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

Purple text represents the responses from measure developers.

Red text denotes developer information that has changed since the last measure evaluation review.

Brief Measure Information

NQF #: 0330

Corresponding Measures:

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

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

De.3. Brief Description of Measure: The measure estimates a hospital-level risk-standardized readmission rate (RSRR) for patients discharged from the hospital with a principal diagnosis of heart failure (HF). Readmission is defined as unplanned readmission for any cause within 30 days of the discharge date for the index admission. Readmissions are classified as planned and unplanned by applying the planned readmission algorithm. The target population is patients aged 65 and over. The Centers for Medicare & Medicaid Services (CMS) annually reports the measure for patients who are 65 years or older and are enrolled in fee-for-service (FFS) Medicare and hospitalized in non-federal hospitals or are patients hospitalized in Veterans Health Administration (VA) facilities.

1b.1. Developer Rationale: The goal of this measure is to improve patient outcomes. 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. Readmissions following HF are influenced by 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; several studies have demonstrated that appropriate, timely, and high-quality treatment can 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.

By providing patients, physicians, hospitals, and policy makers with information about hospital risk-standardized readmission rates following hospitalization for HF, HF readmission is a priority area for outcomes measure development. 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 readmissions. We define readmissions as any inpatient acute care admission, with the exception of certain planned readmissions, within 30 days from the date of discharge from an index admission with a principal discharge diagnosis of HF in patients 65and older. 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.

Additional details are provided in S.5 Numerator Details.

S.6. Denominator Statement: The cohort includes admissions for patients aged 65years 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 is publicly reported by CMS for those patients 65 years and older who are Medicare FFS or VA beneficiaries admitted to non-federal or VA hospitals, respectively.

Additional details are provided in S.7 Denominator Details

S.8. Denominator Exclusions: The 30-day HF readmission measure excludes index admissions for patients:

1. Without at least 30 days of post-discharge enrollment in Medicare FFS (in the case of patients who are not VA beneficiaries);
2. Discharged against medical advice (AMA);
3. Admitted within 30 days of a prior index admission for HF; and
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.

De.1. Measure Type: Outcome

S.17. Data Source: Claims, Enrollment Data, Other

S.20. Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: May 15, 2008 **Most Recent Endorsement Date:** Dec 09, 2016

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, all-cause, 30-day, risk-standardized mortality (RSMR) following HF hospitalization.

Preliminary Analysis: Maintenance of Endorsement

To maintain NQF endorsement, endorsed measures are evaluated periodically to ensure that the measures still meet 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 outcome measure include providing empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service; if these data not available, data demonstrating wide variation in performance, assuming the data are from a robust number of providers and results are not subject to systematic bias. For measures derived from patient report, evidence also should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.

Summary of prior review in 2016

- Developer asserted that readmissions following HF are influenced by complex and critical aspects of care (aspects listed below) and several studies have demonstrated that appropriate, timely, and high-quality treatment can contribute to patient outcomes.
    - communication between providers
    - prevention of and response to complications, patient safety
    - coordinated transitions to the outpatient environment
    - patient safety
- The developer cited a 10 per 1,000 incidence rate of heart failure (HF) in patients 65 years and older, demonstrating the projected continuation as one of the most common discharge diagnoses among the elderly.
- The developer suggested 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.
- Developer offered a depiction of a logic model connecting care processes and elements of patient care with patient outcomes.
- Measuring and reporting readmission rates will inform healthcare providers and facilities about opportunities to improve care, strengthen incentives for quality improvement, increase transparency for consumers, and ultimately improve the quality of care received by Medicare patients.

