patients and to yield readmission rates that more accurately reflect the quality of care that may be obfuscated by a broader readmission measure, such as the CMS Hospital-Wide All-Cause Readmission measure (HWR #1789).

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.*

Alpha testing was conducted between March 2013 and August 2014, with five primary goals:

1) Confirm the validity of the measure;
2) Justify the need for the measure, through comparisons with existing readmission measures;
3) Confirm that the measure is precisely defined to yield more consistent calculations across cancer centers;
4) Demonstrate that the measure is not biased toward particular inpatient settings; and,
5) Refine the measure specifications for use in beta testing.

Alpha testing demonstrated that the measure meets these criteria and produces different rates when compared with the Hospital-Wide All-Cause Unplanned Readmission Measure (#1789), or HWR, measure. In short, the new readmission measure produces a lower 30-day unplanned cancer specific readmissions rate, when compared with the HWR measure.

	Hospital-Wide All-Cause Unplanned Readmission Measure (HWR) Measure			30-Day Unplanned Readmissions for Cancer Patients Measure-Unadjusted Rates		
	Denominator	Numerator	%30-Day Readmission Rate	Denominator	Numerator	%30-Day Readmission Rate
ADCC	65,089	19,288	29.6%	65,089	8,637	13.3%
C4QI	108,562	31,214	28.8%	108,562	14,573	13.4%
Non-C4QI	4,205,079	533,818	12.9%	464,902	60,270	13.0%

Comparison of HWR Measure and 30-Day Unplanned Readmissions for Cancer Patients Measure Results—shows the results of the HWR measure and 30-Day Unplanned Readmissions for Cancer Patients measure, when applied to CY2012 index admissions for three groups: 1) eight participating ADCC member institutions (PCHs); 2) fourteen participating C4QI member institutions (inclusive of the eight participating ADCC member institutions); and, 3) non-C4QI hospitals with data in the UHC CDB/RM. Data source: UHC CDB/RM.

Beta testing was conducted between October 2014 and February 2015 and evaluated the measure's scientific acceptability (validity and reliability), feasibility, and usability. Beta testing results showed that the specified measure was valid, generating global

sensitivity and specificity scores of 0.879 and 0.896, respectively. Testing also demonstrated high reliability for the measure, with a global Cohen´s Kappa coefficient of 0.772. Generally, the feasibility of data collection for the measure is good, with a small number of fields not readily available for data capture from an electronic health record (EHR), the UHC CDB/RM, or other source of administrative data. The usability of the measure was also confirmed, as the measure is applicable for both accountability and performance improvement purposes to reduce unplanned readmissions and to improve patient experience.

Below are the results of data collection for the measure currently embedded in the UHC CDB/RM. The measure is in current use with participating sites monitoring the results for quality and performance improvement efforts. The data have not been risk-adjusted.

30-Day Unplanned Readmissions for Cancer Patients Measure- Unadjusted Rates
Jan 2014-Sept 2015

	Denominator	Numerator	%30-Day Readmission Rate
ADCC	113,263	16,471	14.5%
C4QI	186,322	28,787	15.4%
Non-C4QI	764,368	120,920	15.8%

Updated results for the 30-Day Unplanned Readmissions for Cancer Patients Measure Results—shows the results of the HWR measure and 30-Day Unplanned Readmissions for Cancer Patients measure, when applied to 1Q CY2014-3Q CY2015 index admissions for three groups: 1) ten participating ADCC member institutions (PCHs); 2) seventeen participating C4QI member institutions (inclusive of the ten participating ADCC member institutions); and, 3) non-C4QI hospitals with data in the UHC CDB/RM. Data source: UHC CDB/RM.

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 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.*
In an effort to account for socioeconomic risks associated with readmission, the risk adjustment approach utilized payer status as a proxy for this risk factor. As administrative claims data are used in this approach, payer status appears to offer a more refined level of adjustment than using a three digit zip code as a proxy for this element. Of note, race and gender were evaluated as potential risk factors for this measure, but were not determined to be statistically-significant for this measure.

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.

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, 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.
Cancer is the second leading cause of death in the United States, with nearly 600,000 cancer-related deaths expected in 2015. It was estimated that more than 1.7 million Americans would be diagnosed with cancer in 2015, and nearly 14.5 million Americans with a history of cancer were alive in 2014. Cancer disproportionately affects older Americans, with 78% of all cancers diagnosed in people

55 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 2020.2 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.3 Jencks, et al. estimated that unplanned readmissions cost the Medicare program $17.4 billion in 2004.4 Accordingly, all-cause and disease-specific unplanned readmissions rates have been adopted by 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 Alliance of Dedicated Cancer Centers, or ADCC, and the Comprehensive Cancer Center Consortium for Quality Improvement, or C4QI, (together, Measure Developers) began development of a cancer-specific unplanned readmissions measure: 30-Day Unplanned Readmissions for Cancer Patients. The ADCC is an organization of eleven comprehensive cancer centers that are reimbursed differently by Medicare, and C4QI is a group of nineteen academic medical centers that collaborate to measure and improve the quality of cancer in their centers. 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. 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 care delivery that may be obscured by a broader readmission measure, such as the Hospital-Wide All-Cause Unplanned Readmission Measure (HWR), stewarded by CMS.

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

1. American Cancer Society. Cancer Facts & Figures 2015. 2015. Available at: http://www.cancer.org/acs/groups/content/@editorial/documents/document/acspc-044552.pdf.
2. 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. Jan 19 2011;103(2):117-128.
3. Benbassat J, Taragin M. Hospital readmissions as a measure of quality of health care: advantages and limitations. Arch Intern Med. Apr 24 2000;160(8):1074-1081.
4. Jencks SF, Williams MV, Coleman EA. Rehospitalizations among Patients in the Medicare Fee-for-Service Program. N Engl J Med. April 2, 2009; 360:1418-1428.

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.)*

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):*
Cancer, Cancer : Bladder, Cancer : Breast, Cancer : Colorectal, Cancer : Gynecologic, Cancer : Hematologic, Cancer : Liver, Cancer : Lung, Esophageal, Cancer : Pancreatic, Cancer : Prostate, Cancer : Screening, Cancer : Skin

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Care Coordination : Readmissions, Patient and Family Engagement, Safety, Safety : Complications, Safety : Readmissions

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:** 30-Day_Unplanned_Readmissions_for_Cancer_Patients_Data_Dictionary.xls

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.
This outcome measure demonstrates the rate at which adult cancer patients (=18 years old at the index admission) are readmitted to a PPS-exempt Cancer Hospital (PCH) within 30 days of discharge from an index admission at the same PCH. The numerator includes all eligible patients with a readmission to a PCH within 30 days of the discharge date from an index admission with an admission status of urgent or emergency

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.)*
The time period covers all inpatient discharges from an index admission over an entire calendar year at PCHs (eg. Jan 01, 2014-Dec 31, 2014). The readmission time frame covers 13 months (calendar year with a 1 month tail) to capture those readmissions which may occur within a 30 day time frame from a discharge on Dec. 31.

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.
From the denominator population, 1) all patients with a readmission = 30 days from the discharge date of the index admission; and 2) readmitted with an admission type of ´urgent´ or ´emergency´ on the UB-04 Uniform Bill are included in the numerator population. All patients with a readmission = 30 days from the discharge date of the index admission who were readmitted with a principal diagnosis of ´chemotherapy encounter´ or ´radiation encounter´ are excluded. All patients with a primary diagnosis of metastatic cancer (ICD-10-CM codes: C77.0-C79.9, C7B.0-C7B.8/ICD-9-CM codes: 196.0-198.89, 209.7-209.79) are excluded. The outcome of interest is the rate of = 30 day unplanned readmissions for cancer patients over a specified time frame.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
All adult inpatient admissions with a diagnosis of malignant cancer at PCHs over the defined measurement period. The outcome measure examines the rate of unplanned readmissions within 30 days of discharge of this population.

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

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 denominator population is defined as: 1) All inpatient admissions with a diagnosis of malignant cancer (ICD-10-CM range: C00.0-C96.9, D37.01-D49.9/ICD-9-CM range: 140.00-209.99); 2) = 18 years of age; and 3) a discharge status of ´alive´ for index admission

during the measurement period.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
The following patients are excluded from the denominator population: 1) patients transferred to another acute care facility during the index admission; 2) having missing or incomplete data; 3) admitted to an inpatient hospice bed; and, 4) discharged Against Medical Device (AMA).

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)*
The following index admissions are excluded from the measure denominator:

1. Admissions with missing (incomplete/inaccurate) data;
2. Patients who expired within the term of the index admission;
3. Patients with an admission to an inpatient hospice bed; and,
4. Patients who were discharged against medical advice (AMA), because providers did not have the opportunity to deliver full care and prepare the patient for discharge.

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)*
A logistic regression was applied, using the following risk factors: 1) age less than 40; 2) discharge to hospice; 3) length of stay greater than 3 days; 4) low socioeconomic status; 5) multiple comorbidities; 6) solid tumor; and, 7) Surgical MS-DRG.

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*)

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.)*
Please also refer to the measure flow logic in the data dictionary in Appendix.

Step 1: Denominator population determination

All malignant cancer inpatient admissions within the reporting period at this facility (ICD-10-CM range: C00.0-C96.9, D37.01-D49.9/ICD-9 range:140.00-209.99)

- Patient = 18 years old
- Patient discharged 'alive'
- Without a transfer to another acute care facility
- No missing data
- Not admitted to an inpatient hospice bed
- Not discharged AMA

Step 2: Numerator population determination

From denominator population (Step 1)
- Patient readmitted = 30 days from discharge date of index admission at same facility
- Admit type 'urgent' or 'emergency' on UB-04 Uniform Bill
- Unplanned admission (excludes patients with a principal diagnosis of chemotherapy or radiation encounter)
- Disease has not progressed (excludes patients with a principal diagnosis of metastatic cancer—ICD-10-CM codes: C77.0-C79.9, C7B.0-C7B.8/ICD-9-CM codes 196.0-198.89, 209.7-209.79)

The outcome of interest is unplanned readmission for the target population.

For risk adjustment, n=57,945 index patient admission. After preparation and cleaning, the dataset was split into a model and validation set for modeling. The splitting of the data allows an assessment of the stability of the risk adjustment model. The SAS ranuni function with a seed of "926" was used to split the data. This resulted in an n=28, 978 for risk modeling set and n=28,967 for the validation set.

Through logistic regression modeling, the following 7 risk adjustment factors were found to be the best choices for risk adjustment of this population: 1) Low socioeconomic status (defined as primary payer of: Medicaid, Charity or Self-Pay Uninsured); 2) a surgical MS-DRG; 3) multiple comorbidities (defined as more than one comorbidity based on Elixhauser index (excludes Tumor and Mets comorbidity)); 4) solid tumor (defined as ICD-10-CM codes: C00.0-C80.0 (excluding metastatic/ ICD-9-CM Codes: 140-199 (excluding metastatic); 5) age less than 40; 6) length of stay greater than 3 days; and, 7) discharge to hospice. Note that discharge to a hospice, surgical admissions and solid tumor were all considered to be protective in terms of the likelihood of a readmission. The other included variables represented an increase in the odds of readmission for the patient.

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.
This outcome measure is based on the full population of eligible patients: sampling is not used.

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

S.22. Missing data (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
If patient is missing data, the patient is excluded from the denominator.

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

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.
The measure utilizes inpatient administrative claims data submitted on the UB-04 Uniform Bill, developed by CMS. For most

participating facilities, these data were obtained from the University HealthSystem Consortium (UHC) Clinical Data Base/Resource Manager (CDB/RM) v 1.5.0.10. The UHC CDB/RM is a comparative database containing inpatient and outpatient administrative claims data from more than 100 academic medical centers and other hospitals. Data from an electronic health record (EHR) or other medical record source were also utilized during beta testing. One to six abstractors at each of the participating facilities extracted data from these systems for purposes of beta testing.

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)*
Available in attached appendix at A.1

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/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

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
30-Day_Unplanned_Readmissions_for_Cancer_Patients_Testing_Form-635896777634878661.docx

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: 12/1/2015
Type of Measure:

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

1.3. What are the dates of the data used in testing? For Alpha testing, all adult patients with an index admission date, discharge date and readmission date between Jan 1, 2012-Dec 31, 2012. For the Beta testing phase, all adult patients with an index admission date, discharge date and readmission date between Jan 1, 2014-June 30, 2014.

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

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

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

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

For the alpha testing, a total of 8 Alliance of Dedicated Cancer Center (ADCC) members (which are all PPS-exempt cancer hospitals (PCH)), 14 Comprehensive Cancer Center Consortium for Quality Improvement (C4QI) members, the one National Cancer Institute (NCI) facility not a member of C4QI, and all 100 University HealthSystem Consortium (UHC) members had inpatient administrative claims data analyzed. This group represents a broad range of patient case mix and volume. The results were stratified by institution type: ADCC, C4QI and UHC (excluding C4QI). These institutions represent every locale in the contiguous United States. They are of varying size and varying patient volume.

Of note, all PCH hospitals are members of C4QI. Therefore, the C4QI group is inclusive of PCHs. Similarly, seventeen of the eighteen C4QI member institutions are designated as Cancer Centers by the National Cancer Institute (NCI), due to their research focus. The NCI-Designated Cancer Centers group is largely inclusive of all C4QI member institutions.

For the beta testing phase, a subset of 10 institutions participated; 8 from the ADCC group and 2 from the C4QI group. This phase required manual chart abstraction to confirm the precision of the definition of 'planned' vs. 'unplanned' readmissions.

For development of a risk adjustment methodology, a subset of 6 members of the ADCC were used for computational analysis. These 6 are representative of the group participating in alpha testing.

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*)

For the alpha testing phase:

N=464,902 inpatient admissions with a diagnosis of malignant cancer (ICD-10-CM range: C00.0-C96.9, D37.01-D49.9/ ICD-9-CM range: 140.00-209.99) across the participating institutions. No descriptive characteristics available.

For the beta testing phase:

N= 1,235 patients with an inpatient admission with a diagnosis of malignant cancer (ICD-10-CM range: C00.0-C96.9, D37.01-D49.9/ ICD-9-CM range: 140.00-209.99) and a readmission date within the defined measure period across the participating institutions

- 44% female; 56% male
- 71% < 65 years of age; 27.8% between ages of 65-84
- 54.6% had private insurance coverage; 30.9% covered by Medicare; 10.7% covered by Medicaid

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.

For alpha testing and risk adjustment, the data included all eligible inpatient admissions across the participating institutions during the study period. For beta testing, data sampling through manual chart abstraction across 10 representative institutions on eligible inpatient admissions was conducted. This approach was necessary to confirm the validity and reliability of the measure.

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).

Since this measure is to be implemented based on claims data, many socioeconomic variables that may impact the likelihood of readmission were not included in this model. _____

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*)

The reliability of the proposed unplanned readmission indicator (claims-based indicator) was tested by comparing the level of agreement with the planned/unplanned indicator based on the sample chart review. The following statistical test was performed:

a. A Kappa score was calculated for the overall agreement of the two measures and the facility-level agreement;

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*)

1) Inter-Rater Reliability (Kappa) by Individual Facility: inter-rater reliability analyses (Kappa) were performed to determine consistency between Planned/Unplanned readmission type and inclusion in the measure numerator for individual participating facilities. Kappa scores ranged from 0.080 to 1.000 with asymptotic standard error ranging from 0.000 to 0.113.

Inter-Rater Reliability (Kappa) by Individual Facility

Facility	Kappa	Asymptotic Error Rate	Approx. T[b]	P-value
1	0.878	0.039	10.812	<0.001
2	0.884	0.113	4.449	<0.001
3	0.080	0.074	1.150	.250
4	0.690	0.056	8.840	<0.001
5	0.849	0.044	10.541	<0.001
6	0.437	0.073	5.695	<0.001
7	0.950	0.035	9.466	<0.001

8	0.973	0.019	12.040	<0.001
9	0.966	0.024	11.843	<0.001
10	1.000	0.000	6.557	<0.001
Overall	0.772	0.018	27.162	<0.001

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?*)

Statistical analyses confirmed that the measure is well-defined and precisely specified so that it can be implemented consistently within and across organizations and allow for accurate comparisons to be made.

The results for all ten participating centers as well as individual facilities indicates that, while the measure is reliable, some variability from the average across all participating sites is present when examining scores at the individual facility level. This could be due to:

- Inconsistent application of the definitions of "planned" or "unplanned" among facilities;
- Inconsistent assignment of admission types (*emergency*, *urgent,* and/or *elective*);
- Anomalies in other numerator inclusion/exclusion criteria; and,
- Application of numerator exclusion criteria.

Interpretation

1) Kappa Scoring

A moderate level of agreement (0.772) resulted when Kappa scores across the ten participating facilities were averaged. However, while seven out of the ten participating facilities have Kappa scores above 0.800, three centers had scores ranging from 0.080 to 0.690. This large variability requires further investigation to identify the sources of discrepancy. Variation in applied definitions of "planned" and/or "unplanned" readmissions is one explanation for the widespread Kappa scores. A second source of variation may be the internal facility's guidelines for determining the type of admission. Third, some variation may be due to numerator exclusion criteria (i.e., admissions with a primary diagnosis of chemotherapy or radiation therapy encounter or progression of disease). Finally, one of the participating facilities does not utilize the UHC CDB/RM for patient records and the measure calculation. Their attempt to replicate the definition of the indicator, as defined within the UHC CDB/RM, is a potential source of variation in the Kappa score

---

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)*

Face validity evaluates whether the measure specifications and exclusions align with the existing evidence. Likewise, it evaluates the inclusivity of the measure vis-à-vis the target population. Finally, it evaluates whether the measure exclusions are needed to prevent distortion within the measure. Currently, there are only a few peer-reviewed publications that are specific to readmissions in cancer patients. Thus, the measure's face validity was confirmed through written and oral communications between the Measure Developers and clinicians across C4QI during alpha testing.

Criterion-related validity demonstrates if the new measure specifications actually measure true 30-day unplanned readmissions for cancer patients. This was demonstrated after alpha testing, using CY2012 data for participating facilities to demonstrate the appropriateness of the denominator population and numerator inclusion and exclusion criteria.

Admission status variables for the *30-Day Unplanned Readmissions for Cancer Patients* measure are defined using the Type of Admission/Visit on the CMS UB-04 Uniform Bill. The following definitions were applied:

- ***Emergency (code 1)****: The patient required immediate medical intervention as a result of severe, life threatening or potentially disabling conditions. Generally, the patient was admitted through the emergency room.*
- ***Urgent (code 2):*** *The patient required immediate attention for the care and treatment of a physical or mental disorder. Generally, the patient was admitted to the first available, suitable accommodation.*
- ***Elective (code 3)****: The patient's condition permitted adequate time to schedule the availability of a suitable accommodation.*

For the purpose of the measure, these categories were grouped into "planned" and "unplanned" admission visits:

- **Planned readmissions** are those within 30 days of discharge from an acute care hospital that are a scheduled part of the patient's plan of care. Planned readmissions are not counted as outcomes in this measure.

- **Unplanned readmissions** are defined as an acute clinical event experienced by a patient that requires urgent re-hospitalization. Unplanned readmissions are counted as outcomes in this measure. For the purpose of the *30-Day Unplanned Readmissions for Cancer Patients* measure, unplanned readmissions include those with an "emergency" or "urgent" Type of Admission/Visit.

90.9% of unplanned readmissions that were reviewed during beta testing were included in the new *30-Day Unplanned Readmission for Cancer Patients* measure. Moreover, 86.2% of planned readmissions were not included, indicating that the new measure is accurately capturing 30-day unplanned readmissions in cancer patients

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

Cross-Tabulation of Planned/Unplanned Readmissions and Inclusion in the New Readmission Measure

		Numerator Inclusion		Total
		No	Yes	
Planned	Count	508	59	567
	% of entries within Planned Readmissions	89.6%	10.4%	100.0%
	% included in the numerator of the measure	86.2%	9.1%	45.9%
Unplanned	Count	81	587	668
	% of entries within Unplanned Readmissions	12.1%	87.9%	100.0%
	% included in the numerator of the measure	13.8%	90.9%	54.1%
Total	Count	589	646	1235
	% of entries within Planned/Unplanned Readmissions	47.7%	52.3%	100.0%
	% included in the numerator of the measure	100.0%	100.0%	100.0%

Sensitivity:	0.879
Specificity:	0.896

Cross-tabulation—a sensitivity of 0.879 and a specificity of 0.896 were found when comparing Planned/Unplanned readmission assignment with numerator inclusion criteria, using a sample size of 1,235 patients across all ten participating facilities.

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?*)

For purposes of beta testing, participating facilities sampled patients with a hospital admission discharge within Q1 and Q2 of CY2014 who experienced at least one readmission (planned or unplanned) within 30 days of discharge of an eligible index admission. *30-Day Unplanned Readmissions for Cancer Patients* specifies that only unplanned readmissions within 30 days of discharge of an index admission are included in the numerator, as long as those cases do not qualify for the specified exclusion categories. However, by including both unplanned and planned readmission types in sampling for beta testing and analyses, sensitivity and specificity could be calculated to ascertain the accuracy and consistent application of "planned" and "unplanned" definitions at participating facilities. The number of "planned" readmissions that were not included in the measure numerator represents the specificity of the measure. The overall specificity of planned readmissions across all ten participating facilities was 0.896. Further examination of facility beta testing data submissions is required to identify reasons for numerator inclusion of 59 (10.4%) "planned" readmission cases. The number of "unplanned" readmissions that were included in the measure numerator represents the sensitivity of the measure or percentage of "true positive" values. The sensitivity of the measure across the ten participating facilities is 0.879. The inclusion of 10.4% of planned readmission cases in the numerator as well as the exclusion of 12.1% of unplanned readmission cases from the numerator will be investigated further in CY2015 to determine if this

is an issue with the reliability of the underlying claims data, an issue with the measure specifications, as written, or application of the measure's exclusion criteria.

Beta testing results showed that the specified measure is valid, generating global sensitivity and specificity scores of 0.879 and 0.896, respectively. Testing also demonstrated high reliability for the measure, with a global Cohen's Kappa coefficient of 0.772. Generally, the feasibility of data collection for the measure is good, with a small number of fields not readily available for data capture from an electronic health record (EHR), the University HealthSystem Consortium (UHC) Clinical Data Base/Resource Manager (CDB/RM), or other source of administrative data. The usability of the measure is also confirmed, as the measure is applicable for both accountability and performance improvement purposes to reduce unplanned readmissions and to improve patient experience.

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 for this measure were determined by the measure development work group as the measure was being considered and developed. The measure is intended to identify cancer patients with unplanned readmissions. This is a differentiation from the *Hospital-Wide All Cause Unplanned Readmissions* (HWR) stewarded by CMS. Thus the exclusion of all admissions without a malignant cancer diagnosis (ICD-10-CM range: C00.0-C96.9, D37.01-D49.9/ICD-9-CM range: 140.00-209.99) to capture the appropriate population in the measure was instituted. Other exclusions include: pediatric patients, patients who die during index admission, patients admitted to a hospice bed, patients discharged to another acute care facility or discharged against medical advice (AMA).