Changes to evidence from last review

☐ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☒ **The developer provided updated evidence for this measure:**

Updates:

- The developer indicated the assessment of standardized readmissions rates following hospitalization for HF as a priority area for outcomes measure development. Furthermore, the developer presented findings that demonstrate the lifetime risk of HF estimated at 1 in 5 at 40 years of age, and the prevalence in the aging US population is expected to increase by 46% by 2030 (Heidenreich 2013).
- The developer cited an American Heart Association policy statement on the impact of heart failure in the United States stating that the total direct medical costs of HF were estimated at $30.7 billion in 2012 and the projected increase would be approximately 127% to $69.7 billion by 2030.
- The developer cited a report that found transitional care models that prioritize effective collaboration and communication within and across providers/facilities and patient tracking demonstrate significant hospital readmissions reductions after AMIs. Central to the reference is the pilot transitional care

program (Transitions Across Care Settings [TRACS]), which used the Coleman Care Transitions Intervention model and aimed to reduce 30-day readmissions to lower than the national averages for an initial target population of inpatients with pneumonia, congestive heart failure, and acute myocardial infarction diagnoses. During the course of the TRACS pilot, readmission rates were 0% for acute myocardial infarction and 7.1% for congestive heart failure, both of which are significantly lower than the national average.

- Developer added two items to the first two components of its logic model:
    - Delivery of timely, high-quality care
    - Improved healthcare support and management

Question for the Committee:

- Does the Standing Committee agree that the evidence demonstrates interventions that providers can implement towards the improvement of patient outcomes and measure performance?

Guidance from the Evidence Algorithm

BOX 1: Measure an outcome (Yes) → BOX 2: Empirical evidence to support the relationship to a at least one structure or process (Yes) → PASS

Preliminary rating for evidence: ☒ Pass ☐ No Pass

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

Maintenance measures – increased emphasis on gap and variation

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

The developer provides performance data from four periods, covering a total of 1,286,352 admissions and 4,642 hospitals.

- The data show that during the measurement period of July 2016 – June 2019, heart failure readmission rates ranged from a minimum of 16.7% to a maximum of 31.2%, with the 10th percentile at 20.3%, the 50th percentile at 21.9%, and the 90th percentile at 23.7%.

Distribution of Hospital Heart Failure RSRRs over Different Time Periods (All Hospitals)

Results for each data year Characteristic [07/2016-06/2017] | [07/2017-06/2018] | [07/2018-06/2019] | [07/2016-06/2019]

Number of Hospitals: 4537 | 4516 | 4483 | 4642

Number of Admissions: 420222 | 434269 | 431861 | 1286352

Mean (SD): 21.9(1.1) | 22(1.1) | 22(1.1) | 22(1.4)

Range (Min-Max): 17.3-28.8 | 17.1-29.6 | 17.1-28.3 | 16.7-31.2

Minimum//17.3//17.1//17.1//16.7

10th percentile//20.7//20.8//20.8//20.3

20th percentile//21.2//21.3//21.3//21.0

30th percentile//21.4//21.6//21.6//21.4

40th percentile//21.7//21.8//21.7//21.6

50th percentile//21.8//21.9//21.9//21.9

60th percentile//22.0//22.1//22.1//22.1

70th percentile//22.2//22.4//22.3//22.4

80th percentile//22.5//22.7//22.6//22.9

90th percentile//23.1//23.3//23.2//23.7

Maximum//28.8//29.6//28.3//31.2

Disparities

To support the assessment of potential disparities, the developer provides performance scores (using July 2016 - June 2019 data) for hospitals by proportion of dual eligible patients and performance scores for hospitals according to proportion of patients with AHRQ SES Index Score in the lower and upper social risk quartiles.