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*)
These exclusions are performed in the denominator collection process and during validation to confirm the patient meets all eligibility parameters.

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 rationale of the measure is to collect unplanned readmission information specifically on cancer patients. Inclusion of any patient outside the target population would 1) substantially skew results and 2) be redundant to the HWR cited above.

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 7 potential risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

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

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*)

Potential risk adjustors for *30-Day Unplanned Readmissions for Cancer Patients* were identified by the following methods:

1. Review of the literature to determine which patient-level risk adjustors were included in risk- adjusted NQF-endorsed and CMS measures
2. Convening a group of cancer-hospital analytics experts with experience in creating readmission prediction models; and,
3. Convening a group of physician subject-matter experts from the cancer hospitals

This effort resulted in a listing of potential patient-level variables for inclusion in the risk adjustment model (see list below). Since there are only eleven cancer hospitals, it was not practical to include hospital-level adjustors, such as hospital size or teaching status, in the model. The listing of potential risk adjustors was then compared to the data elements available from administrative claims data. Since this measure is to be implemented based on claims data, many socioeconomic variables that may impact the likelihood of readmission were not included in this model. 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.

Included in Modeling
Surgical vs. non-Surgical admissions
ICU vs. non-ICU admissions
LOS
Admission from Emergency Department (vs. not admitted from Emergency Department)
Age
Gender
Payer (proxy for socioeconomic status)
Discharge disposition (home or hospice vs. other)
Tumor type (based on claims data)
Count of Comorbidity via Elixhauser Comorbidity Index
Race
Metastatic disease

Not Available in Data Used for Modeling
Marital status
Prior hospitalizations (only cancer-related admissions available)
Distance patient lives from the hospital

Not Well-Defined in Claims Data
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
BMI
Under the Hospital's Control
Discharge on a weekday vs. weekend

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

The risk factors chosen for analysis are shown in the table below along with the results.

Observed Readmission Rates by Potential Risk Adjustors

Variable	Readmission Rate		Included in Model?
	Present	Absent	
Age Less Than 40	20.3%	14.8%	Yes
Discharge to Hospice	2.8%	16.1%	Yes
Length of Stay Greater than 3 days	17.8%	11.6%	Yes
Low Socioeconomic Status	20.6%	15.0%	Yes
Multiple Comorbidities	18.2%	13.7%	Yes
Solid Tumor	10.8%	17.3%	Yes
Surgical MS-DRG	9.9%	19.1%	Yes
ICU Utilization	14.4%	15.6%	No (p >0.05)
Male	15.4%	15.6%	No (p >0.05)
White	15.1%	15.8%	No (p >0.05)
Metastatic Disease	14.1%	16.5%	No (inconsistent relationship)
Emergency Room Patient	18.7%	14.7%	No (not in all providers)
Bone Marrow Transplant Status	18.1%	15.4%	No (not in all providers)

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)
This initial analysis demonstrated that risk adjustment is needed for the *30-Day Unplanned Readmissions for Cancer Patients* measure. Patient-specific demographic and clinical factors were identified as appropriate risk adjustment factors to support comparable comparisons across cancer centers. Because this initial model utilized approximately eleven months of claims data from six centers, the actual variables and coefficients may not be appropriate for application to the full complement of PCHs. It is expected that the actual variables included in the model and the coefficients will change when all eleven cancer hospitals are submitting data and over time. Nonetheless, this analysis allowed the ADCC to propose a valid risk adjustment methodology and risk adjusters for this measure.

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 most common type of risk adjustment used in healthcare outcomes measures is based on the observed rate of the outcome compared to the expected rate. The expected rate is based on patient and, sometimes, hospital demographics. If the observed rate is much higher than the expected rate, then the performance of the hospital would be considered poor. Conversely, if the observed rate is much lower than the expected rate, then the

performance of the hospital would be better than expected. A risk-adjusted rate may also be formulated for comparing hospital performance as presented in **Figure 2**.

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

Figure 2. Risk-Adjusted Rate Formula

Alternative methodologies include indirect or direct standardization and hierarchical modeling. Applying indirect or direct standardization to models that include multiple risk factors is cumbersome. Standardization techniques are typically applied in creating age-adjusted rates in epidemiological studies. Hierarchical modeling is applied when there are multiple levels of risk adjustors. For instance, hierarchical modeling may be used to adjust for variables at the hospital level (e.g., teaching hospital, specialization, etc.) and patient level (e.g., race, payer, diagnosis, etc.). Although some hospital-level variables may be of interest in risk-adjusting readmission rates, the proposed cancer measure will only be applied to eleven hospitals at this point. Therefore, it would not be possible to estimate hospital level adjusters with an acceptable level of precision.

Statistical modeling is typically used to estimate the expected rate for a hospital. In this case, logistic regression was used to estimate the probability of an unplanned readmission that meets the proposed definition outlined in this report based on the risk factors included in the model. The probability of readmission was then summed over the index admissions for each hospital to calculate the expected readmission rate. Prior to fitting the logistic regression model, the dataset was randomly divided into a fit (i.e., model set) and validation set. This strategy allowed the fitting of the model to be tested for robustness or generalization by comparing the c-statistic resulting from using the fit model on the validation dataset. A similar value for the c-statistic (or Area Under the Curve) was indicative of a stable model that may be used for risk adjustment.

The logistic model was fit using SAS/STAT software, Version 9.4 (SAS Institute, Inc. 2015) using the 'stepwise' option and maximum likelihood estimation (MLE). Prior to inclusion in the model, the potential association between the various risk adjustors was assessed by calculating the tetrachoric correlation. Variables 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 (highly-correlated risk factors). No variables were excluded due to multicollinearity. The logistic model diagnostics, such as the c-statistic, ROC, Hosmer and Lemeshow Goodness of Fit Test, the likelihood ratio test and Akaike Information Criterion (AIC), were all collected and analyzed prior to selecting the risk adjustment model. (Bewick V 2005)

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*)**:**
The c-statistic for the modeling data set was 0.6572 and 0.6554 for the validation dataset. The Brier Score was 0.1271 for the modeling data set and 0.1253 for the validation dataset. The stability of these fit statistics supports the validity of the model.

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

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

2b4.9. Results of Risk Stratification Analysis:

Logistic Model Coefficients and Odds Ratios

Parameter	Model Coefficients		Odds Ratio Estimates		
	Estimate	P-value	Point Estimate	95% Wald Confidence Limits	
Intercept	-2.7774	<.0001			
Low Socioeconomic Status	0.108	0.0001	1.241	1.112	1.385
Surgical MS-DRG	-0.3521	<.0001	0.495	0.459	0.533
Multiple Comorbidities	0.1211	<.0001	1.274	1.192	1.362
Solid Tumor	-0.1979	<.0001	0.673	0.620	0.731
Age Less Than 40	0.1264	<.0001	1.288	1.169	1.418
Length of Stay Greater than 3 days	0.2718	<.0001	1.722	1.603	1.850
Discharge to Hospice	-1.0511	<.0001	0.122	0.090	0.166

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*)
This initial analysis demonstrated that risk adjustment is needed for the *30-Day Unplanned Readmissions for Cancer Patients* measure. Patient-specific demographic and clinical factors were identified as appropriate risk adjustment factors to support comparable comparisons across cancer centers. The results demonstrate this risk adjustment strategy is appropriate to the level of population tested.

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)*

Using the Logistic model in in 2b.4.9, the expected readmission rate for each hospital was calculated. Those values were then compared to the observed readmission rate. The aggregate rate for all six hospitals was used as the "standard" rate. 95% confidence intervals were then calculated to determine if the risk-adjusted rate for each hospital was different from the standard rate

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*)

Hospital Level Results

Hospital	Observed Readmission Rate	Expected Readmission Rate	Standardized Readmission Rate (SRR) = obs/exp	Risk-Adjusted Readmission Rate (RARR)	95% Confidence Interval	Comparison to Standard (15.6%)
A	13.9%	16.2%	0.861	13.6%	(12.5%, 14.6%)	Better
B	14.0%	15.1%	0.929	14.6%	(14.1%, 15.2%)	Better
C	15.8%	15.5%	1.018	16.0%	(15.2%, 16.8%)	Not Different
D	10.9%	13.8%	0.787	12.4%	(11.2%, 13.6%)	Better
E	17.7%	16.3%	1.083	17.0%	(16.5%, 17.5%)	Worse
F	15.2%	16.4%	0.927	14.6%	(13.6%, 15.6%)	Not Different

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 results indicate this risk adjustment approach will demonstrate statistically meaningful differences among facilities. The approach creates a standardized benchmark measure for expected unplanned readmissions among cancer patients for comparison across all PCHs.

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)

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*)

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*)

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*)

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*)

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*)

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)*

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.**

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.

The data required to calculate the 30-Day Unplanned Readmissions for Cancer Patients measure using the current specifications are readily available using the UHC CDB/RM and other sources of administrative claims data. To minimize differences in data capture across participating centers, data element definitions were aligned through the use of data dictionaries, where available (American College of Surgeons NCDB PUF: http://ncdbpuf.facs.org/?q=print-pdf-all; and, CMS Research Data Distribution Center LDS Inpatient SNF Claim Record Data Dictionary, Version November 2009: https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets/downloads/SAFldsSNFNov2009.pdf).

Some data fields proposed for collection during beta testing presented a greater challenge for centers to obtain. Any modifications to the current measure specifications that would require routine collection of the following variables would create additional reporting burden:

- Documentation or record of patient enrollment in a clinical trial;
- Documentation or record of appointments (regardless of setting), which occur between the patient's index admission and readmission;
- Identification of the ICD-9-CM to ICD-10-CM coding "crosswalk" used at the facility;
- Consistent location of documentation/record of metastatic disease development; and,
- Independent confirmation of "emergency," "urgent," and "elective" Type of Admission/Visit on the UB-04 Uniform Bill

through manual assessment of admissions as "planned" or "unplanned."

Beta testing highlighted some discrepancies in "emergency," "urgent," and "elective" Type of Admission/Visit reported via administrative claims data. The Measure Developers performed a preliminary investigation into these discrepancies during CY2015 to identify potential contributing factors. It was expected that this might lead to practice or coding changes among PCHs and non-PCH member institutions of C4QI, which would, in turn, necessitate additional reliability testing for those sites. Preliminary data suggests some improvement, though more formal investigation and testing is required. This will be discussed in CY2016.

To maximize this measure's utility for purposes of benchmarking, the risk adjustment will need to be automated. This is particularly important for centers with higher proportions of patients with hematologic cancers or lower socioeconomic status. This will be an area of focus for the measure developers in 2016.

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.

Planned	Current Use (for current use provide URL)
Public Reporting	Quality Improvement with Benchmarking (external benchmarking to multiple organizations) UHC www.uhc.edu Quality Improvement (Internal to the specific organization) City of Hope http://www.cityofhope.org/homepage Sylvester Comprehensive Cancer Care http://sylvester.org/ Seattle Cancer Care Alliance http://www.seattlecca.org/

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

1) UHC: External Benchmarking

a. ADCC using as an active measure of ADCC member hospitals' performance on cancer specific readmission rates. Able to compare the cancer-specific rate to all hospitals' readmission rates

b. C4QI members using as an active measure of C4QI member hospitals' performance on cancer specific readmission rates. Able to compare the cancer-specific rate to all hospitals' readmission rates

c. Nationwide

2) City of Hope: Quality Improvement
a. Using as a monthly quality improvement report
b. Southern California (LA)- 1 facility

3) Sylvester Comprehensive Cancer Care
a. Using for care decisions in discharge planning
b. Southern Florida-1 facility

4) Seattle Cancer Care Alliance
a. Comparing all-cause readmission data to cancer-specific readmission data to demonstrate sensitivity of treating cancer patients as a separate category for systems reporting.
b. Pacific Northwest-5 Seattle area facilities

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 on the 2013/2014 MUC list and received conditional support from the MAP pending NQF endorsement. It is our expectation this measure will be included in future rule-making; potentially as early as the FY 2017 Hospital Inpatient PPS rule making.

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**

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.**

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.**
No unintended negative consequences were identified in the alpha and beta testing phase. This is a passive surveillance approach with no attached intervention.

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

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?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Where possible, the 30-Day Unplanned Readmissions for Cancer Patients measure was harmonized with the HWR measure. For example, the 30-Day Unplanned Readmissions for Cancer Patients measure excludes from the denominator index admissions where the patient is transferred to another facility or where the patient is discharged AMA. Two important distinctions—where harmonization was not possible—also exist•The first distinction relates to the facility where the patient was readmitted. The HWR measure captures readmissions at any acute care facility, regardless of where the index admission occurred. Conversely, the 30-Day Unplanned Readmissions for Cancer Patients measure captures readmissions at the reporting facility only. Therefore, the 30-Day Unplanned Readmissions for Cancer Patients measure will not capture patients discharged from a PCH and readmitted within 30 days at another facility. Because the PCHs do not have access to inpatient administrative claims data for other hospitals, the numerator was defined to include unplanned readmissions at the reporting facility (PCH) only. :• Another important distinction is how unplanned readmissions are defined within the measures. The HWR measure defines unplanned readmissions based on procedure codes and discharge diagnosis categories. However, the 30-Day Unplanned Readmissions for Cancer Patients measure utilizes the Type of Admission/Visit submitted on the UB-04 Uniform Bill. Readmissions in cancer patients are not always predictable based on procedure code and discharge diagnosis. Therefore, the Measure Developers chose to leverage the Type of Admission/Visit submitted on the UB-04 Uniform Bill, which is coded based on provider notes in the medical record.

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:** 30-Day_Unplanned_Readmissions_for_Cancer_Patients_Supplemental_Materials.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Seattle Cancer Care Alliance
Co.2 **Point of Contact:** Barbara, Jagels, bjagels@seattlecca.org, 206-288-2127-
Co.3 **Measure Developer if different from Measure Steward:** Alliance of Dedicated Cancer Centers (ADCC)
Co.4 **Point of Contact:** Terry, Fisher, tfisher3@mdanderson.org, 713-563-2694-

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.
Workgroup Member
Denise Morse, MBA
Laura Crocitto, MD, MHA
Merle Smith, RN, BSBA, MSN
Steve Flaherty, MPH
Jennifer Snide, MS
Steve Power, MBA
Audrey Holland, PhD
Joyce M. Kane, MSN, RN, CPHQ, RHIT, CTR
Ron Walters, MD, MBA, MS, MHA
Tracy Spinks, BBA
Colleen Tallant, MS
Doug Browning, MBA
Steve Martin, MD
Sarah Berger, MBA
Paul Hendrie, MD, PhD (Workgroup Chair)
Keith D. Eaton, MD, PhD
Barb Jagels, RN, MHA, CPHQ
Tracy Kusnir-Wong, MBA
Gloria (Gigi) Campos, , MSIE
Laurian Walters, BS
Carl R. Schmidt, MD
Joseph M. Flynn, MD
Kristen Johnson, MHA
Linda Lane, RHIA, CPHQ
Afsaneh Barzi, MD, PhD
Stephanie Buia Amport, MBA, CPHQ
Lisa Truini-Pittman, RN, MPH
Susan White, PhD, RHIA, CHDA Facility
City of Hope
City of Hope
City of Hope
Dana-Farber
Dartmouth
Duke
Duke
Sidney Kimmel Cancer Center at Johns Hopkins
MD Anderson
MD Anderson
MD Anderson
MD Anderson
MSK

MSK
Seattle Cancer Care Alliance
Seattle Cancer Care Alliance
Seattle Cancer Care Alliance
Seattle Cancer Care Alliance
Sylvester
Sylvester
The James
The James
The James
The James
USC Norris
Yale
Yale
Health Policy Analytics, LLC

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:** 06, 2015

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?** 06, 2016

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

Ad.7 **Disclaimers:** N/A

Ad.8 **Additional Information/Comments:**

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 #: 0171
De.2. **Measure Title:** Acute Care Hospitalization During the First 60 Days of Home Health
Co.1.1. **Measure Steward:** Centers for Medicare & Medicaid Services
De.3. **Brief Description of Measure:** Percentage of home health stays in which patients were admitted to an acute care hospital during the 60 days following the start of the home health stay.
1b.1. **Developer Rationale:** see attachment "Importance to Report" for a tabular presentation of these data

Hospital readmissions are a national priority for Medicare recipients, based on evidence that 20% of all Medicare beneficiaries who were hospitalized had a return hospital stay within 30 days. In 2004, this cost the Medicare program $17.4 billion (1). Within home health care, an analysis of Medicare claims shows that 14 percent of home health patients are rehospitalized within 30 days of the start of home health care. There is limited research on the extent to which these hospital readmissions are avoidable within home health care. One study reporting on patients with heart failure found that more than 40% of the 30 day rehospitalizations may have been avoidable (2). In addition, studies of Medicare patients in general provided evidence for interventions that reduce the need for hospital care within a substantial proportion of these Medicare beneficiaries (1;3). Moreover, there are a number of national initiatives, both governmental (e.g. Quality Improvement Organizations, National Priorities Partnership and CMS) and through private foundations (e.g. Institute for Healthcare Improvement), addressing this issue. Thus there is room for improvement and this is a national priority issue.

S.4. **Numerator Statement:** Number of home health stays for patients who have a Medicare claim for an unplanned admission to an acute care hospital in the 60 days following the start of the home health stay.
S.7. **Denominator Statement:** Number of home health stays that begin during the 12-month observation period.
S.10. **Denominator Exclusions:** The following are excluded:

1) Home health stays for patients who are not continuously enrolled in fee-for-service Medicare for the 60 days following the start of the home health stay or until death.
2) Home health stays that begin with a Low Utilization Payment Adjustment (LUPA) claim.
3) Home health stays in which the patient receives service from multiple agencies during the first 60 days.
4) Home health stays for patients who are not continuously enrolled in fee-for-service Medicare for the 6 months prior to the home health stay.

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Administrative claims
S.26. **Level of Analysis:** Facility

IF Endorsement Maintenance – Original Endorsement Date: Mar 31, 2009 **Most Recent Endorsement Date:** Aug 10, 2012

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?** Not currently included in a composite measure.

Maintenance of Endorsement -- 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. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

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.

Summary of evidence provided by the developer:

- The developer describes evidence of processes that can be undertaken to reduce acute care hospitalization (ACH) use including care coordination, physician follow up, pharmacist involvement, transitional care, telehealth and a variety of home health care specific evidence-based strategies from the Quality Improvement Organizations (medication management, care provision (frontloading visits), patient education strategies, falls prevention and other topics).

Question for the Committee:

- *Is there at least one process that the provider undertake to improve the measure results?*

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.

- The developer provides data on the distribution of performance of this measure for four years (2011, 2012, 2013, and 2014). These data note that the average risk-adjusted acute care hospitalizations for 2014 was 14.8%; and the 25th percentile was 12.7% and 75th percentile was 16.8%. These distribution of agency performance has a standard deviation of 3.3%.

Disparities

- Risk adjusted measure score by race/ethnicity for 2014 showed no difference between the White and Black populations at both 16.1% rate. The hispanic population had the lowest hospitalization rate at 13.7%.
- The risk adjusted measure score by disability status for 2014 showed a 2.1% difference in acute hospitalization rate between people with disability and with no disability, at 17.6% and 15.5% respectively.

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)

1. Importance to Measure and Report

1a. Evidence to Support Measure Focus

Comments: **limited published validity. Better data that total readmissions can be reduced and not the readmission rate (from CMS demonstrations project).

**This measure is up for maintenance of endorsement.

This is a utilization outcome ACH for which there exists some evidence of multiple processes that have impacts on ACH utilization.

Appears to be supported by the rationale.

1b. Performance Gap

Comments: **There is variation across facilities. Assuming this variation is due to performance then there is a performance gap

**Yes. There is a small performance gap. Interquartile range 4.1

Small differences between disabled and non-disabled populations. In comparison with white populations, hispanics had lower ACH utilization.

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

Comments: **NA

**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.

- This measure calculates the percentage of home health stays in which patients were admitted to an acute care hospital during the 60 days following the start of the home health stay.
- This is a health outcome measure and the level of analysis is facility.
- The denominator is the number of home health stays that begin during the 12-month observation period.
- The numerator is the number of home health stays for patients who have a Medicare claim for an unplanned admission to an acute care hospital in the 60 days following the start of the home health stay.
- A home health stay is defined as a sequence of home health payment episodes separated from other home health payment episodes by at least 60 days.
- The 60 day time window is calculated by adding 60 days to the "from" date in the first home health claim in the series of home health claims that comprise the home health stay. Acute care hospitalization occurs (and the home health stay is included in the numerator) if the patient has at least one Medicare inpatient claim from short term or critical access hospitals during the 60 day window.
- The data sources for this measure may include Medicare Home Health Claims, Medicare Inpatient Claims, Medicare Part A and B claims, and the Medicare Enrollment Database (EDB).
- The measure's time window is 12 months.
- The measure is risk-adjusted using a statistical risk model (see details below).

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:
This measure was endorsed in the Care Coordination Phase 2 Project. All agencies with at least 20 home health stays beginning between 1/1/2010 and 12/31/2010 were included in the reliability analysis, because only information for agencies with at least 20 episodes is publicly reported. Of the 10,125 agencies with any home health stays in 2010, 8,567 agencies met the threshold for the Acute Care Hospitalization measure. Reliability testing methods and results are the same as provided at initial endorsement.

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

A beta-binomial distribution was fitted for all agencies. The beta-binomial method was developed for provider level measures reported as rates, and it allows one to calculate an agency level "reliability score," interpreted as the percent of variance due to the difference in measure score among providers.

Results of reliability testing

The developer notes that the distribution of national reliability scores shows that the majority of agencies have a reliability score greater than 0.871 and that this implies their performance can likely be distinguished from other agencies. This can be interpreted as 87% of the variance is due to differences among providers, and 13% of the variance is due to measurement error or sampling uncertainty.

Guidance from the Reliability Algorithm

Question 1: Submitted specifications are precise, unambiguous, and complete.
Question 2: Empirical reliability testing was conducted using a beta-binomial distribution.
Question 3: Empirical validity testing of patient-level data was conducted.
Question 4: Reliability testing was conducted with computed performance measure scores for each measure
Question 5: The beta binomial method was appropriate for assessing the proportion of variability due to real differences among measured entities.
Question 6: The distribution of national reliability scores (percent of variance due to the difference in measure score among providers at the national level) shows the majority of agencies have a reliability score greater than 0.871, implying that their performance can likely be distinguished from other agencies (i.e., performance on this measure is unlikely to be due to measurement error or insufficient sample size, but is instead due to true differences between the agency and other agencies as it substantially exceeds within agency variation). The distribution of hospital referral region (HRR) reliability scores (percent of variance due to the difference in measure score among providers at the HRR level) for this measure also shows that at least 50% of agencies have a reliability score greater than 0.772, suggesting that between agency variation substantially exceeds within agency variation even at the HRR level.