Distribution of 30-day HF RSRRs by Proportion of Dual Eligible Patients:

- Data Source: Medicare FFS claims, VA claims and Medical Beneficiary Summary File (MBSF) data
- Dates of Data: July 2016 through June 2019
- Variation in RSRRs across hospitals (with at least 25 cases) by proportion of patients with social risk//
- Description of Social Risk Variable//Dual Eligibility
- Quartile//Q1//Q4
- Social Risk Proportion (%)// (0-8.42)//(35.14-100)
    - # of Hospitals//927//928
    - 100%Max//27.0//31.2
    - 90%//23.4//24.3
    - 75%//22.4//23.2
    - 50%//21.6//22.3
    - 25%//20.7//21.5
    - 10%//19.8//20.7
    - 0%Min//16.7//18.4

Distribution of 30-day HF RSRRs by Proportion of Patients with AHRQ SES Index Scores:

- Data Source: Medicare FFS claims, VA claims and The American Community Survey (2013-2017) data
- Dates of Data: July 2016 through June 2019
- Variation in RSRRs across hospitals (with at least 25 cases) by proportion of patients in lower and upper social risk quartiles//
- Description of Social Risk Variable //AHRQ SES Index
- Quartile//Q1//Q4
- Social Risk Proportion (%)// (0-10.22)//(24.46-100)
    - # of Hospitals//928//928
    - 100%Max//31.2//28.8
    - 90%//23.8//24.3
    - 75%//22.6//23.1
    - 50%//21.7//22.3
    - 25%//20.7//21.5
    - 10%//19.7//20.9
    - 0%Min//16.7//17.9

Questions for the Committee:

- *Does the SC agree that 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: For all measures (structure, process, outcome, patient-reported structure/process), empirical data are required. How does the evidence relate to the specific structure, process, or outcome being measured? Does it apply directly or is it tangential? How does the structure, process, or outcome relate to desired outcomes? For maintenance measures –are you aware of any new studies/information that changes the evidence base for this measure that has not been cited in the submission? For measures derived from a patient report: Measures derived from a patient report must demonstrate that the target population values the measured outcome, process, or structure.

- Used for payment
- I am not aware of studies not cited for evidence base, but I am concerned with one study cited regarding transitional care models and impact on readmissions - it resulted in a 7% readmission rate, which while lower than the national average, was based on a review of the model in place with a single system and a limited number of patients. It would be useful to see more expansive studies demonstrating ability to impact CHF readmissions.
- No concerns
- Maintenance measure with additional evidence to support this review--pass for evidence
- I am not aware of any new studies/information that changes the evidence base for this measure.
- I am not aware of any new information related to this measure.
- Evidence may not be all inclusive of patients with CHF. Evidence is restricted to patients over age 65 and this probably skews readmissions for heart failure to a sicker group.

1b. Performance Gap: Was current performance data on the measure provided? How does it demonstrate a gap in care (variability or overall less than optimal performance) to warrant a national performance measure? Disparities: Was data on the measure by population subgroups provided? How does it demonstrate disparities in the care?

- 10th-90th percentile is 20.3-23.7, 3.4 percentage point diff, 15% diff
- Variability demonstrated, though study cited as evidence for measurement suggests some level of readmissions for CHF may be reasonable.
- No concerns
- Yes, current performance data provided and there is little to no improvement over the course of the last 3 years and there is continued need for this measure. The disparities are captured by health plan; however it may be helpful to capture at a more detailed level so you can see the impact on various ages, sex, race, etc.
- Yes. Variability presents an opportunity for improvement. Disparities noted by Proportion of Dual Eligible Patients, across hospitals (with at least 25 cases) by proportion of patients with social risk// Description of Social Risk Variable//Dual Eligibility, by Proportion of Patients with AHRQ SES Index Scores
- Performance data demonstrated a gap in care and need for a national performance measure. Dual eligible status and AHRQ SES index was used to measure social risk factors (SFRs). No statistical significance found at the patient level.

- Measure aims to identify high- (or low-) performing institutions while adjusting for case-mix and patient risks. There are clearly some outliers in the patient sample, possibly amenable to interventions to improve readmission rates.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

2b. Validity: Testing; Exclusions; Risk-Adjustment; Meaningful Differences; Comparability; Missing Data

Reliability

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. For maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures.

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 – less emphasis if no new testing data provided.