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?*

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**

- This measure calculates the number of home health stays for patients who have a Medicare claim for an unplanned admission to an acute care hospital in the 60 days following the start of the home health stay.
- As a rationale for measuring this health outcome, the developers suggest that there strategies that can be undertaken to reduce acute care hospitalizations including care coordination, physician follow up, hospital discharge planning and a variety of home health care specific evidence-based strategies including medication management, care provision (frontloading visits), patient education strategies, and falls prevention.

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 developer did not conduct additional validity testing of the measure elements noting that CMS audits a sample of claims for acute inpatient hospitalizations as a part of the annual payment error calculations.
- The developers tested the validity of the measure through the use of payment error audits. The developers justified this during the prior review by stating that there is no reason to believe hospital would be more likely to have erroneous claims for home health patients than for others.

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:

- Audit of claims data

Validity testing results:

- Of a 2010 audit of 2,454 claims for Acute Inpatient Hospitalizations, there was only one case where the hospital had no record of the acute hospitalization admission.

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?*

Other specific question of the validity testing?

2b3-2b7. Threats to Validity

2b3. Exclusions:

- Patients in the following categories are excluded from the measure:
  - Home health stays for patients who are not continuously enrolled in fee-for-service Medicare for the 60 days following the start of the home health stay or until death.
  - Home health stays that begin with a Low Utilization Payment Adjustment (LUPA) claim.
  - Home health stays in which the patient receives service from multiple agencies during the first 60 days.
  - Home health stays for patients who are not continuously enrolled in fee-for-service Medicare for the 6 months prior to the home health stay.

- The developer notes that the exclusion criteria are based on either data requirements for calculating the measure (continuous enrollment in fee-for-service Medicare) or clear attribution of the measure to the home health agency (LUPAs and change of provider).
- To determine the impact of exclusions, the developer examined overall frequencies and proportions of the total cohort excluded for each exclusion criteria.
- The number and percentage of patients excluded for each criterion are as follows:
    - 126,480 stays (4%) were excluded because the patient was not continuously enrolled in fee-for-service Medicare during the numerator window or until death.
    - 275,342 stays (9%) were excluded because the first claim in the stay was a LUPAs.
    - 37,733 stays (1%) were excluded because the beneficiary changed agencies during the numerator window.
    - 116,757 stays (4%) were excluded because the patient was not continuously enrolled in fee-for-service Medicare for six month look-back period used to calculate hierarchical condition categories (HCCs).
    - 22,621 acute care hospitalizations (0.9%) were not counted toward the measure numerator because they were determined to be planned hospitalizations.

Questions for the Committee:

- *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: **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

- This measure employs a multinomial logit model.
- Variables included in the model include prior care setting (e.g., outpatient emergency room, inpatient acute, psychiatric facility, etc.), health status (measured using HCCs and all remaining CCs), demographic information (measured using age-gender interactions), enrollment status (ESRD and disability), and interactions between these factors.
- To determine which risk factors should be included in the risk adjustment model, a Wald test of joint restrictions was used. Variables that were significant at a level of 0.05 for either outcomes in at least 70 percent of the bootstrap samples were included in the final risk adjustment model.
- The developer calculated counterfactuals to show the impact of each risk factor. Each risk factor has an associated counterfactual value that can be interpreted as the population value of the measure if all patients in the population had the risk factor but had the observed distribution of all other risk factors. The counterfactual represents the relative impact of each risk factor on the outcome.

Conceptual analysis of the need for SDS adjustment

- The developer found that while a recent review (Goodridge et al. Socioeconomic disparities in home health care service access and utilization: A scoping review 2012: *International J. Nursing Studies* 49(10); 1310-19) found that persons of lower socioeconomic status are not disadvantaged in terms of home health care service, findings from the literature support a link between SDS factors and emergency department use and hospital readmission.
- The developer notes that in the home health setting, the 60-day period for hospitalization occurs while the patient is living in their own home, increasing the likelihood that non-medical factors, including geographic location and economic resources, will have an impact on acute care use.
- Specifically, the developer found evidence in the literature of an impact of factors such as race and ethnicity, lower income, living alone, and lower levels of education on ED use and hospital readmission.

Empirical analysis of SDS factors:

- The developer notes that Data for race/ethnicity, disability status, rural location, sex, and Medicaid dual status were readily available through the enrollment database (EDB) and analyzed during the measure development process.

- The developer performed univariate analyses by race/ethnicity, disability status, rural location, and sex.
- The developer does not recommend controlling for SDS factors at this time.
- The results are summarized in the following tables:

Distribution of performance rates, by sex

Sex	2011		2012		2013		2014	
	Observed	Risk Adjusted	Observed	Risk Adjusted	Observed	Risk Adjusted	Observed	Risk Adjusted
Male	18.8%	17.3%	18.5%	17.6%	18.0%	17.6%	17.4%	17.4%
Female	16.3%	15.0%	15.9%	15.2%	15.5%	15.2%	14.9%	15.1%

Distribution of Performance Rates, by Race/Ethnicity

Race/Ethnicity	2011		2012		2013		2014	
	Observed	Risk Adjusted	Observed	Risk Adjusted	Observed	Risk Adjusted	Observed	Risk Adjusted
White	17.3%	15.9%	17.0%	16.2%	16.5%	16.2%	16.0%	16.1%
Black	17.7%	15.8%	17.4%	16.0%	17.0%	16.2%	16.7%	16.1%
Hispanic	13.6%	14.2%	13.9%	14.5%	12.7%	14.1%	11.6%	13.7%
Other	15.6%	15.3%	15.2%	15.5%	14.7%	15.3%	14.1%	15.2%

Distribution of Performance Rates, by Disability Status

Disability Status	2011		2012		2013		2014	
	Observed	Risk Adjusted	Observed	Risk Adjusted	Observed	Risk Adjusted	Observed	Risk Adjusted
Yes	19.0%	17.3%	18.6%	17.6%	18.1%	17.7%	17.7%	17.6%
No	16.7%	15.4%	16.3%	15.7%	15.9%	15.6%	15.3%	15.5%

Distribution of Performance Rates, by Urban/Rural Status

Urban/Rural Status	2011		2012		2013		2014	
	Observed	Risk Adjusted	Observed	Risk Adjusted	Observed	Risk Adjusted	Observed	Risk Adjusted
Urban	17.3%	15.8%	16.7%	16.1%	16.3%	16.1%	15.7%	15.9%
Rural	16.7%	15.9%	17.5%	16.0%	17.1%	16.1%	16.5%	16.1%

Risk model diagnostics

- To assess the overall performance of their risk-adjustment model, the developers computed several summary statistics, including:
  - *Area under the receiver operating characteristic (ROC) curve* (also known as a c-statistic, which measures the probability that the model's prediction of the outcome is better than chance)
  - *Predictive ability* (the model's ability to distinguish high-risk subjects from low-risk subjects)
  - *Over-fitting indices (model calibration)* (to ensure that the model is not only describing the relationship between predictive variables and outcome in the development dataset but also providing valid predictions in new patients)
- The developer used a cross-validation method to test for over-fitting. The statistics computed to test over-fitting include c-statistics, a calibration statistic, and a discrimination statistic expressed in terms of predictive ability.
- A version of the area under the receiver operating curve statistic, also known as the c-statistic, was calculated

for each individual logit and for the model overall. The c-statistic was calculated for each individual logit and the model overall. The c-statistic is 0.693.
    - A c-statistic of .69 means that for 69% of all possible pairs of patients-one who was hospitalized and one who was not-the model correctly assigned a higher probability to those who were hospitalized. Generally, a c-statistic of at least 0.70 is considered acceptable.
- To compute the calibration statistic, the vector of coefficients is estimated from the model on the development sample. These coefficients are then multiplied with the matrix of covariates from the validation sample to give a scalar linear predictor for the probability of an event for a given observation in the validation sample. A logistic regression is then estimated on the validation sample with an intercept and one covariate, the linear predictor. Values of the intercept far from 0 and values of the coefficient on Z far from 1 provide evidence of over-fitting. The calibration statistic for ACH produced an intercept of -0.005 and a coefficient of 0.996.
    - The developers notes with t-statistics of 0.598 and 0.656, these values are not significantly different from 0 and 1, respectively, at the 95% confidence level.
- The developer computed cross-validation statistics by looking at the difference between the 10th percentile of predicted probabilities for an event and contrasting this with the 90th percentile.
    - In the development sample, the range of predicted probabilities for ACH was 8 to 31%. In the verification sample, this range was identical at 8 to 31%.

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?*
-

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

- The distribution of risk adjusted agency rates was analyzed to determine the inter-quartile range and the 90th vs. 10th percentile differences.
    - 10th percentile 11.3%
    - 25th percentile 14.3%
    - 50th percentile 17.2%
    - 75th percentile 19.9%
    - 90th percentile 22.9%
    - Inter-quartile range (75th – 25th) = 19.9 – 14.3 = 5.6%
    - 90th – 10th percentile = 22.9 – 11.3 = 11.6%

Question for the Committee:

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

2b6. Comparability of data sources/methods:

- N/A. The measure uses a single data source.

2b7. Missing Data

- N/A.

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

Comments: **None.

Specs consistent with evidence.

2a2. Reliability Testing

Comments: **good reliability

**Reliability testing was performed on all HHAs with at least 20 stays in 2010, representing approximately 85% of the total sample. This appears reasonable from a practical perspective.

the developers used the betabinomial method and reported the reliability score distribution at the agency and the HRR (hospital referral region level). Reliability scores at both levels of analysis appear to support that variation in quality occurs because of differences between agencies.

2b2. Validity Testing

Comments: **Unclear validity as a measure of quality.

**Empirical validity testing was performed via a MCR claims audit, representing 2454 ACH claims. Of this group, only 1 showed no admission. (there were however, 9.5% claims with some degree of payment error, but since the main outcome is ACH, the developers state that this may represent the upper bound for possible admissions that are not medically necessary).

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **Differences may not be a measure of quality

**Exclusions: No specific objections. Exclusions mainly because of missing/duplicative data required to calculate the measure--from a practical perspective LUPA stays may indicate that the HHA did not have the ability to fully implement the careplan and thus are appropriate for exclusion. Likewise AMA stays, multiple HHA utilization (?interrupted stays). Exclusion of planned ACH is appropriate relative to the measure. These exclusions are based on claims data and should not add additional burden.

Risk adjustment: used a multinomial logit model. c statistic was 0.69--acceptable. conceptual rationale for SDS variables was presented. Testing of risk adjustment was appropriate-- in the development and verification samples, the results were nearly identical.

In terms of meaningful differences, the risk adjusted interquartile range was 5.6%--a small but arguably significant difference, and there is a 11.6% difference between the 90th and 10th percentiles.

SDS relationship variables were not included.

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.

The developer states:

- All measure elements are generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition. All data are directly obtained from Medicare hospital claim dates.
- In electronic forms, ALL data elements are in defined fields in electronic claims.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*

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

Committee pre-evaluation comments
Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **Highly feasible

**No issues with feasibility as this uses administrative claims 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
Publicly reported? ☒ **Yes** ☐ **No**

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

Accountability program details

- Public Reporting
  - Home Health Compare: http://www.cms.gov/HomeHealthCompare/search.aspx
- Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
  - Home Health Quality Initiative: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HomeHealthQualityInits/index.html

Improvement results
The developer reports that:

- Between the CY2011 and the CY2014 measurement period,
  - At the agency level, mean risk-adjusted performance rate on this measure increased from 14.7 percent to 14.8 percent
  - At the population level, the risk-adjusted performance rate has remained stable across all population groups
- Additional information on the geographic area and number and percentage of accountable entities and patients included in this analysis is shown in the Attachment: Importance to Report [HYPERLINK]

Feedback :

- Included in the 2012 MAP Care Coordination/Hospice and Palliative Care Family of Measures.
- MAP included this measure on the Care Coordination family of measures due to its focus on patients receiving home care services and are subsequently hospitalized or visit the ED. MAP recommended that similar measures be developed for other post-acute and long-term care settings.
- In general, MAP prefers outcome measures over process and structural measures.

***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

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **Highly usable. May be useful to decrease costs, not as clear regarding quality.

**Already being used for public reporting: home health compare and home health quality initiative.

Criterion 5: Related and Competing Measures

Related or competing measures

- 2380 : Rehospitalization During the First 30 Days of Home Health
    - During the endorsement of #2380, the Standing Committee reviewed this issue. The previous findings were:
        - This measure competes directly with measure 0171 Acute Care Hospitalization—Percentage of Home Health stays in which patients were admitted to an acute care hospital during the 60 days following the start of the Home Health stay.
        - The measure specifications for measure 0171 and measure 2380 were harmonized along several measure dimensions, including Data source, Population, Denominator Exclusions, Numerator, and Risk Adjustment methodology.
        - The developers of this measure contended that there are differences that justify having two separate measures. Whereas measure 0171 evaluates patient admission to an acute care hospital during the 60 days following the start of a Home Health stay (regardless of whether or not this stay was preceded by an inpatient hospitalization), measure 2380 evaluates readmission to the hospital within 30 days after starting Home Health care for patients who were recently discharged from an inpatient setting. Home Health agencies can track their performance on both utilization measures to gain an accurate picture of how much acute care is being used by their patients. Additionally, measure 2380 is an outcome measure that assesses the efficacy of care coordination as patients transition from inpatient acute care to outpatient Home Health services. In contrast, measure 0171 assesses the efficacy of clinical care provided to all patients, as indicated by rates of hospitalization after entry into Home Health services.
        - These are distinct domains of care under the CMS Quality Strategy and reflect related but distinct care quality concepts. This is not the only setting in which CMS has developed paired readmission and hospitalization measures. Such measures exist for end-stage renal disease (ESRD), and such pairings are being considered in other care settings as well.
    - In the review of #2380 the committee agreed it was sufficiently different to continue endorsement on both. The committee reviewed the developer's justification and agreed with the assessment that both measures should be endorsed.
    - According to NQF guidance, since #0171 was not reviewed in the project, the committee did not make a recommendation with regards to these two competing measures.

Harmonization

- The developer states that this measure is "harmonized with the Rehospitalization measures (NQF numbers 2505 and 2380) and with CMS' Hospital-Wide All-Cause Unplanned Readmission (HWR) measure (NQF 1789) in the definition of unplanned hospitalizations."'
- The developer adds that this measure differs from other post-acute hospital readmission measures due to the unique nature of home health care as a post-acute setting.
- The developer states that this measure is risk adjusted using patient-level predicted probabilities calculated

from a multinomial logistic regression. Risk factors that are accounted for include demographics and health status as measured by both CMS' Hierarchical Condition Categories (HCCs) found on claims in the previous six months. The differences of this measure from other post-acute care measures arise from the unique nature of home care as well as from a desire for harmonization across home health quality measures."
***Questions for the Committee**:* • Does the committee confirm its prior decision that #0171 and #2380 are sufficiently different to maintain endorsement on both measures?

Pre-meeting public and member comments

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0171

Measure Title: Acute care hospitalization

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: 1/29/2016

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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: Click here to name the health outcome

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

Process-outcome (as ACH is a utilization outcome). There is evidence of processes that can be undertaken to reduce ACH use including care coordination, physician follow up, pharmacist involvement, transitional care, telehealth and a variety of home health care specific evidence-based strategies from the Quality Improvement Organizations (medication management, care provision (frontloading visits), patient education strategies, falls prevention and other topics).

Related question and response from previous submission:

1c.1 Structure-Process-Outcome Relationship *(Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process-health outcome; intermediate clinical outcome-health outcome)***:**

Process-outcome (as ACH is a utilization outcome). There is evidence that there are strategies that can be undertaken to reduce ACH use including care coordination, physician follow up, hospital discharge planning and a variety of home health care specific evidence-based strategies from the Quality Improvement Organizations (medication management, care provision (frontloading visits), patient education strategies, falls prevention and other topics).

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

ACH is a multifactorial issue. Recent evidence indicates that a community-wide care transitions program reduced rehospitalization using the Best Practice Intervention Package from the QIOs (Markley et al. 2012).

These findings are discordant with older research by Schade et al. who found no differences in ACH rate for participating versus non-participating home health care agencies.

Reference List

(1) Markley J, Sabharwal K, Wang Z, Bigbee C, Whitmire L. A community-wide quality improvement project on patient care transitions reduces 30-day hospital readmissions from home health agencies. Home Healthc Nurse 2012;30:E1-E11.

(2) Schade CP, Esslinger E, Anderson D, Sun Y, Knowles B. Impact of a national campaign on hospital readmissions in home care patients. Int J Qual Health Care 2009;21:176-182.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

Not Applicable

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

Not Applicable

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Not Applicable

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Not Applicable

Related question and response from previous submission:

1c.16 Quote verbatim, the specific guideline recommendation *(Including guideline # and/or page #)***:**

No guidelines were identified for this measure:

A search of guideline.gov with the terms “hospitalization” and ‘rehospitalization” did not return any relevant guidelines.

Systematic reviews and meta-analyses:

A PubMed Search using the term “rehospitalization” and the limits of meta-analysis or practice guideline returned for the last three years returned 5 results, none of which were relevant. A search within 5 years returned 12 results, none of which were relevant.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

Not Applicable

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

Not Applicable

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

Not Applicable

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

Not Applicable

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

Not Applicable

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

Not Applicable

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

Not Applicable

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

Not Applicable

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Not Applicable

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

(1) Hansen LO, Young RS, Hinami K, Leung A, Williams MV. Interventions to reduce 30-day rehospitalization: a systematic review. Ann Intern Med 2011;155:520-528.

(2) Kim H, Thyer BA. Does transitional care prevent older adults from rehospitalization? A review. J Evid Inf Soc Work 2015;12:261-271.

Related question and response from previous submission:

1c.15 Citations for Evidence other than Guidelines*(Guidelines addressed below)*:

Reference List

(1) Peterson-Sgro K. Reducing acute care hospitalization and emergent care use through home health disease management: one agency´s success story. Home Healthc Nurse 2007; 25(10):622-627.

(2) Schade CP, Esslinger E, Anderson D, Sun Y, Knowles B. Impact of a national campaign on hospital readmissions in home care patients. Int J Qual Health Care 2009; 21(3):176-182.

(3) Silver MP, Ferry RJ, Edmonds C. Causes of unplanned hospital admissions: implications for practice and policy. Home Healthc Nurse 2010; 28(2):71-81.

(4) Daley CM. A hybrid transitional care program. Crit Pathw Cardiol 2010; 9(4):231-234.

(5) Russell D, Rosati RJ, Sobolewski S, Marren J, Rosenfeld P. Implementing a transitional care program for high-risk heart failure patients: findings from a community-based partnership between a certified home healthcare agency and regional hospital. J Healthc Qual 2011; 33(6):17-24.

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Neither were graded

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

A PubMed Search using the term "rehospitalization" and the limits of meta-analysis or practice guideline returned for the last five years returned 6 results, two of which were relevant.

We are presenting the results from Hansen as it had the broadest scope for interventions. Hansen and colleagues (2011) reviewed the literature and reviewed 43 studies on interventions to reduce 30 day rehospitalization. Because the study interventions were insufficiently described, they were unable to perform a meta-analysis. Additionally, 24 of the 43 tested only a single component intervention and of those, only 7 were randomized. The authors conclude that there is insufficient evidence on discrete interventions or bundles of interventions that will "reliably" reduce rehospitalization.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

"Study quality was assessed using an adapted version of the Cochrane Effective Practice and Organisation of Care (EPOC) risk of bias tool. Nine criteria were assessed including randomisation, allocation concealment, similarity of baseline characteristics, outcome assessment, handling of missing data, and likelihood of contamination between study groups."

Related question and response from the previous submission:

1c.6 Quality of Body of Evidence (*Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b) directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events):* While the five studies are home health care specific, two are quality improvement studies versus "research" as they do not include control groups or sufficient scientific rigor to allow for determination of the effects of the interventions. The study by Schade et al is more rigorous, uses an observational study design and matches agencies on factors that may have influenced the results. However, the diffusion of the intervention to the "non-participating" agencies made it impossible to determine whether the QIO best practice program materials were effective. The studies by Russell and Daley were more rigorous but were conducted in single home health care agencies, raising concerns about the extent to which the findings will be generalizable to other agencies. As well, the study by Russell used two different time periods for the control and intervention groups.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

Not Applicable

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

1975-2011

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

"There were 16 randomised controlled trials (RCTs), with sample sizes ranging from 34 to 835 patients. There were 20 quasi-experimental/cohort studies and seven non-controlled before-and-after studies (although the quality assessment table showed 14 controlled clinical studies and 13 non-controlled studies)."

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Inadequate quality according to the authors. Not sufficient information or numbers of studies to undertake a meta-analysis.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

Meta-analysis not performed.

Related question and response from previous submission:

1c.7 Consistency of Results <u>across Studies</u> *(Summarize the consistency of the magnitude and direction of the effect):*
The consistency of the findings are mixed, primarily because there are variations in what interventions agencies use.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

No harms were noted in the Hansen report

Related question and response from previous submission:

1c.8 Net Benefit *(Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms)***:**

It is difficult to use the evidence to determine net benefit as the largest study (Schade et al) found no difference while the two agency-specific studies found small to moderate effects for patients with heart failure.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

Not Applicable

1a.8 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.

Not Applicable

1a.8.1 What process was used to identify the evidence?

Not Applicable

1a.8.2. Provide the citation and summary for each piece of evidence.

Not Applicable

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
0171_Evidence_Form_2016_2-23-16.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)*
see attachment "Importance to Report" for a tabular presentation of these data

Hospital readmissions are a national priority for Medicare recipients, based on evidence that 20% of all Medicare beneficiaries who were hospitalized had a return hospital stay within 30 days. In 2004, this cost the Medicare program $17.4 billion (1). Within home health care, an analysis of Medicare claims shows that 14 percent of home health patients are rehospitalized within 30 days of the start of home health care. There is limited research on the extent to which these hospital readmissions are avoidable within home health care. One study reporting on patients with heart failure found that more than 40% of the 30 day rehospitalizations may have been avoidable (2). In addition, studies of Medicare patients in general provided evidence for interventions that reduce the need for hospital care within a substantial proportion of these Medicare beneficiaries (1;3). Moreover, there are a number of national initiatives, both governmental (e.g. Quality Improvement Organizations, National Priorities Partnership and CMS) and through private foundations (e.g. Institute for Healthcare Improvement), addressing this issue. Thus there is room for improvement and this is a national priority issue.

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.*
see attachment "Importance to Report" for a tabular presentation of these data

No performance scores reported in the literature

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.**
see attachment "Importance to Report" for a tabular presentation of these data

Care coordination is one strategy that has been identified nationally by the National Priorities Partnership to address these high rates of hospital care. Models of care coordination and transitional care have been identified and tested in RCTs and are currently being tested in national demonstration projects with expectations that health care reform activities will incorporate care coordination for persons at high risk of hospitalization and rehospitalization (4). While there has been limited testing of these models within the existing home health care system, there is evidence of effectiveness: Daley reported a small study (N = 89 patients with heart failure [HF]) where care coordination resulted in a reduction in hospitalization rate beyond that expected (15% versus 20%) (5). Russell and colleagues provide preliminary findings on a care transition project within home health care that provided a 57% less likely need for hospital care for persons with HF (6). Finally Markley and colleagues, as part of the CMS Care Transitions project, identified that home health care agencies using care coordination, had 30 day rehospitalization rates of 16.5%, 5% lower than the average overall rate for the community. (7)

In addition to care transition interventions, there is evidence that strategies like telehealth (TH) may be beneficial in reducing

hospitalizations among home health care patients although the evidence on effectiveness of TH is more mixed in meta-analyses (8-12). Complicating the understanding of effectiveness of TH in home health care is that much TH research is done outside the existing home health care system. Another strategy that reduces the likelihood of rehospitalization for home health care patients is prompt physician follow up after a hospital stay. Wolff et al (13) found that 77.6% of home health recipients who received at least one physician evaluation and management visit during their home health stay were discharged to the community (rather than transferred to an inpatient facility) while only 70.6% of patients who did not receive physician visits were discharged to the community, suggesting that increasing physician visits may be cost effective.

One study tested the use of the Care Transitions Measure (CTM), as used in hospital discharge planning, to try to predict rehospitalization in 495 patients from one home health care agency. The CTM did not predict rehospitalization (14). Another study examined the use of frontloading in reducing rehospitalization among 4500 randomly selected patients from 5 home health care agencies. Propensity scoring was used to address confounding variables. The researchers found that frontloading had no impact on the rate of rehospitalization (15).

(1) Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. N Engl J Med 2009 Apr 2;360(14):1418-28.
(2) Madigan EA, Gordon NH, Fortinsky RH, Koroukian SM, Pina I, Riggs JS. Rehospitalization in a national population of home health care patients with heart failure. Health Serv Res 2012 Dec;47(6):2316-38.
(3) Schade CP, Esslinger E, Anderson D, Sun Y, Knowles B. Impact of a national campaign on hospital readmissions in home care patients. Int J Qual Health Care 2009 Jun;21(3):176-82.
(4) Boult C, Green AF, Boult LB, Pacala JT, Snyder C, Leff B. Successful models of comprehensive care for older adults with chronic conditions: evidence for the Institute of Medicine´s "retooling for an aging America" report. J Am Geriatr Soc 2009 Dec;57(12):2328-37.
(5) Daley CM. A hybrid transitional care program. Crit Pathw Cardiol 2010 Dec;9(4):231-4.
(6) Russell D, Rosati RJ, Sobolewski S, Marren J, Rosenfeld P. Implementing a transitional care program for high-risk heart failure patients: findings from a community-based partnership between a certified home healthcare agency and regional hospital. J Healthc Qual 2011 Nov;33(6):17-24.
(7) Markley J, Sabharwal K, Wang Z, Bigbee C, Whitmire L. A community-wide quality improvement project on patient care transitions reduces 30-day hospital readmissions from home health agencies. Home Healthc Nurse 2012 Mar;30(3):E1-E11.
(8) Bowles KH, Holland DE, Horowitz DA. A comparison of in-person home care, home care with telephone contact and home care with telemonitoring for disease management. J Telemed Telecare 2009;15(7):344-50.
(9) Polisena J, Coyle D, Coyle K, McGill S. Home telehealth for chronic disease management: a systematic review and an analysis of economic evaluations. Int J Technol Assess Health Care 2009 Jul;25(3):339-49.
(10) Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K. Home telehealth for diabetes management: a systematic review and meta-analysis. Diabetes Obes Metab 2009 Oct;11(10):913-30.
(11) Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K, et al. Home telehealth for chronic obstructive pulmonary disease: a systematic review and meta-analysis. J Telemed Telecare 2010;16(3):120-7.
(12) Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K, et al. Home telemonitoring for congestive heart failure: a systematic review and meta-analysis. J Telemed Telecare 2010;16(2):68-76.
(13) Wolff JL, Meadow A, Boyd CM, Weiss CO, Leff B. Physician evaluation and management of Medicare home health patients. Med Care 2009 Nov;47(11):1147-55.

(14) Ryvicker M, McDonald MV, Trachtenberg M, Peng TR, Sridharan S, Feldman PH. Can the care transitions measure predict rehospitalization risk or home health nursing use of home healthcare patients? J Healthc Qual 2013;35:32-40.

(15) O´Connor M, Hanlon A, Bowles KH. Impact of frontloading of skilled nursing visits on the incidence of 30-day hospital readmission. Geriatr Nurs 2014;35:S37-S44.

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.*
see attachment "Importance to Report" for a tabular presentation of these data

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.
see attachment "Importance to Report" for a tabular presentation of these data

Not reported in literature.

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, High resource use, Patient/societal consequences of poor quality

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.**

Hospital readmissions are a national priority for Medicare recipients, based on evidence that 20% of all Medicare beneficiaries who were hospitalized had a return hospital stay within 30 days. In 2004, this cost the Medicare program $17.4 billion (1). Within home health care, an analysis of Medicare claims shows that 14 percent of home health patients are rehospitalized within 30 days of the start of home health care. There is limited research on the extent to which these hospital readmissions are avoidable within home health care: one study reporting on patients with heart failure found that more than 40% of the 30 day rehospitalizations may have been avoidable (2).

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

(1) Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. N Engl J Med 2009 Apr 2;360(14):1418-28.

(2) Madigan EA, Gordon NH, Fortinsky RH, Koroukian SM, Pina I, Riggs JS. Rehospitalization in a national population of home health care patients with heart failure. Health Serv Res 2012 Dec;47(6):2316-38.

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.)*

Not applicable

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):*

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

Care Coordination, Overuse

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

https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-

Instruments/HomeHealthQualityInits/HHQIQualityMeasures.html

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:** Data_Dictionaries_ffs_inst_and_non-inst_claims-635895196660789022.xls

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
There are no new changes made to the measure since the last annual maintenance which occurred on October 1, 2015. In the previous maintenance period two minor changes were made to the measures: (1) the title of the measure was changed to improve clarity and (2) recalibration of the risk adjustment model coefficients using data from January 1, 2013 to December 31, 2013.

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.
Number of home health stays for patients who have a Medicare claim for an unplanned admission to an acute care hospital in the 60 days following the start of the home health stay.

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.)*
12 month data collection period, and updated quarterly.

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.
The 60 day time window is calculated by adding 60 days to the "from" date in the first home health claim in the series of home health claims that comprise the home health stay. Acute care hospitalization occurs (and the home health stay is included in the numerator) if the patient has at least one Medicare inpatient claim from short term or critical access hospitals (identified by CMS Certification Number ending in 0001-0879, 0800-0899, or 1300-1399) during the 60 day window.
Inpatient claims for planned hospitalizations are excluded from the measure numerator. Planned hospitalizations are defined using the same criteria as the Yale Hospital-Wide All-Cause Unplanned Readmission Measure. Specifically, admissions are categorized as "planned" based on AHRQ Procedure and Condition CCS as well as other sets of ICD-9-CM procedure codes. These admissions are excluded unless they have a discharge condition category considered "acute or complication of care," which is defined using AHRQ Condition CCS. The definitions of AHRQ CCS can be found here:
http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp#download

The AHRQ CCS that define planned hospitalizations are found below and are AHRQ Procedure CCS unless otherwise noted.

AHRQ CCS Description
45 PTCA
254 Rehabilitation (Condition CCS)
84 Cholecystectomy and common duct exploration
157 Amputation of lower extremity
44 CABG
78 Colorectal resection
51 Endarterectomy; vessel of head and neck
113 Transurethral resection of prostate
99 Other OR Gastrointestinal therapeutic procedures
48 Insertion; revision; replacement; removal of cardiac pacemaker or cardioverter/defibrillator

45 Maintenance chemotherapy (Condition CCS)
211 Therapeutic radiology for cancer treatment
3 Laminectomy; excision intervertebral disc
43 Heart valve procedures
152 Arthroplasty knee
158 Spinal fusion
55 Peripheral vascular bypass
52 Aortic resection; replacement or anastomosis
36 Lobectomy or pneumonectomy
153 Hip replacement; total and partial
60 Embolectomy and endarterectomy of lower limbs
85 Inguinal and femoral hernia repair
104 Nephrectomy; partial or complete
1 Incision and excision of CNS
124 Hysterectomy; abdominal and vaginal
167 Mastectomy
10 Thyroidectomy; partial or complete
114 Open prostatectomy
74 Gastrectomy; partial and total
119 Ooporectomy; unilateral and bilateral
154 Arthroplasty other than hip or knee
ICD-9-CM procedure codes 30.5, 31.74, 34.6 Radial laryngectomy, revision of tracheostomy, scarification of pleura
166 Lumpectomy; quadrantectomy of breast
64 Bone marrow transplant
105 Kidney transplant
176 Other organ transplantation
ICD-9-CM procedure codes 94.26, 94.27 Electroshock therapy

Discharge AHRQ Condition CCS considered “acute or complication of care” are listed below.

AHRQ CCS Description
237 Complications of device; implant or graft
106 Cardiac dysrhythmias
Condition CCS 207, 225, 226, 227, 229, 230, 231, 232 Fracture
100 Acute myocardial infarction
238 Complications of surgical procedures or medical care
108 Congestive heart failure; nonhypertensive
2 Septicemia (except in labor)
146 Diverticulosis and diverticulitis
105 Conduction disorders
109 Acute cerebrovascular disease
145 Intestinal obstruction without hernia
233 Intracranial injury
116 Aortic and peripheral arterial embolism or thrombosis
122 Pneumonia (except that caused by TB or sexually transmitted disease)
131 Respiratory failure; insufficiency; arrest (adult)
157 Acute and unspecified renal failure
201 Infective arthritis and osteomyelitis (except that caused by TB or sexually transmitted disease)
153 Gastrointestinal hemorrhage
130 Pleurisy; pneumothorax; pulmonary collapse
97 Peri-; endo-; and myocarditis; cardiomyopathy
127 Chronic obstructive pulmonary disease and bronchiectasis
55 Fluid and electrolyte disorders
159 Urinary tract infection
245 Syncope
139 Gastroduodenal ulcer (except hemorrhage)
160 Calculus of urinary tract

112 Transient cerebral ischemia

S.7. Denominator Statement *(Brief, narrative description of the target population being measured)*
Number of home health stays that begin during the 12-month observation period.

S.8. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

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)*
A home health stay is a sequence of home health payment episodes separated from other home health payment episodes by at least 60 days. Each home health payment episode is associated with a Medicare home health (HH) claim, so home health stays are constructed from claims data using the following procedure.

1.First, retrieve HH claims with a "from" date (FROM_DT) during the 12-month observation period or the 120 days prior to the beginning of the observation period and sequence these claims by "from" date for each beneficiary.
2.Second, drop claims with the same "from" date and "through" date (THROUGH_DT) and claims listing no visits and no payment. Additionally, if multiple claims have the same "from" date, keep only the claim with the most recent process date.
3.Third, set Stay_Start_Date(1) equal to the "from" date on the beneficiary's first claim. Step through the claims sequentially to determine which claims begin new home health stays. If the claim "from" date is more than 60 days after the "through" date on the previous claim, then the claim begins a new stay. If the claim "from" date is within 60 days of the "through" date on the previous claim, then the claim continues the stay associated with the previous claim.
4.Fourth, for each stay, set Stay_Start_Date(n) equal to the "from" date of the first claim in the sequence of claims defining that stay. Set Stay_End_Date(n) equal to the "through" date on the last claim in that stay. Confirm that Stay_Start_Date(n+1) – Stay_End_Date(n) > 60 days for all adjacent stays.
5.Finally, drop stays that begin before the 12-month observation window.

Note the examining claims from the 120 days before the beginning of the 12-month observation period is necessary to ensure that stays beginning during the observation period are in fact separated from previous home health claims by at least 60 days.

S.10. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*
The following are excluded:

1) Home health stays for patients who are not continuously enrolled in fee-for-service Medicare for the 60 days following the start of the home health stay or until death.
2) Home health stays that begin with a Low Utilization Payment Adjustment (LUPA) claim.
3) Home health stays in which the patient receives service from multiple agencies during the first 60 days.
4) Home health stays for patients who are not continuously enrolled in fee-for-service Medicare for the 6 months prior to the home health stay.

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)*
Four types of home health stays are excluded from the measure denominator:

1.Home health stays for patients who are not continuously enrolled in fee-for-service Medicare for the 6 months prior to the start of the home health stay, for the 60 days following the start of the home health stay, or until death.
• Both enrollment status and beneficiary death date are identified using the Medicare Enrollment Database (EDB).

2.Home health stays that begin with a Low Utilization Payment Adjustment (LUPA) claim.
• Exclude the stay if LUPAIND = L for the first claim in the home health stay.

3.Home health stays in which the patient receives service from multiple agencies during the first 60 days.
• Define Initial_Provider = PROVIDER on the first claim in the home health stay.
• If Intial_Provider does not equal PROVIDER for a subsequent claim in the home health stay AND if the "from" date of the

subsequent claim is within 60 days of Stay_Start_Date, then exclude the stay.

4.Home health stays for patients who are not continuously enrolled in fee-for-service Medicare for the 6 months prior to the start of the home health stay.
•Enrollment status is identified using the Medicare Enrollment Database (EDB).

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)*
Not applicable

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)*
Multinomial logit with outcomes of "No acute event", "Emergency Department without Hospitalization", and "Acute Care Hospitalization".

Risk factors include:

Prior Care Setting –
The main categories are community (i.e., no prior care setting), outpatient emergency room, inpatient-acute (IP-acute), inpatient rehabilitation facility (IRF), psychiatric facility, long-term care hospital (LTCH), and skilled nursing facility (SNF). The hierarchy of setting is SNF, most recent inpatient stay (including IP-acute, IRF, LTCH, and psychiatric facility), outpatient ER, and community. Acumen used the five cohorts from the Yale Hospital-Wide All-Cause Unplanned Readmission Measure to segregate the IP-acute category. The five cohorts are:
1.Surgery/Gynecology: admissions likely cared for by surgical or gynecological teams, based on AHRQ procedure categories;
2.Cardiorespiratory: admissions treated by the same care teams with very high readmission rates, such as for pneumonia, chronic obstructive pulmonary disease, and heart failure;
3.Cardiovascular: admissions treated by separate cardiac or cardiovascular team in large hospitals, such as for acute myocardial infarctions;
4.Neurology: admissions for neurological conditions, such as stroke, that may be treated by a separate neurology team in large hospitals; and
5.Medicine: admissions for all other non-surgical patients.

These cohorts were designed to account for differences in readmission risk for surgical and non-surgical patients.
Finally, the IP-acute categories and the SNF category were further refined by length of stay. Each of the five IP-acute categories are separated into stays of length 0 to 3 days, 4 to 8 days, and 9 or more days, while the SNF categories are split into stays of length 0 to 13, 14 to 41, and 42 and more days. A patient cared for in both a skilled nursing facility and an inpatient hospital during the 30 days prior to starting home health care is included in the skilled nursing categories and not the inpatient categories. The length of stay is determined from the last inpatient or skilled nursing stay prior to beginning home health care.

Age and Gender Interactions –
Age is subdivided into 12 bins for each gender: aged 0-34, 35-44, 45-54, five-year age bins from 55 to 95, and a 95+ category. Using a categorical age variable allows the model to account for the differing effects of age and gender. Age is determined based on the patient's age at Stay_Start_Date.

CMS Hierarchical condition categories (HCCs) –
HCCs were developed for the risk adjustment model used in determining capitation payments to Medicare Advantage plans and are calculated using Part A and B Medicare claims. While the CMS-HHC model uses a full year of claims data to calculate HCCs, for these measures, we use only 6 months of data to limit the number of home health stays excluded due to missing HCC data. All 2012 HCCs and CCs that are not hierarchically ranked that were statistically significant predictors of ACH and ED use are included in the model.

Details of the CMS-HCC model and the code lists for defining the HCCs can be found here:

https://www.cms.gov/MedicareAdvtgSpecRateStats/06_Risk_adjustment.asp

A description of the development of the CMS-HCC model can be found here:
https://www.cms.gov/HealthCareFinancingReview/Downloads/04Summerpg119.pdf

ESRD and Disability Status –
Original End Stage Renal Disease (ESRD) and current ESRD status are included as risk factors. Original disabled status and male, and original disabled status and female, are also included. Medicare beneficiaries with ESRD or disabled status represent a fundamentally different health profile.

Interaction Terms –
All interaction terms included in the 2012 HCC risk adjustment models that were statistically significant predictors of ED Use and ACH were included. Interaction terms account for the additional effect two risk factors may have when present simultaneously, which is more than the additive effect of each factor separately.

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.
Provided in response box S.15a

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HomeHealthQualityInits/Downloads/Claims-Based-ACH-and-ED-Use-Measures-Technical-Documentation-and-Risk-Adjustment.zip

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.)*
1.Construct Home Health Stays from HH Claims (see 2a1.7 for details)
2.Identify numerator window (60 days following Stay_Start_Date) for each stay and exclude stays for patients who are not continuously enrolled in fee-for-service Medicare during the numerator window or until patient death.
3.Exclude stays that begin with a LUPA or that involve a provider change during the numerator window
4.Link stays to enrollment data by beneficiary.
5.Exclude stays for patients who are not continuously enrolled in fee-for-service Medicare during the 6 months prior to Stay_Start_Date.
6.Calculate demographic risk factors for each stay (age, gender, etc.) using enrollment data.
7.Link to Part A and Part B claims for 6 months prior to Stay_Start_Date for each beneficiary
8.Calculate prior care setting indicators, HCCs, and HCC interactions.
9.Link to Inpatient (IP) claims from Short Stay and Critical Access hospitals (excluding planned hospitalizations - see 2a1.3 for details) for numerator window (60 days following Stay_Start_Date)
10.Set Hospital Admission indicator (Hosp_Admit = 1) if any IP claims are linked to the stay in step 9.
11.Using coefficients from the multinomial logit risk model and risk factors calculated in steps 6 and 8, calculate the predicted probability of being included in the measure numerator for each stay (Pred_Hosp). Additionally calculate the average of Pred_Hosp across all stays that are included in the measure denominator (not excluded in steps 3 or 5) and call this value National_pred_Hosp.
12.Calculate observed and risk adjusted rates for each home health agency (Initial_Provider):
a.Calculate the observed rate of Acute Care Hospitalization as the fraction all (non-excluded) HH Stays with that agency as Initial_Provider that are also included in the measure numerator (Hosp_Admit = 1). Call the value Agency_obs_Hosp.
b.Calculate the agency predicted rate of Acute Care Hospitalization by taking the average of Pred_ Hosp across all (non-excluded) stays with that agency as Initial_Provider. Call this value Agency_pred_Hosp.

c.Calculate the risk adjusted rate of Acute Care Hospitalization using the following formula: Agency_riskadj_Hosp = National_pred_Hosp + (Agency_obs_Hosp – Agency_pred_Hosp). If an agency's calculated risk adjusted rate is negative, that agency will have a publicly reported rate of 0%

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 at measure-specific web page URL identified in S.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.
Not applicable

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.
Not applicable

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Not applicable

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

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.
Denominator: Medicare Home Health Claims
Numerator: Medicare Inpatient Claims
Exclusions: Medicare Home Health Claims, Medicare Enrollment Data
Risk Factors: Medicare Enrollment Data, Medicare Part A & B Claims

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)*
Available at measure-specific web page URL identified in S.1

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)*
Home Health
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.)*
Not applicable

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
0171_MeasureTesting_02-19-16.doc

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 0171	**NQF Project:** Care Coordination Project

2. RELIABILITY & 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. **(evaluation criteria)**

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate field. Supplemental materials may be referenced or attached in item 2.1. See guidance on measure testing.

2a2. Reliability Testing. (*Reliability testing was conducted with appropriate method, scope, and adequate demonstration of reliability*.)

2a2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**

All agencies with at least 20 home health stays beginning between 1/1/2010 and 12/31/2010 were included in the reliability analysis, because only information for agencies with at least 20 episodes is publicly reported. Of the 10,125 agencies with any home health stays in 2010, 8,567 agencies met the threshold for the Acute Care Hospitalization measure. For the national analysis, a beta-binomial distribution was fitted using all agencies. For the HHR (hospital referral region) analysis described below, separate beta-binomials were fitted for each of 306 HHRs, using only those agencies in the HHR. It is worth noting that even the agencies that are in HRRs with only two agencies have high reliability scores, because these small HRR agencies tend to service many home health patients relative to the rest of the country.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)***:**

Reliability analysis of this measure follows the beta-binomial method described in "The Reliability of Provider Profiling: A Tutorial" by John L. Adams. The beta-binomial method was developed for provider level measures reported as rates, and it allows one to calculate an agency level "reliability score," interpreted as the percent of variance due to the difference in measure score among providers. Thus, a reliability score of .80 signifies that 80% of the variance is due to differences among providers, and 20% of the variance is due to measurement error or sampling uncertainty. A high reliability score implies that performance on a measure is unlikely to be due to measurement error or insufficient sample size, but rather due to true differences between the agency and other agencies. Each agency receives an agency specific reliability score which depends on both agency size, agency performance on the measure, and measure variance for the relevant comparison group of agencies. The observed rates of acute care hospitalization, rather than the risk adjusted rates, were used for this analysis as the assumptions of this method are only appropriate for observed rates.

In addition to calculating reliability scores at the national level, we also calculated agency reliability scores at the level of

hospital referral regions (HRRs), because the HRR grouping more adequately captures the types of comparisons health care consumers are likely to make. HRRs are region designations determined in the Dartmouth Atlas of Health Care study, and they represent regional health care markets for tertiary medical care that generally requires the service of a major referral center. They are aggregated hospital service areas (HSAs) and thus aggregated local health care markets. The HRRs are used to determine categories of sufficient size to make comparisons while still capturing the local set of HHA choices available to a beneficiary.

2a2.3 Testing Results *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**

Distribution of Within National Reliability Scores by Case Volume for Agencies with At Least 20 Stays

Number of Stays		Mean	Min	10th	25th	Median	75th	90th	Max
All Agencies	0.831	0.336	0.623	0.756	0.871	0.938	0.969	1.000	
20 to 99	0.706	0.336	0.521	0.623	0.721	0.791	0.858	1.000	
100 to 499	0.894	0.739	0.825	0.863	0.903	0.932	0.948	0.992	
>500	0.970	0.938	0.954	0.961	0.970	0.980	0.987	0.999	

The distribution of national reliability scores (percent of variance due to the difference in measure score among providers at the national level) shows the majority of agencies have a reliability score greater than 0.871, implying that their performance can likely be distinguished from other agencies (i.e., performance on this measure is unlikely to be due to measurement error or insufficient sample size, but is instead due to true differences between the agency and other agencies as it substantially exceeds within agency variation).