Validity

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 – less emphasis if no new testing data provided.

2b2-2b6. Potential threats to validity should be assessed/addressed.

Complex measure evaluated by Scientific Methods Panel? ☒ **Yes** ☐ **No**

Evaluators: NQF Scientific Methods Panel
SMP Rating:
R: H-0; M-7; L-1; I-0 (Pass)
V: H-2; M-5; L-1; I-0 (Pass)
Methods Panel Review (Combined)
Methods Panel Evaluation Summary:
This measure was reviewed by the Scientific Methods Panel and discussed on the call. A summary of the measure and the Panel discussion is provided below.

Reliability

- 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.
- Developer conducted reliability testing at the measure score level by calculating the intra-class correlation coefficient (ICC) using a split sample (i.e. test-retest) method 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 ICC was calculated for hospitals with 25 or more admission.
    - Total admissions: 1,286,352 admissions were included in the analysis of a 3-year period
    - 642,047 admissions from 4,593 hospitals in one half and 644,305 admissions from 4,642 hospitals in the other half.
    - ICC: 0.587 agreement between the two independent assessments of the RSRR for each hospital

- Developer also assessed facility-level reliability levels for hospitals with at least 25 admissions using the signal-to-noise testing method. Demonstrating moderate agreement, the median reliability score was 0.57, ranging from 0.14 to 0.96. The 25th and 75th percentiles were 0.31 and 0.75, respectively.

Validity

- The developer conducted validity testing at the performance measure score level by assessing HF readmission scores correlation with other measures that target the same domain of quality for the same or similar populations.
    - After substantial literature review and consultations with experts in the field, three external hospital quality measures were identified and used in this examination of performance in the HF readmission measure scores (RSRRs):
        - Hospital Star Rating readmission group score
        - Overall Hospital Star Rating
        - HF Excess Days in Acute Care (EDAC)

- The comparison assessment results aligned with the developer's predictions. Demonstrating the strongest association was the correlation between HF RSRRs and the Star-Rating readmissions scores. The correlations were as follows:
        - Correlation between HF RSRRs and Star-Rating readmissions score: -0.585
    - The data supports the suggestion that hospitals with lower HF RSRRs are more likely to have higher Star-Rating readmission scores.
        - The correlation between HF RSRRs and Star-Rating summary score: -0.378
    - The data supports the suggestion that hospitals with lower HF RSRRs are more likely to have higher Star-Rating summary scores.
        - The correlation between HF RSRRs and HF EDAC scores: 0.574
    - The data suggests that hospitals with lower HF RSRRs are more likely to have lower HF EDAC scores.
- The developer also measured validity via medical record validation in the original measure development stage by comparing the HF readmission measure against a medical record measure that they developed. Please refer to the developer submission data on the **Validity criterion in the Scientific Acceptability section** for full information.

Risk Adjustment

- The developer risk-adjusted for 37 risk factors; social risk factors (SRF; dual eligibility and ASPE SES index) were tested but not included in the final specification.
- The developer conducted a decomposition analysis to assess the independent effects of the SRF variables at the patient level and the hospital level.
- For both factors, there was a considerably greater hospital-level effect, compared with the patient-level effect, leading the developer to suggest that any patient-level adjustment alone would also adjust for quality differences between hospitals.
- Considering these findings, as well as the lack of an impact on measure scores, the conceptual model, and the fact that CMS adjusts this measure at the program level, the developer did not to include these SRFs in the risk adjustment model.