Distribution of Within HHR Reliability Scores by Case Volume for Agencies with At Least 20 Stays

Number of Stays		Mean	Min	10th	25th	Median	75th	90th	Max
All Agencies	0.727	0.074	0.435	0.607	0.772	0.881	0.938	1.000	
20 to 99	0.598	0.074	0.331	0.456	0.607	0.748	0.849	1.000	
100 to 499	0.778	0.133	0.587	0.702	0.808	0.882	0.927	0.991	
> 500	0.899	0.373	0.804	0.869	0.919	0.954	0.974	0.996	

The distribution of HRR reliability scores (percent of variance due to the difference in measure score among providers at the HRR level) for this measure also shows that at least 50% of agencies have a reliability score greater than 0.772, suggesting that between agency variation substantially exceeds within agency variation even at the HRR level.

2b. VALIDITY. Validity, Testing, <u>including all Threats to Validity</u>: H☐ M☐ L☐ I ☐

2b1.1 Describe how the measure specifications *(measure focus, target population, and exclusions)* **are consistent with the evidence cited in support of the measure focus (***criterion 1c)* **and identify any differences from the evidence:**

CMS chose to respecify the Acute Care Hospitalization measure with Medicare claims data to enhance the validity and reliability of this measure. The measure population is limited to fee-for-service (FFS) Medicare beneficiaries, ensuring that Medicare claims are filed for all covered services. The measure numerator is a broad measure of utilization (Acute Care Hospitalization) that can be cleanly identified using claims data. Because claims form the basis of Medicare payments, CMS invests significant resources in validating claims submissions prior to payment.

2b2. Validity Testing. (*Validity testing was conducted with appropriate method, scope, and adequate demonstration of*

validity.)

2b2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**

As CMS audits a sample of claims for acute inpatient hospitalizations as part of annual payment error calculations, additional validity testing of measure elements has not been conducted. The annual payment error calculation for 2010 involved a sample of Medicare claims that were then compared to medical records and included 2,454 claims for Acute Inpatient Hospitalizations.

2b2.2 Analytic Method *(Describe method of validity testing and rationale; if face validity, describe systematic assessment)***:**

Review of 2010 Medicare CERT Report. Available at: https://www.cms.gov/CERT/Downloads/Medicare_FFS_2010_CERT_Report.pdf

2b2.3 Testing Results *(Statistical results, assessment of adequacy in the context of norms for the test conducted; if face validity, describe results of systematic assessment)***:**

Of the sampled claims, the hospital had no record of seeing the patient in only one case. It is possible that an extremely small fraction of claims represent care that did not occur, but this problem is clearly not widespread. For acute inpatient hospital claims reviewed, 9.5% had some type of payment error. Payment error analysis can also shed light on cases where the patient was hospitalized, but the hospitalization was not medically necessary. Payment errors include insufficient documentation, meaning the reviewers can't determine if the treatment (including hospital admission) was medically necessary, and medical necessity errors. In some cases, the reviewers determined that the patient's medical condition did not require admission to an acute inpatient hospital. Thus 9.5% represents an upper bound on the extent to which Medicare claims document hospitalizations that were not medically necessary.

POTENTIAL THREATS TO VALIDITY. (*All potential threats to validity were appropriately tested with adequate results.*)

2b3. Measure Exclusions. (*Exclusions were supported by the clinical evidence in 1c or appropriately tested with results demonstrating the need to specify them.*)

2b3.1 Data/Sample for analysis of exclusions *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**

All home health stays (constructed from Medicare HH claims for Medicare certified HH agencies) beginning in 2010. Prior to applying exclusions, there were 3,069,749 such stays.

2b3.2 Analytic Method (*Describe type of analysis and rationale for examining exclusions, including exclusion related to patient preference)***:**

Denominator exclusion frequencies.

Exclusion criteria are based on either data requirements for calculating the measure (continuous enrollment in fee-for-service Medicare) or clear attribution of the measure to the home health agency (LUPAs and change of provider). We present the frequency of each type of exclusion.

Impact of planned hospitalization exclusion across agencies.

For the exclusion of planned hospitalizations from the measure numerator, we examine how planned hospitalizations were distributed across agencies and the impact on agency observed rates of ACH.

2b3.3 Results *(Provide statistical results for analysis of exclusions, e.g., frequency, variability, sensitivity analyses):*

126,480 stays (4%) were excluded because the patient was not continuously enrolled in fee-for-service Medicare during the numerator window (60 days after Stay_Start_Date) or until death.

275,342 stays (9%) were excluded because the first claim in the stay was a LUPAs.

37,733 stays (1%) were excluded because the beneficiary changed agencies during the numerator window.

116,757 stays (4%) were excluded because the patient was not continuously enrolled in fee-for-service Medicare for six month look-back period used to calculate HCCs.

22,621 acute care hospitalizations (0.9%) were not counted toward the measure numerator because they were determined to be planned hospitalizations.

The table below shows the effect of excluding planned hospitalizations on the ACH rates at agencies with at least 20 home health stays. The first column presents the percent of HH stays with planned hospitalizations, while the second column presents the percent of all ACH that are classified as planned. Though on average, 5.5% of ACH are planned hospitalizations, as many as 100% of hospitalizations at an agency are planned. While many agencies would not be affected by excluding planned hospitalization, a number would see marked decreases in their ACH rates. Excluding planned hospitalizations from the measure numerator results in a more equitable evaluation of agencies who serve a large number of patients with planned hospital care. Importantly, this exclusion does not decrease the variation in ACH rates.

Rate of Planned Hospitalizations at HHAs with at least 20 HH Stays

Planned Hospitalizations/all HH Stays		Planned Hospitalizations/all ACH
Number of Agencies	8,567	
Mean Rate	0.9%	5.5%
Standard Deviation	1.1%	7.7%
Minimum Rate	0.0%	0.0%
10th Percentile	0.0%	0.0%
25th Percentile	0.0%	0.0%
Median	0.7%	3.9%
75th Percentile	1.3%	7.4%
90th Percentile	2.1%	12.9%
95th Percentile	2.9%	18.2%
99th Percentile	4.7%	33.3%
Maximum Rate	12.7%	100.0%

2b4. Risk Adjustment Strategy. (*For outcome measures, adjustment for differences in case mix (severity) across measured entities was appropriately tested with adequate results.*)

2b4.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**

Of the 2,513,454 home health stays in 2010, a random 80% sample without replacement was chosen to calibrate the multinomial logit model and to estimate counterfactuals. The remaining 20% of the stays were used to cross-validate the model. The same multinomial logit model is used to predict both this measure and Emergency Department Use without Hospitalization.

2b4.2 Analytic Method (*Describe methods and rationale for development and testing of risk model or risk stratification including selection of factors/variables)***:**

Risk factors included in the model include prior care setting (e.g., outpatient emergency room, inpatient acute, psychiatric facility, etc.), health status (measured using HCCs and all remaining CCs), demographic information (measured using age-gender interactions), enrollment status (ESRD and disability), and interactions between these factors.

To determine which risk factors should be included in the risk adjustment model, a Wald test of joint restrictions was used. First, 700 bootstrap samples of a randomly chosen 80% sampling without replacement of the full data set were taken. A Wald test was performed which determined if the change in both outcomes associated with each covariate was significantly different from zero. Variables that were significant at a level of 0.05 for either outcomes in at least 70 percent of the bootstrap samples were included in the final risk adjustment model.

Calculation of counterfactuals to show impact of each risk factor. Each risk factor has an associated counterfactual value that can be interpreted as the population value of the measure if all patients in the population had the risk factor but had the observed distribution of all other risk factors. The counterfactual represents the relative impact of each risk factor on the outcome.

Goodness of fit statistics were then calculated for the calibrated model and the 20% sample was used for cross-validation.

2b4.3 Testing Results *(Statistical risk model: Provide quantitative assessment of relative contribution of model risk factors; risk model performance metrics including cross-validation discrimination and calibration statistics, calibration curve and risk decile plot, and assessment of adequacy in the context of norms for risk models. Risk stratification: Provide quantitative assessment of relationship of risk factors to the outcome and differences in outcomes among the strata)***:**

Among HH stays in 2010, the population average for Acute Care Hospitalization was 17.9%. The counterfactuals indicate the percentage point change in the outcome that the risk factor is associated with. For example, prior emergency department use is associated with a 4.12 percentage point increase in the risk of acute care hospitalization. This represents a 23.0% increase over the population average rate of 17.9%.

Prior Care Setting (omitted category: Community)	Marginal Effect
ED Use without Hospitalization	4.12
Short Term IP, 0-4 Days, Yale Medicine Cohort	4.45
Short Term IP, 0-4 Days, Yale Neurology Cohort	2.06

Short Term IP, 0-4 Days, Yale CRF Cohort		5.62	
Short Term IP, 0-4 Days, Yale Surgery Cohort		-0.66	
Short Term IP, 0-4 Days, Yale CVD Cohort		3.35	
Short Term IP, 4-9 Days, Yale Medicine Cohort		5.53	
Short Term IP, 4-9 Days, Yale Neurology Cohort		4.05	
Short Term IP, 4-9 Days, Yale CRF Cohort		6.94	
Short Term IP, 4-9 Days, Yale Surgery Cohort		1.18	
Short Term IP, 4-9 Days, Yale CVD Cohort		5.77	
Short Term IP, 9+ Days, Yale Medicine Cohort		7.88	
Short Term IP, 9+ Days, Yale Neurology Cohort		6.20	
Short Term IP, 9+ Days, Yale CRF Cohort			8.87
Short Term IP, 9+ Days, Yale Surgery Cohort		5.29	
Short Term IP, 9+ Days, Yale CVD Cohort			6.47
Inpatient, IRF		1.67	
Inpatient, LTCH		2.74	
Inpatient, Psych	3.12		
Skilled Nursing, 0-13 days		2.38	
Skilled Nursing, 14-41days		1.61	
Skilled Nursing, 42+ days		1.45	

mographics (omitted: 65-70 Male)	Marginal Effect
0-34 Years, Female	3.09
0-34 Years, Male	0.18
35-45, Female	1.05
35-45, Male	0.02
45-55, Female	-0.25
45-55, Male	-0.80
55-60, Female	-0.45
55-60, Male	-0.40
60-65, Female	-0.49
60-65, Male	-1.09
65-70, Female	-0.21
70-75, Female	0.01
70-75, Male	0.62

75-80, Female	0.51
75-80, Male	1.13
80-85, Female	1.27
80-85, Male	2.16
85-90, Female	2.29
85-90, Male	2.93
90-95, Female	3.14
90-95, Male	3.99
95+, Female	3.66
95+, Male	5.15

HCCs

Due to space constraints, counterfactuals for all HCCs are not reported. Included below are the marginal effects for several common HCCs. The attachment found in 2a1.17 includes the marginal effects for all HCCs.

HCC	Marginal Effect
Congestive Heart Failure	3.20%
Chronic Obstructive Pulmonary Disease	2.44%
Diabetes with Renal Manifestation	2.11%

Interactions included in the model are the interactions from the 2008 and 2012 HCC models that were statistically significant predictors of ACH and ED use.

Interaction	Marginal Effect
Artificial Openings * Pressure Ulcer	-2.03
Bacterial Pneumonia * Pressure Ulcer	-1.77
Cancer * Immune Disorders	-0.69
CHF * COPD	-0.92
COPD * CRF	1.51
Disabled * Chronic Pancreatitis	2.91
Disabled * Severe Hematological Disorders	2.16
Disabled * Alcohol Psychosis	3.76
Disabled * Alcohol Dependence	1.55
Disabled * Multiple Sclerosis	-1.77
Disabled * CHF	0.59
Disabled * Pressure Ulcer	0.89
Diabetes * CHF	-1.25
Diabetes * CVD	-0.64
Renal Failure * CHF	-1.49

Renal Failure * CHF * Diabetes	-1.33
Schizophrenia * CHF	2.15
Schizophrenia * Seizure	2.14
Sepsis * CRF	-1.89

Note that a number of interaction terms are negative. The total impact of having two condition (for example, CHF and COPD) is the sum of the coefficients on each condition and the coefficient of the interaction term. Negative coefficients on the interaction terms mean that while a patient with both conditions is more likely to have ACH than a patient with only one of the condition, just adding the effect of each condition will overstate ACH likelihood.

In order to test for over-fitting, a cross-validation method was used in which simple random sampling without replacement split the dataset into an 80% development sample comprising 2,010,764 stays and a 20% verification sample comprising 502,690 stays. The statistics computed to test over-fitting include c-statistics, a calibration statistic, and a discrimination statistic expressed in terms of predictive ability.

A version of the area under the receiver operating curve statistic, also known as the c-statistic, was calculated for each individual logit and for the model overall. This extension of the c-statistic averages pair-wise comparisons to reduce the multi-class form to the standard two-class case. The c-statistic measures the ability of a risk adjustment model to differentiate between outcomes without resorting to an arbitrary cutoff point. For ACH the c-statistic is 0.693, which is identical to the validation sample value of 0.693. The Total AUC for the model in the development sample is 0.654, which is similar to the verification sample value of 0.653.

To compute the calibration statistic, the vector of coefficients is estimated from the model on the development sample. These coefficients are then multiplied with the matrix of covariates from the validation sample to give a scalar linear predictor for the probability of an event for a given observation in the validation sample. A logistic regression is then estimated on the validation sample with an intercept and one covariate, the linear predictor. Values of the intercept far from 0 and values of the coefficient on Z far from 1 provide evidence of over-fitting. The calibration statistic for ACH produced an intercept of -0.005 and a coefficient of 0.996. With t-statistics of 0.598 and 0.656, these values are not significantly different from 0 and 1, respectively, at the 95% confidence level.

Cross-validation discrimination statistics were computed by looking at the difference between the 10th percentile of predicted probabilities for an event and contrasting this with the 90th percentile. In the development sample, the range of predicted probabilities for ACH was 8 to 31%. In the verification sample, this range was identical at 8 to 31%.Among HH stays in 2010, the population average for Acute Care Hospitalization was 17.9%. The counterfactuals indicate the percentage point change in the outcome that the risk factor is associated with. For example, prior emergency department use is associated with a 4.12 percentage point increase in the risk of acute care hospitalization. This represents a 23.0% increase over the population average rate of 17.9%.

Prior Care Setting (omitted category: Community)	Marginal Effect
ED Use without Hospitalization	4.12
Short Term IP, 0-4 Days, Yale Medicine Cohort	4.45
Short Term IP, 0-4 Days, Yale Neurology Cohort	2.06
Short Term IP, 0-4 Days, Yale CRF Cohort	5.62
Short Term IP, 0-4 Days, Yale Surgery Cohort	-0.66
Short Term IP, 0-4 Days, Yale CVD Cohort	3.35
Short Term IP, 4-9 Days, Yale Medicine Cohort	5.53
Short Term IP, 4-9 Days, Yale Neurology Cohort	4.05

Short Term IP, 4-9 Days, Yale CRF Cohort		6.94	
Short Term IP, 4-9 Days, Yale Surgery Cohort		1.18	
Short Term IP, 4-9 Days, Yale CVD Cohort		5.77	
Short Term IP, 9+ Days, Yale Medicine Cohort		7.88	
Short Term IP, 9+ Days, Yale Neurology Cohort		6.20	
Short Term IP, 9+ Days, Yale CRF Cohort			8.87
Short Term IP, 9+ Days, Yale Surgery Cohort		5.29	
Short Term IP, 9+ Days, Yale CVD Cohort			6.47
Inpatient, IRF		1.67	
Inpatient, LTCH		2.74	
Inpatient, Psych	3.12		
Skilled Nursing, 0-13 days		2.38	
Skilled Nursing, 14-41days		1.61	
Skilled Nursing, 42+ days		1.45	

Demographics (omitted: 65-70 Male)	Marginal Effect
0-34 Years, Female	3.09
0-34 Years, Male	0.18
35-45, Female	1.05
35-45, Male	0.02
45-55, Female	-0.25
45-55, Male	-0.80
55-60, Female	-0.45
55-60, Male	-0.40
60-65, Female	-0.49
60-65, Male	-1.09
65-70, Female	-0.21
70-75, Female	0.01
70-75, Male	0.62
75-80, Female	0.51
75-80, Male	1.13
80-85, Female	1.27
80-85, Male	2.16
85-90, Female	2.29

85-90, Male	2.93
90-95, Female	3.14
90-95, Male	3.99
95+, Female	3.66
95+, Male	5.15

HCCs

Due to space constraints, counterfactuals for all HCCs are not reported. Included below are the marginal effects for several common HCCs. The attachment found in 2a1.17 includes the marginal effects for all HCCs.

HCC	Marginal Effect
Congestive Heart Failure	3.20%
Chronic Obstructive Pulmonary Disease	2.44%
Diabetes with Renal Manifestation	2.11%

Interactions included in the model are the interactions from the 2008 and 2012 HCC models that were statistically significant predictors of ACH and ED use.

Interaction	Marginal Effect
Artificial Openings * Pressure Ulcer	-2.03
Bacterial Pneumonia * Pressure Ulcer	-1.77
Cancer * Immune Disorders	-0.69
CHF * COPD	-0.92
COPD * CRF	1.51
Disabled * Chronic Pancreatitis	2.91
Disabled * Severe Hematological Disorders	2.16
Disabled * Alcohol Psychosis	3.76
Disabled * Alcohol Dependence	1.55
Disabled * Multiple Sclerosis	-1.77
Disabled * CHF	0.59
Disabled * Pressure Ulcer	0.89
Diabetes * CHF	-1.25
Diabetes * CVD	-0.64
Renal Failure * CHF	-1.49
Renal Failure * CHF * Diabetes	-1.33
Schizophrenia * CHF	2.15

Schizophrenia * Seizure	2.14
Sepsis * CRF	-1.89

Note that a number of interaction terms are negative. The total impact of having two condition (for example, CHF and COPD) is the sum of the coefficients on each condition and the coefficient of the interaction term. Negative coefficients on the interaction terms mean that while a patient with both conditions is more likely to have ACH than a patient with only one of the condition, just adding the effect of each condition will overstate ACH likelihood.

In order to test for over-fitting, a cross-validation method was used in which simple random sampling without replacement split the dataset into an 80% development sample comprising 2,010,764 stays and a 20% verification sample comprising 502,690 stays. The statistics computed to test over-fitting include c-statistics, a calibration statistic, and a discrimination statistic expressed in terms of predictive ability.

A version of the area under the receiver operating curve statistic, also known as the c-statistic, was calculated for each individual logit and for the model overall. This extension of the c-statistic averages pair-wise comparisons to reduce the multi-class form to the standard two-class case. The c-statistic measures the ability of a risk adjustment model to differentiate between outcomes without resorting to an arbitrary cutoff point. For ACH the c-statistic is 0.693, which is identical to the validation sample value of 0.693. The Total AUC for the model in the development sample is 0.654, which is similar to the verification sample value of 0.653.

To compute the calibration statistic, the vector of coefficients is estimated from the model on the development sample. These coefficients are then multiplied with the matrix of covariates from the validation sample to give a scalar linear predictor for the probability of an event for a given observation in the validation sample. A logistic regression is then estimated on the validation sample with an intercept and one covariate, the linear predictor. Values of the intercept far from 0 and values of the coefficient on Z far from 1 provide evidence of over-fitting. The calibration statistic for ACH produced an intercept of -0.005 and a coefficient of 0.996. With t-statistics of 0.598 and 0.656, these values are not significantly different from 0 and 1, respectively, at the 95% confidence level.

Cross-validation discrimination statistics were computed by looking at the difference between the 10th percentile of predicted probabilities for an event and contrasting this with the 90th percentile. In the development sample, the range of predicted probabilities for ACH was 8 to 31%. In the verification sample, this range was identical at 8 to 31%.

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment: NA

2b5. Identification of Meaningful Differences in Performance. (*The performance measure scores were appropriately analyzed and discriminated meaningful differences in quality.*)

2b5.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**

Medicare certified agencies with at least 20 home health stays beginning between 1/1/2010 and 12/31/2010 and meeting the measure denominator criteria. There were 8,567 such agencies (85% of the 10,125 agencies with at least one stay beginning in 2010). The average size agency had 248 home health stays included in the measure numerator, while the median size agency had 102 home health stays.

2b5.2 Analytic Method *(Describe methods and rationale to identify statistically significant and practically/meaningfully differences in performance)***:**

The distribution of risk adjusted agency rates was analyzed to determine the inter-quartile range and the 90th vs. 10th percentile differences.

2b5.3 Results *(Provide measure performance results/scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:

Risk Adjusted Agency Rates for Agencies with At Least 20 Stays:

10th percentile 11.3%

25th percentile 14.3%

50th percentile 17.2%

75th percentile 19.9%

90th percentile 22.9%

Inter-quartile range (75th – 25th) = 19.9 – 14.3 = 5.6%

90th – 10th percentile = 22.9 – 11.3 = 11.6%

While the accounting for differences in case-mix (risk adjustment) narrows the distribution in rates of Acute Care Hospitalization somewhat, an agency at the 75th percentile still has a risk adjusted rate of Acute Care Hospitalization that is 5.6 percentage points higher than an agency at the 25th percentile, meaning the poorer quality agency experiences many more hospitalizations than the better agency.

2b6. Comparability of Multiple Data Sources/Methods. (*If specified for more than one data source, the various approaches result in comparable scores.*)

2b6.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**

NA - single data source

2b6.2 Analytic Method *(Describe methods and rationale for testing comparability of scores produced by the different data sources specified in the measure)***:**

NA - single data source

2b6.3 Testing Results *(Provide statistical results, e.g., correlation statistics, comparison of rankings; assessment of adequacy in the context of norms for the test conducted)***:**

NA - single data source

2c. Disparities in Care: H☐ M☐ L☐ I☐ NA☐ (*If applicable, the measure specifications allow identification of*

disparities.)
2c.1 If measure is stratified for disparities, provide stratified results *(Scores by stratified categories/cohorts)*: NA **2c.2 If disparities have been reported/identified (e.g., in 1b), but measure is not specified to detect disparities, please explain:**
2.1-2.3 Supplemental Testing Methodology Information:
Steering Committee: Overall, was the criterion, *Scientific Acceptability of Measure Properties*, met? **(*Reliability and Validity must be rated moderate or high*) Yes☐ No☐** **Provide rationale based on specific subcriteria:**
If the Committee votes No, STOP

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 by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition

If other: Directly from Medicare hospital claim dates

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.**

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.

Implementing claims-based measures such as this one requires extensive familiarity with Medicare claims and enrollment data. Because multiple types of claims are used, beneficiaries must be linked across claim types and enrollment files. Additionally, different types of claims suffer from different submission lags. Thus it is important to use the most up-to-date claims data possible in calculating claims based measures. For public reporting, this measure will be updated quarterly on a rolling basis. While the latest quarter in the observation window may have slightly lower rates of Acute Care Hospitalization due to claims delay, these events will be captured in the next quarterly update.

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)*.

Not applicable

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)
	Public Reporting Home Health Compare http://www.cms.gov/HomeHealthCompare/search.aspx Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Home Health Quality Initiative http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HomeHealthQualityInits/index.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

The Home Health Compare website is a federal government website managed by the Centers for Medicare and Medicaid Services (CMS). It provides information to consumers about the quality of care provided by Medicare-certified home health agencies throughout the nation. The measures reported on Home Health Compare includes all Medicare-certified agencies with at least 20 home health quality episodes. In CY 2014, there were 9,345 agencies the met the measure denominator criteria for reporting, representing 2,714,575 episodes of care nationally. (c.f. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HomeHealthQualityInits/Downloads/Quality-of-Patient-Care-Star-Ratings-Methodology-Report-updated-5-11-15.pdf)

CMS´ Home Health Quality Initiative "Outcome Quality Measure Report" provides all Medicare-certified home health agencies with opportunities to use outcome measures for outcome-based quality improvement. The report allows agencies to benchmark their performance against agencies across the state and nationally, as well as their own performance from prior time periods. All Medicare-certified home health agencies can access their Outcome Quality Measure Reports via CMS´ online CASPER system.

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?*)

Not applicable

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.*)

Not applicable

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**

Between the CY2011 and the CY2014 measurement period,
- At the agency level, mean risk-adjusted performance rate on this measure increased from 14.7 percent to 14.8 percent
- At the population level, the risk-adjusted performance rate has remained stable across all population groups

Additional information on the geographic area and number and percentage of accountable entities and patients included in this analysis is shown in the Attachment: Importance to Report

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.**
The small increase in the Acute Care Hospitalizations is likely to be attributed to agency activities in prior years that have resulted in performance plateaus or slight increases. In other words, agencies took the actions that had the most impact earlier and there has been a slight adverse change. There are anecdotal reports that the "average" home health care patient in the US is older than in the past and there are more women, living alone, with multiple chronic conditions. All of these factors increase the likelihood of rehospitalization.

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.**
A key issue in using this measure to accurately identify performance at the home health agency level regards attribution. Two decisions were made to assure proper attribution. First, the numerator window was synchronized to the length of home health prospective payment episodes (60 days) and home health stays beginning with low utilization payment episodes were excluded. This means that stays included in the measure were those in which the HHA was paid to provide appropriate home health care to the patient during the measurement period. Second, stays in which the patient changed home health providers during the numerator window were also excluded from measurement. Although provider switches often follow acute care utilization (ED use or hospitalization) and may reflect patient or caregiver dissatisfaction with the initial provider, we chose to exclude all HH stays with multiple providers during the numerator window. This ensures that agencies that do not have sufficient time to impact a patient's health are not penalized for that patient's outcomes.

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)
2380 : Rehospitalization During the First 30 Days of Home Health

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?

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**

The home health (HH) Acute Care Hospitalization (ACH) and Emergency Department Use (ED-Use) without Hospitalization measures are harmonized with the Rehospitalization measures (NQF numbers 2505 and 2380) and with CMS' Hospital-Wide All-Cause Unplanned Readmission (HWR) measure (NQF 1789) in the definition of unplanned hospitalizations . They differ from other post-acute hospital readmission measures, however, in the definition of eligible post-acute stays, in the risk adjustment approach, and by measuring emergency department use as an outcome. The differences arise due to the unique nature of home health care as a post-acute setting. The ACH and ED-Use measures were initially developed and later leveraged to construct the Rehospitalization measures by further restricting the ACH and ED-Use measures' eligible population by requiring a prior proximal inpatient hospital stay within 5 days from the start of HH. Finally, both pairs of measures are risk adjusted using patient-level predicted probabilities calculated from a multinomial logistic regression. Risk factors that are accounted for in both pairs of measures include demographics and health status as measured by both CMS' Hierarchical Condition Categories (HCCs) found on claims in the previous six months. The Rehospitalization measures leverage the prior proximal inpatient hospital claim to obtain the patient's Diagnosis Related Group (DRG) and also risk adjust for the Activities of Daily Living (ADL) fields on the Outcome and Assessment Information Set (OASIS) assessment of the initial home health stay. The risk-adjusted rates for the ACH and ED-Use measures are publicly reported. However, due to a large number of relatively small home health agencies treating previously hospitalized patients, the measure developer determined that reporting home health agencies' risk-adjusted rates could lead to misleading conclusions, since small home health agencies' risk-adjusted rates tend to be unstable. Therefore, the risk-adjusted rates for the home health Rehospitalization measures are publicly reported as categorizations (i.e., "Better than Expected", "Same as Expected", and "Worse than Expected"). While the Acute Care Hospitalization and Emergency Department Use without Hospitalization measures differ from other post-acute care measures in some regards, these differences arise from the unique nature of home care as well as from a desire for harmonization across home health quality measures.

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.)

Not applicable; there are no other measures that report acute care hospitalization rates for home health patients.

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:** Importance_to_Report_0171.docx

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services
Co.2 **Point of Contact:** Sophia, Chan, Sophia.Chan@cms.hhs.gov, 410-786-5050-
Co.3 **Measure Developer if different from Measure Steward:** Centers for Medicare & Medicaid Services
Co.4 **Point of Contact:** Theresa, White, Theresa.White@cms.hhs.gov, 410-786-2394-

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.

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2012
Ad.3 **Month and Year of most recent revision:** 10, 2015
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?** 06, 2016

Ad.6 **Copyright statement:**
Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:** Continuation of response to S2b:

Enrollment Database:
https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/IdentifiableDataFiles/DenominatorFile.html

Note: The Denominator File contains data on all Medicare beneficiaries enrolled and/or entitled in a given year. It is an abbreviated version of the Enrollment Data Base (EDB) (selected data elements).

Following is in response to NQF´s comment on February 12, 2016 that the ICD-10 code conversion is missing from the January 29, 2016 submission:
The specifications for measure 0171 depend on AHRQ CCS codes for specific denominator exclusions and HCCs and DRGs for risk adjustment. All three code groupings already have ICD-10 specifications; therefore, it was not necessary for the measure developer to construct a crosswalk from specific ICD-9 to ICD-10 codes.

Continuation of response 4.1:

Planned Use:
Proposed to be used in the HH VBP program as part of outcome measures:
http://www.wha.org/Data/Sites/1/reimbursement/2016HHAProposedRuleBrief.pdf

Response to 5.a1 under "Related and Competing Measures" tab:
Yes, the measure specifications are harmonized with 2380 Rehospitalization During the First 30 Days of Home Health.
No, the measure specifications are not harmonized with 1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR).

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 #: 0173
De.2. **Measure Title:** Emergency Department Use without Hospitalization During the First 60 Days of Home Health
Co.1.1. **Measure Steward:** Centers for Medicare & Medicaid Services
De.3. **Brief Description of Measure:** Percentage of home health stays in which patients used the emergency department but were not admitted to the hospital during the 60 days following the start of the home health stay.
1b.1. **Developer Rationale:** see attachment "Importance to Report" for a tabular presentation of these data

Rationale for this measure:
Emergency department use that does not lead to hospital admission may be for conditions that could have been treated in the outpatient setting or at home. LaCalle and Rabin1 identify persons over 65 years as among the "frequent users." Frequent users have been found to have a primary care physician but have trouble getting access to the primary care physician in a timely manner.

There is evidence for strategies that can be undertaken to reduce emergency department use without readmission among community dwelling elderly. Strategies include care coordination, primary care access (i.e. physician follow up), telehealth and a variety of home health care specific evidence-based strategies from the Quality Improvement Organizations such as medication reconciliation, care provision [frontloading visits], patient education strategies, falls prevention and other topics. Note that many of the latter QI strategies have not been studied via research in home health care patients.

(1) LaCalle E, Rabin E. Frequent users of emergency departments: the myths, the data, and the policy implications. Ann Emerg Med 2010, 56(1):42-48.

S.4. **Numerator Statement:** Number of home health stays for patients who have a Medicare claim for outpatient emergency department use and no claims for acute care hospitalization in the 60 days following the start of the home health stay.
S.7. **Denominator Statement:** Number of home health stays that begin during the 12-month observation period.
S.10. **Denominator Exclusions:** The following are excluded:

1) Home health stays for patients who are not continuously enrolled in fee-for-service Medicare for the 60 days following the start of the home health stay or until death.
2) Home health stays that begin with a Low Utilization Payment Adjustment (LUPA) claim.
3) Home health stays in which the patient receives service from multiple agencies during the first 60 days.
4) Home health stays for patients who are not continuously enrolled in fee-for-service Medicare for the 6 months prior to the home health stay.

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Administrative claims
S.26. **Level of Analysis:** Facility

IF Endorsement Maintenance – Original Endorsement Date: Mar 31, 2009 **Most Recent Endorsement Date:** Aug 10, 2012

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?** Not currently included in a composite measure.

Maintenance of Endorsement -- 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. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

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.

Summary of evidence:

- The developer describes multiple strategies that can be undertaken to reduce the use of emergency department without readmission among community dwelling elderly.
- Citing 5 studies published in peer-reviewed journals, evidence-based strategies identified are telehealth and increasing primary care access. Other interventions to reduce ED use with mixed results include geriatric nursing assessment, home care follow up, educational interventions and cost sharing.
- With ED visits identified as "gateway" encounters for hospital readmissions, the developer cited a recent 2013 study that found an instance in 1 hospital, where a readmission measure without a return to the ED portion, would miss 54% of all ED use following hospitalization. With 88% of unscheduled admission coming through the ED, there is definite potential for this measure that measures ED use AND inpatient readmission to identify potential areas for improving care.

Question for the Committee:

- *The developer attests the underlying evidence for the measure has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat discussion and vote on Evidence?*

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.

- The developer provide data on the distribution of risk-adjusted performance on this measure for 2011-2014.
- The average risk-adjusted performance is 11.9%, with the 25th percentile performance at 11.1% and the 75th performance at 12.5%.

Disparities

- Risk-adjusted measure scores based on gender and race did not show considerable differences in admission rates.
- Risk adjusted measure score by race/ethnicity showed the group demonstrate the Hispanic subgroup with a rate of 11.5% admissions. There is only a slight difference between the White and Black populations at 12% and 12.3% respectively.
- Risk adjusted measure score by disability status showed a 3.8% difference between people with disability and with none, at 14.9% and 11.1% respectively.

Questions for the Committee:

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

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

Committee pre-evaluation comments

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

1a. Evidence to Support Measure Focus

Comments: **This seems to be an appropriate look at utilization in the post-acute space. Whether it is appropriate to look at readmissions alone and ED visits alone vs. hospital utilization in general is an interesting topic.

1b. Performance Gap

Comments: **It does appear, at least at the tails, that there is a performance gap.

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

Comments: **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.

- This measure calculates the percentage home health stays in which patients used the emergency department but were not admitted to the hospital during the 60 days following the start of the home health stay
- This is a health outcome measure and the level of analysis is facility.
- The denominator is the number of home health stays that begin during the 12-month observation period.
- The numerator is the number of home health stays for patients who have a Medicare claim for outpatient emergency department use and no claims for acute care hospitalization in the 60 days following the start of the home health stay.
- The data sources for this measure may include Medicare Home Health Claims, Medicare Inpatient Claims, Medicare Part A and B claims, and the Medicare Enrollment Database (EDB).
- The measure's time window is 12 months.
- The measure is risk-adjusted using a statistical risk model (see details below)

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:

This measure was endorsed in the Care Coordination Phase 2 Project. All agencies with at least 20 home health stays beginning between 1/1/2010 and 12/31/2010 were included in the reliability analysis, because only information for agencies with at least 20 episodes is publicly reported. Of the 10,125 agencies with any home health stays in 2010, 8,567 agencies met the threshold for the Emergency Department Use without Hospitalization measure. Reliability testing methods and results are the same as provided at initial endorsement.

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

A beta-binomial distribution was fitted for all agencies. The beta-binomial method was developed for provider level measures reported as rates, and it allows one to calculate an agency level "reliability score," interpreted as the percent of variance due to the difference in measure score among providers.

Results of reliability testing

The developer notes that the distribution of national reliability scores shows that the majority of agencies have a reliability score greater than 0.818 and that this implies their performance can likely be distinguished from other agencies. This can be interpreted as approximately 82% of the variance is due to differences among providers, and 12% of the variance is due to measurement error or sampling uncertainty.

Guidance from the Reliability Algorithm

Question 1: Submitted specifications are precise, unambiguous, and complete.
Question 2: Empirical reliability testing was conducted using a beta-binomial distribution.
Question 3: Empirical validity testing of patient-level data was conducted.
Question 4: Reliability testing was conducted with computed performance measure scores for each measure
Question 5: The beta binomial method was appropriate for assessing the proportion of variability due to real differences among measured entities.
Question 6: The distribution of national reliability scores (percent of variance due to the difference in measure score among providers at the national level) shows the majority of agencies have a reliability score greater than 0.818, implying that their performance can likely be distinguished from other agencies (i.e., performance on this measure is unlikely to be due to measurement error or insufficient sample size, but is instead due to true differences between the agency and other agencies as it substantially exceeds within agency variation). The distribution of hospital referral region (HRR) reliability scores (percent of variance due to the difference in measure score among providers at the HRR level) for this measure also shows that at least 50% of agencies have a reliability score greater than 0.709, suggesting that between agency variation substantially exceeds within agency variation even at the HRR level.

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?*

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**

- This measure calculates the number of home health stays for patients who have a Medicare claim for outpatient emergency department use and no claims for acute care hospitalization in the 60 days following the start of the home health stay.
- As a rationale for measuring this health outcome, the developers suggest there is evidence for strategies that can be undertaken to reduce emergency department use without readmission among community dwelling elder including care coordination, primary care access (i.e. physician follow up), telehealth, medication reconciliation, care provision [frontloading visits], patient education strategies, and falls prevention.

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 developer did not conduct additional validity testing of the measure elements noting that CMS audits a sample of claims for acute inpatient hospitalizations as a part of the annual payment error calculations.
- During the previous review the Committee clarified that observation stays are included in this measure. The developer also clarified during the previous review that this is an all-cause measure.

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:

- Audit of claims data

Validity testing results:

- Of a 2010 audit of ***31,766 Part B claims***, there was 0.2% (801) claims that can patient record could not be found.

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?*
- *Other specific question of the validity testing?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developer notes that the exclusion criteria are based on either data requirements for calculating the measure (continuous enrollment in fee-for-service Medicare) or clear attribution of the measure to the home health agency (LUPAs and change of provider).
- To determine the impact of exclusions, the developer examined overall frequencies and proportions of the total cohort excluded for each exclusion criteria.
- The number and percentage of patients excluded for each criterion are as follows:
  - 126,480 stays (4%) were excluded because the patient was not continuously enrolled in fee-for-service Medicare during the numerator window or until death.
  - 275,342 stays (9%) were excluded because the first claim in the stay was a LUPAs.
  - 37,733 stays (1%) were excluded because the beneficiary changed agencies during the numerator window.
  - 116,757 stays (4%) were excluded because the patient was not continuously enrolled in fee-for-service Medicare for six month look-back period used to calculate hierarchical condition categories (HCCs).

Questions for the Committee:

- *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: **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

- This measure employs a multinomial logit model.
- Variables included in the model include prior care setting (e.g., outpatient emergency room, inpatient acute, psychiatric facility, etc.), health status (measured using HCCs and all remaining CCs), demographic information (measured using age-gender interactions), enrollment status (ESRD and disability), and interactions between these factors.
- To determine which risk factors should be included in the risk adjustment model, a Wald test of joint restrictions was used. Variables that were significant at a level of 0.05 for either outcomes in at least 70 percent of the bootstrap samples were included in the final risk adjustment model.
- The developer calculated counterfactuals to show the impact of each risk factor. Each risk factor has an associated counterfactual value that can be interpreted as the population value of the measure if all patients in the population had the risk factor but had the observed distribution of all other risk factors. The counterfactual represents the relative impact of each risk factor on the outcome.

Conceptual analysis of the need for SDS adjustment

- The developer found that while a recent review (Goodridge et al. Socioeconomic disparities in home health care service access and utilization: A scoping review 2012: *International J. Nursing Studies* 49(10); 1310-19) found that persons of lower socioeconomic status are not disadvantaged in terms of home health care service, findings from the literature support a link between SDS factors and emergency department use and hospital readmission.
- The developer notes that in the home health setting, the 60-day period for hospitalization occurs while the patient is living in their own home, increasing the likelihood that non-medical factors, including geographic location and economic resources, will have an impact on acute care use.
- Specifically, the developer found evidence in the literature of an impact of factors such as race and ethnicity, lower income, living alone, and lower levels of education on ED use and hospital readmission.

Empirical analysis of SDS factors:

- The developer notes that Data for race/ethnicity, disability status, rural location, sex, and Medicaid dual status were readily available through the enrollment database (EDB) and analyzed during the measure development process.
- The developer performed univariate analyses by race/ethnicity, disability status, rural location, and sex.
- The developer does not recommend controlling for SDS factors at this time.
- The results are summarized in the following tables:

Distribution of Performance Rates, by Sex

	2011		2012		2013		2014	
Sex	**Observed**	**Risk Adjusted**	**Observed**	**Risk Adjusted**	**Observed**	**Risk Adjusted**	**Observed**	**Risk Adjusted**
Male	10.6%	11.2%	11.1%	11.6%	11.5%	11.6%	11.8%	11.7%
Female	10.8%	11.6%	11.4%	11.9%	11.9%	12.1%	12.1%	12.1%

Distribution of Performance Rates, by Race/Ethnicity

	2011		2012		2013		2014	
Race/Ethnicity	**Observed**	**Risk**	**Observed**	**Risk**	**Observed**	**Risk**	**Observed**	**Risk**

		Adjusted		Adjusted		Adjusted		Adjusted
White	10.6%	11.5%	11.2%	11.8%	11.6%	11.9%	11.9%	12.0%
Black	12.3%	11.8%	12.9%	12.1%	13.3%	12.2%	13.7%	12.3%
Hispanic	9.6%	11.2%	10.1%	11.4%	10.4%	11.5%	10.3%	11.5%
Other	8.7%	10.8%	9.0%	11.1%	9.2%	11.1%	9.4%	11.1%

Distribution of Performance Rates, by Disability Status

Disability Status	2011		2012		2013		2014	
	Observed	Risk Adjusted	Observed	Risk Adjusted	Observed	Risk Adjusted	Observed	Risk Adjusted
Yes	13.4%	14.1%	14.1%	14.6%	14.5%	14.8%	14.7%	14.9%
No	10.0%	10.7%	10.5%	11.0%	10.9%	11.0%	11.2%	11.1%

Distribution of Performance Rates, by Urban/Rural Status

Urban/Rural Status	2011		2012		2013		2014	
	Observed	Risk Adjusted	Observed	Risk Adjusted	Observed	Risk Adjusted	Observed	Risk Adjusted
Urban	10.7%	11.5%	11.0%	11.8%	11.4%	11.9%	11.7%	12.0%
Rural	10.9%	11.4%	13.1%	11.8%	13.7%	12.0%	13.8%	12.1%

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?*

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

- The distribution of risk adjusted agency rates was analyzed to determine the inter-quartile range and the 90th vs. 10th percentile differences.
  - 10th percentile 4.9%
  - 25th percentile 7.1%
  - 50th percentile 9.4%
  - 75th percentile 11.9%
  - 90th percentile 14.6%
  - Inter-quartile range (75th – 25th) = 11.9 – 7.1 = 4.8%
  - 90th – 10th percentile = 14.6 – 4.9 = 9.7%

Question for the Committee:

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

2b6. Comparability of data sources/methods:

- N/A. The measure uses a single data source.

2b7. Missing Data

- N/A.

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

Comments: **The specifications are consistent with the evidence.

2a2. Reliability Testing

Comments: **The testing suggests that the data is reliable and most attributable to provider factors.

2b2. Validity Testing

Comments: **The test sample was adequate. There is sufficient validity. The 10/90 and 25/75 differences appear significant.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **No

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.

The developer states:

- Data elements are in defined fields in electronic claims and are generated and used by healthcare personnel during the provision of care.
- Claims tend to suffer from submission lags so it is important to use the most up-to-date claims data in calculating measure rate.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*

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

Committee pre-evaluation comments
Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **The data is from claims and can be feasibly used.

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**
OR
Planned use in an accountability program? ☐ **Yes** ☐ **No**

Accountability program details

- The measure is currently used for public reporting on Home Health Compare.
- The measure is currently used in quality improvement program – the Home Health Quality Initiative.
- The developer provides links to both web sites mentioned above.

Improvement results

- The developer states that between the 2011 and 2014 measurement period, the mean risk-adjusted performance rate on this measure increased from 11.4% to 11.9%, at the agency level. Rate has remained stable across all population groups at the population level.

Unexpected findings (positive or negative) during implementation

- The developer states that "The small increase in the Acute Care Hospitalizations is likely to be attributed to agency activities in prior years that have resulted in performance plateaus or slight increases. In other words, agencies took the actions that had the most impact earlier and there has been a slight adverse change. There are anecdotal reports that the "average" home health care patient in the US is older than in the past and there are more women, living alone, with multiple chronic conditions. All of these factors increase the likelihood of rehospitalization."

Potential harms

- The developer states that attribution at the home health agency level is a key issue in reporting this measure. The developer reported that for data reported, they excluded home health stays that involved multiple providers within the measurement window so as not to penalize agencies for a patient's outcome whose health care they have not had sufficient time to impact at the time of episode.

Feedback :

- Included in the 2012 MAP Care Coordination/Hospice and Palliative Care Family of Measures
- MAP included this measure on the Care Coordination family of measures due to its focus on patients receiving home care services and are subsequently hospitalized or visit the ED. MAP recommended that similar measures be developed for other post-acute and long-term care settings. In general, MAP prefers outcome measures over process and structural measures.

***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

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **This is being publically reported on Hospital Compare and this seems appropriate. How this measure is used, in conjunction with other measures such as readmission measures, is key to its utility.

Criterion 5: Related and Competing Measures

Related or competing measures

- 2505: Emergency Department Use without Hospital Readmission during the First 30 Days of Home Health
  - During the endorsement of #2505, the Standing Committee reviewed this issue. The previous findings were:
    - This measure competes directly with measure 0171 Acute Care Hospitalization—Percentage of Home Health stays in which patients were admitted to an acute care hospital during the 60 days following the start of the Home Health stay.
    - The measure specifications for measure 0173 and measure 2505 were harmonized along several measure dimensions, including Data source, Population, Denominator Exclusions, Numerator, and Risk Adjustment methodology.
    - The developers of this measure contended that there are differences that justify having two separate measures. patient admission to an emergency department (without hospitalization) during the 60 days following the start of Home Health stay, measure 2505 evaluates admission to the emergency department (without hospital readmission) within 30 days after starting Home Health care for patients who were recently discharged from an inpatient setting. Home Health agencies can track their performance on both utilization measures to gain an accurate picture of how much acute care is being used by their patients. Additionally, measure 2505 is an outcome measure that assesses the efficacy of care coordination as patients transition from inpatient acute care to outpatient Home Health services. In contrast, measure 0173 assesses the efficacy of clinical care provided to all patients, as indicated by rates of ED use after entry into Home Health services.
  - According to NQF guidance, since #0173 was not reviewed in the project, the committee did not make a recommendation with regards to these two competing measures.