Questions for the Committee regarding reliability:

- *Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?*
- *Do you have any concerns regarding the distribution of reliability scores?*
- *The Scientific Methods Panel is satisfied with the reliability testing for the measure. Does the Committee think there is a need to discuss and/or vote on reliability?*

Questions for the Committee regarding validity:

- *Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?*
- *Do you agree with the developer's approach to social risk factor adjustment?*
- *The Scientific Methods Panel is satisfied with the validity analyses for the measure. Does the Committee think there is a need to discuss and/or vote on validity?*

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

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

Committee Pre-evaluation Comments:
Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

2a1. Reliability-Specifications: Which data elements, if any, are not clearly defined? Which codes with descriptors, if any, are not provided? Which steps, if any, in the logic or calculation algorithm or other specifications (e.g., risk/case-mix adjustment, survey/sampling instructions) are not clear? What concerns do you have about the likelihood that this measure can be consistently implemented?

- No issues
- none
- No concerns
- Would rate low-moderate based on the panelist assessment review
- No concerns. The measure is well defined and precisely specified.
- I do not have any concerns about the consistent implementation of this measure.
- Specifications for the causes of heart failure are vague and/or absent. This adds to population heterogeneity and may skew differences in CHF readmission to favor certain subgroups that may not represent all CHF patients.

2a2. Reliability - Testing: Do you have any concerns about the reliability of the measure?

- Yes. Split sample 0.587 median s/n 0.57. Low in terms of differentiating.
- none
- No concerns
- No concerns
- No
- No
- Yes. See Section 5.2a1 above. Specifications for the causes of heart failure are vague and/or absent. This adds to population heterogeneity and may skew differences in CHF readmission to favor certain subgroups that may not represent all CHF patients.

2b1. Validity -Testing: Do you have any concerns with the testing results?

- No

- It says they used a comparison assessment with the Star Ratings Readmission measure group score and with overall Star Ratings – but those ratings use this measure, so it is not surprising that they saw correlations between scores on the measure those scores that use the measure. Would like more details on that approach.
- No concerns
- No concerns (moderate)
- No
- No
- Testing results seem valid for potentially very mixed, possibly skewed, population

2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment) 2b2. Exclusions: Are the exclusions consistent with the evidence? Are any patients or patient groups inappropriately excluded from the measure? 2b3. Risk Adjustment: If outcome (intermediate, health, or PRO-based) or resource use performance measure: Is there a conceptual relationship between potential social risk factor variables and the measure focus? How well do social risk factor variables that were available and analyzed align with the conceptual description provided? Are all of the risk-adjustment variables present at the start of care (if not, do you agree with the rationale provided)? Was the risk adjustment (case-mix adjustment) appropriately developed and tested? Do analyses indicate acceptable results? Is an appropriate risk-adjustment strategy included in the measure?

- SRF do not change ranking or scores in any substantial way when added to RA model.
- Says dual eligibility and ASPE SES index were tested but not included in specs because there was a suggestion that patient-level adjustment alone would adjust for quality differences between hospital. It says that because CMS adjusts at the program level that is sufficient, but CMS does not adjust at the program level – it adjusts for payment purposes and not for measurement and public reporting purposes through the hospital star ratings.
- No concerns
- The AHRQ SES Index Score was used capture risk.
- Yes
- Risk adjustment strategy was included in the measure. SRFs were analyzed at the patient and hospital level by reviewing for dual eligible status and AHRQ SES index
- See the above comments. I am not sure that limiting readmissions to patients older than 65 provides the best metric for heart failure readmissions. This should be tested by comparing a younger population who had readmissions for CHF.

2b4-6. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data) 2b4. Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality? 2b5. Comparability of performance scores: If multiple sets of specifications: Do analyses indicate they produce comparable results? 2b6. Missing data/no response: Does missing data constitute a threat to the validity of this measure?

- No
- no
- No concerns
- N/A
- No
- Hospital Star Rating readmissions and HF RSRRs correlation as indication of quality. No missing data
- Missing data does not seem to be a problem, rather the patients not included in the sample may skew results in the population.

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.

- Measure is supported by administrative claims and enrollment data which the developers note are routinely generated and collected as part of hospitals' billing processes, adding data collection burden to hospitals or providers.
- The original information is coded by a different person other than the person who obtains the original information. The developer indicates that all data elements are in defined fields in electronic claims.

Questions for the Committee:

- *Does the SC have any concerns related to measure feasibility?*

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

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

- No concerns. Administrative data measure
- No feasibility concerns
- No concerns
- Claims data, no concerns
- None
- No concerns. The measure uses administrative claims and enrollment data already collected by hospitals.
- Data accrual is relatively easy since it is based on Medicare data and is encompassed in electronic format.

Criterion 4: Usability and Use

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

4a. Use (4a1. Accountability and Transparency; 4a2. Feedback on measure)

4a. ***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.

4a.1. 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.

Current uses of the measure

Publicly reported? ☒ Yes ☐ No

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

Accountability program details

Public Reporting
Hospital Compare
Payment Program
Hospital Readmission Reduction (HRRP) Program

4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured, and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure.

Feedback on the measure by those being measured or others

The developer maintains an active response and feedback loop for inquiries and comments raised about the implementation, specifications, calculation, etc. of the measure. Submitted questions or comments have included the following:

- Requests for detailed measure specifications including CC-to-ICD-9 code crosswalks, and ICD-10 codes used to define the measure cohort or in the risk-adjustment model;
- Requests for the SAS code and SAS packs used to calculate measure results;
- Request for the national 30-day heart failure readmission rate;
- Queries about how cohorts and outcomes are defined including whether there is overlap in the cohorts and outcomes assessed in the AMI, coronary artery bypass grafting, and stroke mortality measures;
- Questions about how transfers are handled in the measure calculation;
- Queries about how to calculate the measure and to interpret the statistical model including the interpretation of coefficients for risk variables;
- Request for the Measure Calculation Package included in the Hospital Value-Based Program; and;
- Requests for hospital-specific measure information, such as data included in the HSRs.

Additional Feedback:

Feedback received from other stakeholders:

- Requests for detailed measure specifications including the narrative specifications for the measure, CC-to-ICD-9 code crosswalks, and ICD-10 codes used to define the measure cohort or in the risk-adjustment model;
- Requests for the SAS code and SAS packs used to calculate measure results;
- Request for the Measure Calculation Package included in the Hospital Value-Based Program;
- Requests for clarification of how inclusion and exclusion criteria are applied;
- Queries about how cohorts and outcomes are defined;
- Questions about how transfers are handled in the measure calculation;
- Queries about the implementation of the measure in CMS initiatives and payment programs; and
- Requests about hospital-specific measure information including mock Hospital-Specific Reports.

- Additional Feedback (Measure Applications Partnership 2016)
- Addressing cost of care for a heart failure episode aligns with the National Quality Strategy, addressing a condition that is significant driver of cost for the Medicare program;
- Reporting this measure with an indicator of quality of care associated with heart failure, such as the risk-standardized mortality and risk-standardized readmission measures that CMS is already collecting through the Hospital Inpatient Quality Reporting (HIQR) and Hospital Readmissions Reduction Penalty Programs (HRRP), will increase understanding of the value of healthcare services provided, while also illuminating drivers of cost;
- This measure is likely to drive care coordination towards the reduction of readmissions and increase of patient utilization of less costly care;
- Concerns that the measure, one of federal spending, is not a real indicator of "value" for beneficiaries as it does not capture the quality of care; it is a cross cutting measure that is not appropriate for evaluating care exclusively in the inpatient setting;
- Concerns with variability of post-acute care quality and use (due to local market availability, resources, patterns of care, etc.) tied to outcomes (readmissions rates) that are not necessarily within the hospital's control;
- Question about the extent to which this measure will produce actionable information that a hospital and physician could use to identify how to make improvements.
- Opposition to potentially layering condition-specific measures on top of the broad total cost measures that CMS is now using.

***Questions for the Committee**:*

- *Does the SC agree that the performance results have been used to further the goal of high-quality, efficient healthcare?*
- *Is there anything that the Committee wishes to discuss related to the current use of the measure?*

Preliminary rating for Use: ☒ Pass ☐ No Pass

4b. Usability (4a1. Improvement; 4a2. Benefits of measure)

4b. Usability 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.

4b.1 Improvement. Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.

Improvement results

- The developer reports a median HF 30-day, all-cause, RSRR for the HF readmission measure for the 3-year period between July 1, 2016 and June 30, 2019 was 21.9%. The median RSRR increased by 0.1 absolute percentage points from July 2016-June 2017 (median RSRR: 21.8%) to July 2018-June 2019 (median: RSRR: 21.9%)."

Year over year data over a three-year spread indicates little to no change in performance rates:

Results for each data year Characteristic [07/2016-06/2017] | [07/2017-06/2018] | [07/2018-06/2019] | [07/2016-06/2019]

Number of Hospitals: 4537 | 4516 | 4483 | 4642

Number of Admissions: 420222 | 434269 | 431861 | 1286352

Mean (SD): 21.9(1.1) | 22(1.1) | 22(1.1) | 22(1.4)

Range (Min-Max): 17.3-28.8 | 17.1-29.6 | 17.1-28.3 | 16.7-31.2

Minimum//17.3//17.1//17.1//16.7

10th percentile//20.7//20.8//20.8//20.3

20th percentile//21.2//21.3//21.3//21.0

30th percentile//21.4//21.6//21.6//21.4

40th percentile//21.7//21.8//21.7//21.6

50th percentile//21.8//21.9//21.9//21.9

60th percentile//22.0//22.1//22.1//22.1

70th percentile//22.2//22.4//22.3//22.4

80th percentile//22.5//22.7//22.6//22.9

90th percentile//23.1//23.3//23.2//23.7

Maximum//28.8//29.6//28.3//31.2

4b2. Benefits vs. harms. 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).

Unexpected findings (positive or negative) during implementation

- Developer reported that health services researchers have explored potential spillover effects of the HF readmission measure's implementation and reductions in readmissions for non-targeted conditions.
- Several studies support positive spillover effects, as there has been systematic improvement in risk-standardized readmission rates for patients not included in HRRP measures (Carey et al., 2015; Angraal et al., 2018; Demiralp et al., 2018; Sukul et al., 2017; Myers et al., 2020).

Potential harms

- Developer does not offer any examples of potential harms.

***Questions for the Committee**:*

- *Give the year-over-year reported median results, does the SC agree that the performance results can be used to further the goal of high-quality, efficient healthcare?*

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

Committee Pre-evaluation Comments:
Criteria 4: Usability and Use

4a1. Use - Accountability and Transparency: How is the measure being publicly reported? Are the performance results disclosed and available outside of the organizations or practices whose performance is measured? For maintenance measures - which accountability applications is the measure being used for? For new measures - if not in use at the time of initial endorsement, is a credible plan for implementation provided? 4a2. Use - Feedback on the measure: Have those being measured been given performance results or data, as well as assistance with interpreting the measure results and data? Have those being measured or other users been given an opportunity to provide feedback on the measure performance or implementation? Has this feedback has been considered when changes are incorporated into the measure?

- Standard feedback: score, rank, detail on cases included in analysis
- none
- No concerns
- Yes, the measure is being publicly being reported Hospital Compare/ Hospital Readmission Reduction Program and allows for feedback via open comment and MAP feedback.

- Yes
- Current use on various publicly available websites including Hospital Compare. Feedback was reviewed. No changes made to measure.
- I cannot see that patients have any feedback related to measure results and interpretation.

4b1. Usability – Improvement: How can the performance results be used to further the goal of high-quality, efficient healthcare? If not in use for performance improvement at the time of initial endorsement, is a credible rationale provided that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations? 4b2. Usability – Benefits vs. harms: Describe any actual unintended consequences and note how you think the benefits of the measure outweigh them.

- Would like some discussion of evidence on association with increased mortality. The conclusion cited on no sufficient evidence suggests inability to reject null leads to conclusion that should treat finding as if null is value. would like more info on actual magnitudes of observed effects.
- Would like to see discussion from developer of studies suggesting impact of measure in Hospital Readmission Reduction Program on mortality rates. (see Wadhera, et al., Association of the Hospital Readmissions Reduction Program with Mortality Among Beneficiaries Hospitalized for Heart Failure, Acute Myocardial Infarction, and Pneumonia, JAMA, 2018)
- No concerns
- No harm noted, however, the benefits could also positively affect other disease states and reduce readmissions rate.
- The benefits outweigh any unintended consequences.
- Studies showed improvement in the readmission rates for patients hospitalized with diagnosis other than heart failure. Systematic implementation of strategies across diagnosis may reduce overall readmission rates.
- There are very few adverse outcomes related to use of this measure, while measure performance results might improve readmission outcomes. One potential adverse economic result of these measurements might be to keep CHF patients in the hospital longer to avoid penalties related to readmissions - something that has been seen in studies that penalize providers for readmissions.

Criterion 5: Related and Competing Measures

Related or competing measures

- 0229 Hospital 30-Day, All-Cause, Risk-Standardized Mortality Rate (RSMR) Following Heart Failure (HF) Hospitalization
- 0505 Hospital 30-day all-cause risk-standardized readmission rate (RSRR) following acute myocardial infarction (AMI) hospitalization.
- 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
- 2879 Hybrid Hospital-Wide Readmission (HWR) Measure with Claims and Electronic Health Record Data
- 2880 Excess days in acute care (EDAC) after hospitalization for heart failure (HF)
- 2886 Risk-Standardized Acute Admission Rates for Patients with Heart Failure
- 2888 Accountable Care Organization Risk-Standardized Acute Hospital Admission Rate for Patients with Multiple Chronic Conditions

Harmonization

Developer indicated that all measure specifications have been harmonized to the furthest extent possible. The developer did not include non-outcome (e.g., process) measures with the same target population as NQF 330 in its list of related measures. The developer noted the patient exclusion limitations of non-outcome

measures and explained that clinical coherence of the cohort takes precedence over alignment with related non-outcome measure.

Committee Pre-evaluation Comments: Criterion 5: Related and Competing Measures

5. Related and Competing: Are there any related and competing measures? If so, are any specifications that are not harmonized? Are there any additional steps needed for the measures to be harmonized?

- No
- interaction with all-cause readmissions measures merits discussion
- No concerns
- The measures appear to be harmonized.
- Yes and they are harmonized to the extent possible
- Yes. Non-outcome or process measures were not included.
- Competing measures involve a more well-defined patient population. For example, admission following cardiac operations in patients with CHF presents a more homogeneous study group. There are likely other subgroups of CHF patients who might provide a homogeneous study group (e.g., drug-related cardiomyopathy, certain infections, alcohol, or drug abuse)

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 01/21/2021

- Comment by: American Medical Association

  The American Medical Association (AMA) appreciates the opportunity to comment on the NQF Quality Positioning System (QPS) Measure #330: Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization. This is an important measure which captures the unplanned readmission for any reason within 30 days of a patient's discharge from the hospital.

  In reviewing the calculation, we are disappointed to see the minimum measure score reliability result calculated at 0.14 and the intraclass correlation coefficient (ICC) calculated at 0.587, both using a minimum case number of just 25 patients. We believe that measures must meet minimum acceptable thresholds of 0.7 for reliability and require higher case minimums to allow for the overwhelming majority of hospitals to achieve an ICC of 0.6 or higher.

  The AMA is also extremely concerned that the measure developer used the recommendation to exclude social risk factors in the risk adjustment models for measures that are publicly reported as outlined in the recent report to Congress by Assistant