Harmonization

- The developer states that this measure is "harmonized with the Rehospitalization measures (NQF numbers 2505 and 2380) and with CMS' Hospital-Wide All-Cause Unplanned Readmission (HWR) measure (NQF 1789) in the definition of unplanned hospitalizations."'
- The developer adds that this measure differs from other post-acute hospital readmission measures due to the unique nature of home health care as a post-acute setting.
- The developer states that this measure is risk adjusted using patient-level predicted probabilities calculated from a multinomial logistic regression. Risk factors that are accounted for include demographics and health status as measured by both CMS' Hierarchical Condition Categories (HCCs) found on claims in the previous six months. The differences of this measure from other post-acute care measures arise from the unique nature of home care as well as from a desire for harmonization across home health quality measures."

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0173

Measure Title: Emergency Department Use without Hospitalization

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: 1/29/2016

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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: Click here to name the health outcome

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

There are interventions that have been tested to reduce ED use (geriatric nursing assessment, home care follow up, telehealth increasing primary care accessibility, educational interventions and cost sharing) (1-5) with the strongest evidence for telehealth (4;5) and increasing primary care access (3). The remaining interventions have mixed results or are inconclusive (1-3). Note that two systematic reviews in this area are dated but provide comprehensive information (1;2).

(1) Aminzadeh F, Dalziel WB. Older adults in the emergency department: a systematic review of patterns of use, adverse outcomes, and effectiveness of interventions. Ann Emerg Med 2002 Mar;39(3):238-47.

(2) Hastings SN, Heflin MT. A systematic review of interventions to improve outcomes for elders discharged from the emergency department. Acad Emerg Med 2005 Oct;12(10):978-86.

(3) Flores-Mateo G, Violan-Fors C, Carrillo-Santisteve P, Peiro S, Argimon JM. Effectiveness of organizational interventions to reduce emergency department utilization: a systematic review. PLoS One 2012;7(5):e35903.

(4) Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K, et al. Home telehealth for chronic obstructive pulmonary disease: a systematic review and meta-analysis. J Telemed Telecare 2010;16(3):120-7.

(5) Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K, et al. Home telemonitoring for congestive heart failure: a systematic review and meta-analysis. J Telemed Telecare 2010;16(2):68-76.

Related question and response from previous submission:

1c.1 Structure-Process-Outcome Relationship *(Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process-health outcome; intermediate clinical outcome-health outcome)*:

Process-outcome (utilization).

There is evidence that there are strategies that can be undertaken to reduce the use of emergency department including contacting the primary care provider and/or home health care agency as well as telehealth interventions.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

There is very little home health care specific research available on this measure. A single study (Tzeng, 2011) reports on 31 patients from a single home health care agency where the study aim was to determine actions taken prior to seeking ED care. There were 35 ED visits made, as some patients had two ED visits during the study period. More than half the patients (57.1%) contacted their primary care provider prior to seeking ED care while less than one third also contacted the home health care agency (28.6%). Of the 35 ED visits, 20 resulted in admissions to the hospital while 15 visits resulted in the patient being sent home.

The Emergency Department serves an important function in post-acute care that has not been recognized with reports that one third of hospital revisits are missed if ED visits are not included (2). ED visits have been described as "gateway" encounters for hospital re-admissions. A recent study highlighted the importance of measuring both emergency department visits and inpatient readmissions after a hospital discharge. (3) In this study, focused on a single hospital, measures which evaluate readmission to the inpatient setting and do not include a return to the ED would miss 54 percent of all ED use after an inpatient stay. In addition, there has been a change in the proportion of ED visits that lead to hospital stays with more than 88% of unscheduled admissions coming through the ED (4). Thus, measuring ED use as well as inpatient readmission can help identify potential areas to improve care.

(1) Tzeng HM. Preliminary assessment of appropriateness of emergency care service use: actions taken and consultations obtained before emergency care presentation. Home Health Care Serv Q 2011 Jan;30(1):10-23.

(2) Steiner C, Barrett M, Hunter K. Hospital Readmissions and Multiple Emergency Department Visits, in Selected States, 2006-2007: Statistical Brief #90. 2006 Feb.

(3) Rising KL, White LF, Fernandez WG, Boutwell AE. Emergency Department Visits After Hospital Discharge: A Missing Part of the Equation. Ann Emerg Med 2013 Mar 28.

(4) Kocher KE, Dimick JB, Nallamothu BK. Changes in the source of unscheduled hospitalizations in the United States. Med Care 2013;51:689-698.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

Not Applicable

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

Not Applicable

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Not Applicable

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Not Applicable

Relevant question and response from previous submission:

1c.16 Quote verbatim, the specific guideline recommendation *(Including guideline # and/or page #)***:**

A search of the National Guideline Clearinghouse using the terms "emergency department" and "home care services" returned 2 guidelines, none were relevant.

A search of the National Guideline Clearinghouse for "emergency department" and "home care" returned 13 guidelines, none were relevant.

A search of the National Guideline Clearinghouse for "emergency department" and

"home health care" returned 3 guidelines, none were relevant.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

Not Applicable

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

Not Applicable

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

Not Applicable

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

Not Applicable

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

Not Applicable

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

Not Applicable

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

Not Applicable

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

Not Applicable

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

Not Applicable

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

Not Applicable

Relevant question and response from previous submission:

1c.15 Citations for Evidence other than Guidelines*(Guidelines addressed below)*:

Bourbeau J, Julien M, Maltais F, Rouleau M, Beaupre A, Begin R et al. Reduction of hospital utilization in patients with chronic obstructive pulmonary disease: a disease-specific self-management intervention. Arch Intern Med 2003; 163(5):585-591.

Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K. Home telehealth for diabetes management: a systematic review and meta-analysis. Diabetes Obes Metab 2009; 11(10):913-930.

Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K et al. Home telemonitoring for congestive heart failure: a systematic review and meta-analysis. J Telemed Telecare 2010; 16(2):68-76.

Tzeng HM. Preliminary assessment of appropriateness of emergency care service use: actions taken and consultations obtained before emergency care presentation. Home Health Care Serv Q 2011; 30(1):10-23.

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

NOT APPLICABLE

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

Not Applicable

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

Not Applicable

Related question and response from previous submission:

1c.6 Quality of Body of Evidence (*Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b) directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events):* Generally moderate to high quality for the telehealth studies. Telehealth is generally effective at reducing ED use for the diseases in which it has been studied and evaluated in systematic reviews. The studies are RCTs or observational studies with small to large sample sizes, depending on the study.

There are systematic reviews on ED use for community dwelling older people but they are dated (2002, 2005) and not reported here.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

Not Applicable

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

Not Applicable

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

Not Applicable

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Not Applicable

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

Not Applicable

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

Not Applicable

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

Not Applicable

1a.8 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.8.1 What process was used to identify the evidence?

Not Applicable

1a.8.2. Provide the citation and summary for each piece of evidence.

Not Applicable

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
0173_Evidence_Form_2016_2-23-16.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)*
see attachment "Importance to Report" for a tabular presentation of these data

Rationale for this measure:
Emergency department use that does not lead to hospital admission may be for conditions that could have been treated in the outpatient setting or at home. LaCalle and Rabin1 identify persons over 65 years as among the "frequent users." Frequent users have been found to have a primary care physician but have trouble getting access to the primary care physician in a timely manner.

There is evidence for strategies that can be undertaken to reduce emergency department use without readmission among community dwelling elderly. Strategies include care coordination, primary care access (i.e. physician follow up), telehealth and a variety of home health care specific evidence-based strategies from the Quality Improvement Organizations such as medication reconciliation, care provision [frontloading visits], patient education strategies, falls prevention and other topics. Note that many of the latter QI strategies have not been studied via research in home health care patients.

(1) LaCalle E, Rabin E. Frequent users of emergency departments: the myths, the data, and the policy implications. Ann Emerg Med 2010, 56(1):42-48.

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.*
see attachment "Importance to Report" for a tabular presentation of these data

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.**
see attachment "Importance to Report" for a tabular presentation of these data

No home health care specific performance data found

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.*
see attachment "Importance to Report" for a tabular presentation of these data

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.**
see attachment "Importance to Report" for a tabular presentation of these data

None found specific to home health care.

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, High resource use, Patient/societal consequences of poor quality

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.**

Steiner at al, in a Statistical Brief, report that 7.4% of home health patients have reported emergency department use without hospitalization. This information, although dated (2006-2007), is the most recent available. One report from Ontario (Costa et al) indicates that 41% of home care patients in Ontario had one or more ED visit in the six months following the initiation of home care.2 This information has to be interpreted cautiously as the Canadian home care system is very different from that of the United States.

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

Steiner C, Barrett M, Hunter K. Hospital Readmissions and Multiple Emergency Department Visits, in Selected States, 2006-2007: Statistical Brief #90. 2006 Feb.

Costa AP, Hirdes JP, Bell CM et al. Derivation and validation of the detection of indicators and vulnerabilities for emergency room trips scale for classifying the risk of emergency department use in frail community-dwelling older adults. J Am Geriatr Soc 2015;63:763-769.

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.)*

Not applicable

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):*

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

Care Coordination, Overuse

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

https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HomeHealthQualityInits/HHQIQualityMeasures.html

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:** Data_Dictionaries_ffs_inst_and_non-inst_claims.xls

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
There are no new changes made to the measure since the last annual maintenance which occurred on October 1, 2015. In the previous maintenance period two minor changes were made to the measures: (1) the title of the measure was changed to improve clarity and (2) recalibration of the risk adjustment model coefficients using data from January 1, 2013 to December 31, 2013.

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.
Number of home health stays for patients who have a Medicare claim for outpatient emergency department use and no claims for acute care hospitalization in the 60 days following the start of the home health stay.

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.)*
12 month data collection period, updated quarterly.

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.
The 60 day time window is calculated by adding 60 days to the "from" date in the first home health claim in the series of home health claims that comprise the home health stay. If the patient has any Medicare outpatient claims with any ER revenue center codes (0450-0459, 0981) during the 60 day window AND if the patient has no Medicare inpatient claims for admission to an acute care hospital (identified by the CMS Certification Number on the IP claim ending in 0001-0879, 0800-0899, or 1300-1399) during the 60 day window, then the stay is included in the measure numerator.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Number of home health stays that begin during the 12-month observation period.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

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)*
A home health stay is a sequence of home health payment episodes separated from other home health payment episodes by at least 60 days. Each home health payment episode is associated with a Medicare home health (HH) claim, so home health stays are constructed from claims data using the following procedure.

1.First, retrieve HH claims with a "from" date (FROM_DT) during the 12-month observation period or the 120 days prior to the beginning of the observation period and sequence these claims by "from" date for each beneficiary.
2.Second, drop claims with the same "from" date and "through" date (THROUGH_DT) and claims listing no visits and no payment. Additionally, if multiple claims have the same "from" date, keep only the claim with the most recent process date.
3.Third, set Stay_Start_Date(1) equal to the "from" date on the beneficiary's first claim. Step through the claims sequentially to determine which claims begin new home health stays. If the claim "from" date is more than 60 days after the "through" date on the previous claim, then the claim begins a new stay. If the claim "from" date is within 60 days of the "through" date on the previous claim, then the claim continues the stay associated with the previous claim.
4.Fourth, for each stay, set Stay_Start_Date(n) equal to the "from" date of the first claim in the sequence of claims defining that stay. Set Stay_End_Date(n) equal to the "through" date on the last claim in that stay. Confirm that Stay_Start_Date(n+1) –

Stay_End_Date(n) > 60 days for all adjacent stays.
5.Finally, drop stays that begin before the 12-month observation window.

Note the examining claims from the 120 days before the beginning of the 12-month observation period is necessary to ensure that stays beginning during the observation period are in fact separated from previous home health claims by at least 60 days.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
The following are excluded:

1) Home health stays for patients who are not continuously enrolled in fee-for-service Medicare for the 60 days following the start of the home health stay or until death.
2) Home health stays that begin with a Low Utilization Payment Adjustment (LUPA) claim.
3) Home health stays in which the patient receives service from multiple agencies during the first 60 days.
4) Home health stays for patients who are not continuously enrolled in fee-for-service Medicare for the 6 months prior to the home health stay.

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)*
Four types of home health stays are excluded from the measure denominator:

1.Home health stays for patients who are not continuously enrolled in fee-for-service Medicare for the 6 months prior to the start of the home health stay, for the 60 days following the start of the home health stay, or until death.
• Both enrollment status and beneficiary death date are identified using the Medicare Enrollment Database (EDB).

2.Home health stays that begin with a Low Utilization Payment Adjustment (LUPA) claim.
• Exclude the stay if LUPAIND = L for the first claim in the home health stay.

3.Home health stays in which the patient receives service from multiple agencies during the first 60 days.
• Define Initial_Provider = PROVIDER on the first claim in the home health stay.
• If Intial_Provider does not equal PROVIDER for a subsequent claim in the home health stay AND if the "from" date of the subsequent claim is within 60 days of Stay_Start_Date, then exclude the stay.

4.Home health stays for patients who are not continuously enrolled in fee-for-service Medicare for the 6 months prior to the start of the home health stay.
•Enrollment status is identified using the Medicare Enrollment Database (EDB).

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)*
Not applicable

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)*
Multinomial logit with outcomes of "No acute event", "Emergency Department use but no Hospitalization", and "Acute Care Hospitalization".

Risk factors include:

Prior Care Setting –
The main categories are community (i.e., no prior care setting), outpatient emergency room, inpatient-acute (IP-acute), inpatient rehabilitation facility (IRF), psychiatric facility, long-term care hospital (LTCH), and skilled nursing facility (SNF). The hierarchy of

setting is SNF, most recent inpatient stay (including IP-acute, IRF, LTCH, and psychiatric facility), outpatient ER, and community. Acumen used the five cohorts from the Yale Hospital-Wide All-Cause Risk Standardization Readmission Measure to segregate the IP-acute category. The five cohorts are:

1.Surgery/Gynecology: admissions likely cared for by surgical or gynecological teams, based on AHRQ procedure categories;
2.Cardiorespiratory: admissions treated by the same care teams with very high readmission rates, such as for pneumonia, chronic obstructive pulmonary disease, and heart failure;
3.Cardiovascular: admissions treated by separate cardiac or cardiovascular team in large hospitals, such as for acute myocardial infarctions;
4.Neurology: admissions for neurological conditions, such as stroke, that may be treated by a separate neurology team in large hospitals; and
5.Medicine: admissions for all other non-surgical patients.

These cohorts were designed to account for differences in readmission risk for surgical and non-surgical patients.

Finally, the IP-acute categories and the SNF category were further refined by length of stay. Each of the five IP-acute categories are separated into stays of length 0 to 3 days, 4 to 8 days, and 9 or more days, while the SNF categories are split into stays of length 0 to 13, 14 to 41, and 42 and more days. A patient cared for in both a skilled nursing facility and an inpatient hospital during the 30 days prior to starting home health care is included in the skilled nursing categories and not the inpatient categories. The length of stay is determined from the last inpatient or skilled nursing stay prior to beginning home health care.

Age and Gender Interactions –
Age is subdivided into 12 bins for each gender: aged 0-34, 35-44, 45-54, five-year age bins from 55 to 95, and a 95+ category. Using a categorical age variable allows the model to account for the differing effects of age and gender. Age is determined based on the patient's age at Stay_Start_Date.

CMS Hierarchical condition categories (HCCs) –
HCCs were developed for the risk adjustment model used in determining capitation payments to Medicare Advantage plans and are calculated using Part A and B Medicare claims. While the CMS-HHC model uses a full year of claims data to calculate HCCs, for these measures, we use only 6 months of data to limit the number of home health stays excluded due to missing HCC data. All 2012 HCCs and CCs that are not hierarchically ranked that were statistically significant predictors of ACH and ED use are included in the model.

Details of the CMS-HCC model and the code lists for defining the HCCs can be found here:
https://www.cms.gov/MedicareAdvtgSpecRateStats/06_Risk_adjustment.asp

A description of the development of the CMS-HCC model can be found here:
https://www.cms.gov/HealthCareFinancingReview/Downloads/04Summerpg119.pdf

ESRD and Disability Status –
Original End Stage Renal Disease (ESRD) and current ESRD status are included as risk factors. Original disabled status and male, and original disabled status and female, are also included. Medicare beneficiaries with ESRD or disabled status represent a fundamentally different health profile.

Interaction Terms –
All interaction terms included in the 2012 HCC risk adjustment models that were statistically significant predictors of ED Use and ACH were included. Interaction terms account for the additional effect two risk factors may have when present simultaneously, which is more than the additive effect of each factor separately.

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.
Provided in response box S.15a

S.15a. Detailed risk model specifications (*if not provided in excel or csv file at S.2b*)
https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HomeHealthQualityInits/Downloads/Claims-Based-ACH-and-ED-Use-Measures-Technical-Documentation-and-Risk-Adjustment.zip

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.)*
1.Construct Home Health Stays from HH Claims (see 2a1.7 for details)
2.Identify numerator window (60 days following Stay_Start_Date) for each stay and exclude stays for patients who are not continuously enrolled in fee-for-service Medicare during the numerator window or until patient death.
3.Exclude stays that begin with a LUPA or that involve a provider change during the numerator window
4.Link stays to enrollment data by beneficiary
5.Exclude stays for patients who are not continuously enrolled in fee-for-service Medicare during the 6 months prior to Stay_Start_Date.
6.Calculate demographic risk factors for each stay (age, gender, etc.) using enrollment data.
7.Link to Part A and Part B claims for 6 months prior to Stay_Start_Date for each beneficiary
8.Calculate prior care setting indicators, HCCs, and HCC interactions.
9.Link to Inpatient (IP) claims from Short Stay and Critical Access hospitals(excluding planned hospitalizations) for the numerator window (60 days following Stay_Start_Date) – see specifications for the home health Acute Care Hospitalization (NQF 0171) measure for details.
10.Set Hospital Admission indicator (Hosp_Admit = 1) if any IP claims are linked to the stay in step 9. These stays are not included in the ED Use without Hospitalization measure numerator.
11.Link to Outpatient claims with revenue center codes indicating Emergency Department use for the numerator window (60 days following Stay_Start_Date).
12.Set Outpatient ED Use indicator (OP_ED = 1) if any outpatient claims are linked to the stay in step 11.
13.Flag stays for inclusion in the measure numerator (ED_noHosp = 1) if OP_ED =1 and NOT Hosp_Admit = 1.
14.Using coefficients from the multinomial logit risk model and risk factors calculated in steps 6 and 8, calculate the predicted probability of being included in the measure numerator for each stay (Pred_ED_noHosp). Additionally calculate the average of Pred_ED_noHosp across all stays that are included in the measure denominator (not excluded in steps 3 or 5) and call this value National_pred_ED.
15.Calculate observed and risk adjusted rates for each home health agency (Initial_Provider):
a.Calculate the observed rate of Emergency Department Use without Hospitalization as the fraction all (non-excluded) HH Stays with that agency as Initial_Provider that are also included in the measure numerator (ED_noHosp = 1). Call the value Agency_obs_ED.
b.Calculate the agency predicted rate of Emergency Department use without Hospitalization by taking the average of Pred_ED_noHosp across all (non-excluded) stays with that agency as Initial_Provider. Call this value Agency_pred_ED.
c.Calculate the risk adjusted rate of Emergency Department use without Hospitalization using the following formula: Agency_riskadj_ED = National_pred_ED + (Agency_obs_ED – Agency_pred_ED). If an agency's calculated risk adjusted rate is negative, that agency will have a publicly reported rate of 0%

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 at measure-specific web page URL identified in S.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.
Not applicable

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.
Not applicable

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Not applicable

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

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.
Denominator: Medicare Home Health Claims
Numerator: Medicare Inpatient Claims
Exclusions: Medicare Home Health Claims, Medicare Enrollment Data
Risk Factors: Medicare Enrollment Data, Medicare Part A & B Claims

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)*
Available at measure-specific web page URL identified in S.1

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)*
Home Health
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.)*
Not applicable

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
0173_MeasureTesting_02-19-16.doc

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 0173	**NQF Project:** Care Coordination Project

2. RELIABILITY & 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. **(evaluation criteria)**

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate field. Supplemental materials may be referenced or attached in item 2.1. See guidance on measure testing.

2a2. Reliability Testing. (*Reliability testing was conducted with appropriate method, scope, and adequate demonstration of reliability*.)

2a2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**

All agencies with at least 20 home health stays beginning between 1/1/2010 and 12/31/2010 were included in the reliability analysis, because only information for agencies with at least 20 episodes is publicly reported. Of the 10,125 agencies with any home health stays in 2010, 8,567 agencies met the threshold for the Emergency Department Use without Hospitalization measure. For the national analysis, a beta-binomial distribution was fitted using all agencies. For the HHR (hospital referral region) analysis described below, separate beta-binomials were fitted for each of 306 HHRs, using only those agencies in the HHR. It is worth noting that even the agencies that are in HRRs with only two agencies have high reliability scores, because these small HRR agencies tend to service many home health patients relative to the rest of the country.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)***:**

Reliability analysis of this measure follows the beta-binomial method described in "The Reliability of Provider Profiling: A Tutorial" by John L. Adams. The beta-binomial method was developed for provider level measures reported as rates, and it allows one to calculate an agency level "reliability score," interpreted as the percent of variance due to the difference in measure score among providers. Thus, a reliability score of .80 signifies that 80% of the variance is due to differences among providers, and 20% of the variance is due to measurement error or sampling uncertainty. A high reliability score implies that performance on a measure is unlikely to be due to measurement error or insufficient sample size, but rather due to true differences between the agency and other agencies. Each agency receives an agency specific reliability score which depends on both agency size, agency performance on the measure, and measure variance for the relevant comparison group of agencies. The observed rates of ED use, rather than the risk adjusted rates, were used for this analysis as the

assumptions of this method are only appropriate for observed rates.

In addition to calculating reliability scores at the national level, we also calculated agency reliability scores at the level of hospital referral regions (HRRs), because the HRR grouping more adequately captures the types of comparisons health care consumers are likely to make. HRRs are region designations determined in the Dartmouth Atlas of Health Care study, and they represent regional health care markets for tertiary medical care that generally requires the service of a major referral center. They are aggregated hospital service areas (HSAs) and thus aggregated local health care markets. The HRRs are used to determine categories of sufficient size to make comparisons while still capturing the local set of HHA choices available to a beneficiary.

Reference: Adams, John L. The Reliability of Provider Profiling: A Tutorial. Santa Monica, CA: RAND Corporation, 2009. http://www.rand.org/pubs/technical_reports/TR653.

2a2.3 Testing Results *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**

Distribution of Within National Reliability Scores by Case Volume for Agencies with At Least 20 Stays

Number of Stays		Mean	Min	10th	25th	Median	75th	90th	Max
All Agencies	0.770	0.182	0.503	0.666	0.818	0.911	0.957	1.000	
20 to 99	0.610	0.182	0.382	0.503	0.616	0.721	0.804	1.000	
100 to 499	0.848	0.608	0.752	0.804	0.861	0.901	0.925	0.982	
> 500	0.958	0.898	0.934	0.945	0.959	0.973	0.983	0.999	

The distribution of national reliability scores (percent of variance due to the difference in measure score among providers at the national level) shows that the majority of agencies have a reliability score greater than 0.818, implying that their performance can likely be distinguished from other agencies (i.e., performance on this measure is unlikely to be due to measurement error or insufficient sample size, but is instead due to true differences between the agency and other agencies as it substantially exceeds within agency variation).

Distribution of Within HHR Reliability Scores by Case Volume for Agencies with At Least 20 Stays

Number of Stays		Mean	Min	10th	25th	Median	75th	90th	Max
All Agencies	0.674	0.030	0.373	0.528	0.709	0.845	0.918	1.000	
20 to 99	0.533	0.030	0.284	0.391	0.527	0.672	0.775	1.000	
100 to 499	0.728	0.075	0.505	0.637	0.758	0.849	0.901	0.977	
> 500	0.872	0.254	0.753	0.833	0.898	0.944	0.966	0.998	

The distribution of HRR reliability scores (percent of variance due to the difference in measure score among providers at the HRR level) for this measure also shows that at least 50% of agencies have a reliability score greater than 0.709, suggesting that between agency variation substantially exceeds within agency variation even at the HRR level.

2b. VALIDITY. Validity, Testing, including all Threats to Validity: H☐ M☐ L☐ I ☐

2b1.1 Describe how the measure specifications *(measure focus, target population, and exclusions)* **are consistent with**

the evidence cited in support of the measure focus (*criterion 1c)* **and identify any differences from the evidence:**

CMS chose to respecify the Emergency Department Use without Hospitalization measure with Medicare claims data to enhance the validity and reliability of this measure. The measure population is limited to fee-for-service (FFS) Medicare beneficiaries, ensuring that Medicare claims are filed for emergency department services the beneficiary receives. The measure numerator is a broad measure of utilization (Emergency Department Use) that can be cleanly identified using claims data. Because claims form the basis of Medicare payments, CMS invests significant resources in validating claims submissions prior to payment.

2b2. Validity Testing. (*Validity testing was conducted with appropriate method, scope, and adequate demonstration of validity.)*

2b2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**

As CMS audits a sample of claims for Part B services (including outpatient emergency department visits) as part of annual payment error calculations, additional validity testing of measure elements has not been conducted. The annual payment error calculation for 2010 involved a sample of Medicare claims that were then compared to medical records and included 31,766 claims Part B (and an additional 2,454 claims for Acute Inpatient Hospitalizations).

2b2.2 Analytic Method *(Describe method of validity testing and rationale; if face validity, describe systematic assessment)***:**

Review of 2010 Medicare CERT Report. Available at:
https://www.cms.gov/CERT/Downloads/Medicare_FFS_2010_CERT_Report.pdf

2b2.3 Testing Results *(Statistical results, assessment of adequacy in the context of norms for the test conducted; if face validity, describe results of systematic assessment)***:**

Of the sampled Part B claims, the patient record could not be found for 801 (or 0.2%) claims. It is possible that an extremely small fraction of claims represent care that did not occur, but this problem is clearly not widespread. 12.9% had some type of payment error with the bulk of these errors coming from insufficient documentation. It is possible that in some of these cases, reviewers could not determine that emergency department services were utilized or were medically necessary.

While the CERT report calculates the fraction of claims impacted by payment errors only for broad categories of payments and not by clinical setting, it does project the amount of improper payments by both type of error and clinical setting. The report estimates that $1.97 billion of improper payments to hospital outpatient departments resulted from insufficient documentation and $2.64 billion in payment errors resulted from any cause. For comparison, in 2009, total Medicare spending on hospital outpatient services was $34 billion. Thus errors impact only 7.7% of hospital outpatient payments.

POTENTIAL THREATS TO VALIDITY. (*All potential threats to validity were appropriately tested with adequate results.*)

2b3. Measure Exclusions. (*Exclusions were supported by the clinical evidence in 1c or appropriately tested with results demonstrating the need to specify them.*)

2b3.1 Data/Sample for analysis of exclusions *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**

All home health stays (constructed from Medicare HH claims for Medicare certified HH agencies) beginning in 2010. Prior to applying exclusions, there were 3,069,749 such stays.

2b3.2 Analytic Method (*Describe type of analysis and rationale for examining exclusions, including exclusion related to*

patient preference):

Frequencies. Exclusion criteria are based on either data requirements for calculating the measure (continuous enrollment in fee-for-service Medicare) or clear attribution of the measure to the home health agency (LUPAs and change of provider).

2b3.3 Results *(Provide statistical results for analysis of exclusions, e.g., frequency, variability, sensitivity analyses):*

126,480 stays (4%) were excluded because the patient was not continuously enrolled in fee-for-service Medicare during the numerator window (60 days after Stay_Start_Date) or until death.

275,342 stays (9%) were excluded because the first claim in the stay was a LUPAs.

37,733 stays (1%) were excluded because the beneficiary changed agencies during the numerator window.

116,757 stays (4%) were excluded because the patient was not continuously enrolled in fee-for-service Medicare for six month look-back period used to calculate HCCs.

2b4. Risk Adjustment Strategy. (*For outcome measures, adjustment for differences in case mix (severity) across measured entities was appropriately tested with adequate results.*)

2b4.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:

Of the 2,513,454 home health stays in 2010, a random 80% sample without replacement was chosen to calibrate the multinomial logit model and to estimate counterfactuals. The remaining 20% of the stays were used to cross-validate the model. One multinomial logit model was used to predict both this measure and Acute Care Hospitalization.

2b4.2 Analytic Method (*Describe methods and rationale for development and testing of risk model or risk stratification including selection of factors/variables)*:

Risk factors included in the model include prior care setting (e.g., outpatient emergency room, inpatient acute, psychiatric facility, etc.), health status (measured using HCCs and all remaining CCs), demographic information (measured using age-gender interactions), enrollment status (ESRD and disability), and interactions between these factors.

To determine which risk factors should be included in the risk adjustment model, a Wald test of joint restrictions was used. First, 700 bootstrap samples of a randomly chosen 80% sampling without replacement of the full data set were taken. A Wald test was performed which determined if the change in both outcomes associated with each covariate was significantly different from zero. Variables that were significant at a level of 0.05 for both outcomes in at least 70 percent of the bootstrap samples were included in the final risk adjustment model.

Calculation of counterfactuals to show impact of each risk factor. Each risk factor has an associated counterfactual value that can be interpreted as the population value of the measure if all patients in the population had the risk factor but had the observed distribution of all other risk factors. The counterfactual represents the relative impact of each risk factor on the outcome.

Goodness of fit statistics were then calculated for the calibrated model and the 20% sample was used for cross-validation.

2b4.3 Testing Results *(Statistical risk model: Provide quantitative assessment of relative contribution of model risk factors; risk model performance metrics including cross-validation discrimination and calibration statistics, calibration curve and risk decile plot, and assessment of adequacy in the context of norms for risk models. Risk stratification: Provide quantitative*

assessment of relationship of risk factors to the outcome and differences in outcomes among the strata):

Among HH stays in 2010, the population average for Emergency Department Use without Hospitalization was 9.7%. The counterfactuals indicate the percentage point change in the outcome that the risk factor is associated with. For example, emergency department use is associated with a 5.74 percentage point increase in the risk of ED use. This represents a 59.2% increase over the average rate of 9.7%.

Prior Care Setting (omitted category: Community)	Marginal Effect
ED Use without Hospitalization	5.74
Short Term IP, 0-4 Days, Yale Medicine Cohort	2.17
Short Term IP, 0-4 Days, Yale Neurology Cohort	1.78
Short Term IP, 0-4 Days, Yale CRF Cohort	2.28
Short Term IP, 0-4 Days, Yale Surgery Cohort	0.63
Short Term IP, 0-4 Days, Yale CVD Cohort	2.96
Short Term IP, 4-9 Days, Yale Medicine Cohort	1.40
Short Term IP, 4-9 Days, Yale Neurology Cohort	1.15
Short Term IP, 4-9 Days, Yale CRF Cohort	0.84
Short Term IP, 4-9 Days, Yale Surgery Cohort	0.51
Short Term IP, 4-9 Days, Yale CVD Cohort	1.62
Short Term IP, 9+ Days, Yale Medicine Cohort	0.86
Short Term IP, 9+ Days, Yale Neurology Cohort	1.41
Short Term IP, 9+ Days, Yale CRF Cohort	0.53
Short Term IP, 9+ Days, Yale Surgery Cohort	0.79
Short Term IP, 9+ Days, Yale CVD Cohort	1.26
Inpatient, IRF	-0.46
Inpatient, LTCH	-0.02
Inpatient, Psych	3.26
Skilled Nursing, 0-13 days	0.41
Skilled Nursing, 14-41days	-0.23
Skilled Nursing, 42+ days	0.05

Demographics (omitted: 65-70 Male)	Marginal Effect
0-34 Years, Female	8.06
0-34 Years, Male	5.17
35-45, Female	6.59
35-45, Male	4.45

45-55, Female	4.23
45-55, Male	3.11
55-60, Female	2.32
55-60, Male	1.54
60-65, Female	1.23
60-65, Male	0.80
65-70, Female	0.33
70-75, Female	0.03
70-75, Male	-0.07
75-80, Female	0.25
75-80, Male	-0.08
80-85, Female	0.53
80-85, Male	0.14
85-90, Female	0.69
85-90, Male	0.52
90-95, Female	0.98
90-95, Male	0.95
95+, Female	1.05
95+, Male	1.58

HCCs – due to space constraints, counterfactuals for all HCCs are not reported. The marginal effects of several common HCCs are shown below. A full listing of the marginal effects of each HCC can be found in the attachment in 2a1.17.

HCC	Marginal Effect
Asthma	0.81%
Urinary Tract Infection	0.76%
Disorders of Vertebrae/Spinal Discs	0.51%

Interactions included in the model are the interaction from the 2008 and 2012 HCC model that were statistically significant predictors of ED use and ACH.

Interaction	Marginal Effect
Artificial Openings * Pressure Ulcer	-0.01
Bacterial. Pneumonia * Pressure Ulcer	-0.45
Cancer * Immune Disorders	0.06
CHF * COPD	0.12
COPD * CRF	-0.44
Disabled * Chronic Pancreatitis	0.57

Disabled * Severe Hematological Disorders	-1.14
Disabled * Alcohol Psychosis	1.01
Disabled * Alcohol Dependence	1.07
Disabled * Multiple Sclerosis	-0.93
Disabled * CHF	-0.48
Disabled * Pressure Ulcer	-0.77
Diabetes * CHF	0.15
Diabetes * CVD	0.12
Renal Failure * CHF	-0.07
Renal Failure * CHF * Diabetes	-0.05
Schizophrenia * CHF	0.23
Schizophrenia * Seizure	0.29
Sepsis * CRF	0.07

Note that a number of interaction terms are negative. The total impact of having two condition (for example, CHF and Disabled) is the sum of the coefficients on each condition and the coefficient of the interaction term. Negative coefficients on the interaction terms mean that while a patient with both conditions is more likely to have ED Use than a patient with only one of the condition, just adding the effect of each condition will overstate ED Use likelihood.

In order to test for over-fitting, a cross-validation method was used in which simple random sampling without replacement split the dataset into an 80% development sample comprising 2,010,764 stays and a 20% verification sample comprising 502,690 stays. The statistics computed to test over-fitting include c-statistics, a calibration statistic, and a discrimination statistic expressed in terms of predictive ability.

A version of the area under the receiver operating curve statistic, also known as the c-statistic, was calculated for each individual logit and for the model overall. This extension of the c-statistic averages pair-wise comparisons to reduce the multi-class form to the standard two-class case. The c-statistic measures the ability of a risk adjustment model to differentiate between outcomes without resorting to an arbitrary cutoff point. For ED use, the c-statistic for the development sample is 0.632, which is comparable to the validation sample value of 0.631. The Total AUC for the model in the development sample is 0.654, which is similar to the verification sample value of 0.653.

To compute the calibration statistic, the vector of coefficients is estimated from the model on the development sample. These coefficients are then multiplied with the matrix of covariates from the validation sample to give a scalar linear predictor for the probability of an event for a given observation in the validation sample. A logistic regression is then estimated on the validation sample with an intercept and one covariate, the linear predictor. Values of the intercept far from 0 and values of the coefficient on Z far from 1 provide evidence of over-fitting. In our validation sample, the calibration statistic for ED use produced an intercept of -0.017 and a coefficient of 0.992. With t-statistics of 0.854 and 0.819, these values are not significantly different from 0 and 1, respectively, at the 95% confidence level.

Cross-validation discrimination statistics were computed by looking at the difference between the 10th percentile of predicted probabilities for an event and contrasting this with the 90th percentile. In the development sample, the range of

predicted probabilities for ED Use was 5 to 14%. In the verification sample, this was a similar 6 to 14%.

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment:

2b5. Identification of Meaningful Differences in Performance. (*The performance measure scores were appropriately analyzed and discriminated meaningful differences in quality.*)

2b5.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**

Medicare certified agencies with at least 20 home health stays beginning between 1/1/2010 and 12/31/2010 and meeting the measure denominator criteria. There were 8,567 such agencies (85% of the 10,125 agencies with at least one stay beginning in 2010). The average size agency had 248 home health stays included in the measure numerator, while the median size agency had 102 home health stays.

2b5.2 Analytic Method *(Describe methods and rationale to identify statistically significant and practically/meaningfully differences in performance)***:**

The distribution risk-adjusted agency rates was analyzed to determine the inter-quartile range and the 90th vs. 10th percentile differences.

2b5.3 Results *(Provide measure performance results/scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:

Risk Adjusted Agency Rate Distribution for Agencies with At Least 20 Stays:

10th percentile 4.9%

25th percentile 7.1%

50th percentile 9.4%

75th percentile 11.9%

90th percentile 14.6%

Inter-quartile range (75th – 25th) = 11.9 – 7.1 = 4.8%

90th – 10th percentile = 14.6 – 4.9 = 9.7%

An agency at the 75th percentile has a risk adjusted rate of Emergency Department Use without Hospitalization that is 4.8 percentage points higher than that of an agency at the 25th percentile, while an agency at the 90th percentile has a risk adjusted rate that is 9.7 percentage points higher than the rate of hospitalization of an agency at the 10th percentile.

2b6. Comparability of Multiple Data Sources/Methods. (*If specified for more than one data source, the various approaches result in comparable scores.*)

2b6.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**

NA - single data source

2b6.2 Analytic Method *(Describe methods and rationale for testing comparability of scores produced by the different data sources specified in the measure)*:

NA - single data source

2b6.3 Testing Results *(Provide statistical results, e.g., correlation statistics, comparison of rankings; assessment of adequacy in the context of norms for the test conducted)*:

NA - single data source

2c. Disparities in Care: H☐ M☐ L☐ I☐ NA☐ (*If applicable, the measure specifications allow identification of disparities.*)

2c.1 If measure is stratified for disparities, provide stratified results *(Scores by stratified categories/cohorts)*: NA - no stratification

2c.2 If disparities have been reported/identified (e.g., in 1b), but measure is not specified to detect disparities, please explain:

NA

2.1-2.3 Supplemental Testing Methodology Information:

Steering Committee: Overall, was the criterion, *Scientific Acceptability of Measure Properties*, met?

(*Reliability and Validity must be rated moderate or high*) Yes☐ No☐

Provide rationale based on specific subcriteria:

If the Committee votes No, STOP

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 by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition

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.**

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.

Implementing claims-based measures such as this one requires extensive familiarity with Medicare claims and enrollment data. Because multiple types of claims are used, beneficiaries must be linked across claim types and enrollment files. Additionally, different types of claims suffer from different submission lags. Thus it is important to use the most up-to-date claims data possible in calculating claims based measures. For public reporting, this measure will be updated quarterly on a rolling basis. While the latest quarter in the observation window may have slightly lower rates of ED use without Hospitalization, due to claims delay, these events will be captured in the next quarterly update.

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)*.

Not applicable

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)
	Public Reporting Home Health Compare http://www.cms.gov/HomeHealthCompare/search.aspx Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Home Health Quality Initiative http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HomeHealthQualityInits/index.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

The Home Health Compare website is a federal government website managed by the Centers for Medicare and Medicaid Services (CMS). It provides information to consumers about the quality of care provided by Medicare-certified home health agencies throughout the nation. In CY 2014, there were 11,614 agencies the met the measure denominator criteria representing 2,714,575 episodes of care nationally. 80.5 percent of these home health agencies reported 20 or more episodes and were featured on the Home Health Compare website if they met the business rule criteria.
CMS´ Home Health Quality Initiative "Outcome Quality Measure Report" provides all Medicare-certified home health agencies with opportunities to use outcome measures for outcome-based quality improvement. The report allows agencies to benchmark their performance against agencies across the state and nationally, as well as their own performance from prior time periods. All Medicare-certified home health agencies can access their Outcome Quality Measure Reports via CMS´ online CASPER system.

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?*)
Not applicable

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.*)
Not applicable

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**

Between the CY2011 and the CY2014 measurement period,

- At the agency level, mean risk-adjusted performance rate on this measure increased steadily from 11.4 percent to 11.9 percent
- At the population level, the risk-adjusted performance rate has remained stable across all population groups

Additional information on the geographic area and number and percentage of accountable entities and patients included in this analysis is shown in the Attachment: Importance to Report

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.**
The small increase in the ED use without hospitalization is likely to be attributed to agency activities in prior years that have resulted in performance plateaus or slight increases. In other words, agencies took the actions that had the most impact earlier and there has been a slight adverse change. There are anecdotal reports that the "average" home health care patient in the US is older than in the past and there are more women, living alone, with multiple chronic conditions. All of these factors increase the likelihood of rehospitalization.

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.**
A key issue in using this measure to accurately identify performance at the home health agency level regards attribution. Two decisions were made to assure proper attribution. First, the numerator window was synchronized to the length of home health prospective payment episodes (60 days) and home health stays beginning with low utilization payment episodes were excluded. This means that stays included in the measure were those in which the HHA was paid to provide appropriate home health care to the patient during the measurement period. Second, stays in which the patient changed home health providers during the numerator window were also excluded from measurement. Although provider switches often follow acute care utilization (ED use or hospitalization) and may reflect patient or caregiver dissatisfaction with the initial provider, we chose to exclude all HH stays with multiple providers during the numerator window. This ensures that agencies that do not have sufficient time to impact a patient's health are not penalized for that patient's outcomes.

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)
2505 : Emergency Department Use without Hospital Readmission During the First 30 Days of Home Health

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?

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
The home health (HH) Acute Care Hospitalization (ACH) and Emergency Department Use (ED-Use) without Hospitalization measures are harmonized with the Rehospitalization measures (NQF numbers 2505 and 2380) and with CMS' Hospital-Wide All-Cause Unplanned Readmission (HWR) measure (NQF 1789) in the definition of unplanned hospitalizations . They differ from other post-acute hospital readmission measures, however, in the definition of eligible post-acute stays, in the risk adjustment approach, and by measuring emergency department use as an outcome. The differences arise due to the unique nature of home health care as a post-acute setting. The ACH and ED-Use measures were initially developed and later leveraged to construct the Rehospitalization measures by further restricting the ACH and ED-Use measures' eligible population by requiring a prior proximal inpatient hospital stay within 5 days from the start of HH. Finally, both pairs of measures are risk adjusted using patient-level predicted probabilities calculated from a multinomial logistic regression. Risk factors that are accounted for in both pairs of measures include demographics and health status as measured by both CMS' Hierarchical Condition Categories (HCCs) found on claims in the previous six months. The Rehospitalization measures leverage the prior proximal inpatient hospital claim to obtain the patient's Diagnosis Related Group (DRG) and also risk adjust for the Activities of Daily Living (ADL) fields on the Outcome and Assessment Information Set (OASIS) assessment of the initial home health stay. The risk-adjusted rates for the ACH and ED-Use measures are publicly reported. However, due to a large number of relatively small home health agencies treating previously hospitalized patients, the measure developer determined that reporting home health agencies' risk-adjusted rates could lead to misleading conclusions, since small home health agencies' risk-adjusted rates tend to be unstable. Therefore, the risk-adjusted rates for the home health Rehospitalization measures are publicly reported as categorizations (i.e., "Better than Expected", "Same as Expected", and "Worse than Expected"). While the Acute Care Hospitalization and Emergency Department Use without Hospitalization measures differ from other post-acute care measures in some regards, these differences arise from the unique nature of home care as well as from a desire for harmonization across home health quality measures.

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.)
Not applicable; there are no other measures that report emergency department use without hospitalization rates for home health patients.The Home Health Acute Care Hospitalization Measure (NQF# 0171)is specified so that it reports all acute care hospitalizations during the 60-day period following the beginning of the home health stay. This measure is specified so that it only reports emergent care use for patients that are not admitted to an acute care setting. No other measures report Emergent Care use among home health patients.

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:** Importance_to_Report_0173.docx

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services
Co.2 **Point of Contact:** Sophia, Chan, Sophia.Chan@cms.hhs.gov, 410-786-5050-
Co.3 **Measure Developer if different from Measure Steward:** Centers for Medicare & Medicaid Services
Co.4 **Point of Contact:** Theresa, White, Theresa.White@cms.hhs.gov, 410-786-2394-

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.

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2012
Ad.3 **Month and Year of most recent revision:** 10, 2015
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?** 06, 2016

Ad.6 **Copyright statement:**
Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:** Continuation of response to S2b:

Enrollment Database:
https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/IdentifiableDataFiles/DenominatorFile.html

Note: The Denominator File contains data on all Medicare beneficiaries enrolled and/or entitled in a given year. It is an abbreviated version of the Enrollment Data Base (EDB) (selected data elements).

Following is in response to NQF´s comment on February 12, 2016 that the ICD-10 code conversion is missing from the January 29, 2016 submission:
The specifications for measure 0173 depend on AHRQ CCS codes for specific denominator exclusions and HCCs and DRGs for risk adjustment. All three code groupings already have ICD-10 specifications; therefore, it was not necessary for the measure developer to construct a crosswalk from specific ICD-9 to ICD-10 codes.

Continuation of response 4.1:

Planned Use:
Proposed to be used in the HH VBP program as part of outcome measures:
http://www.wha.org/Data/Sites/1/reimbursement/2016HHAProposedRuleBrief.pdf

Response to 5a.1 in "Related and Competing Measures" tab
Yes, the measure specifications are harmonized with 2505 Emergency Department Use without Hospital Readmission During the First 30 Days of Home Health.
No, the measure specifications are not harmonized with 1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR).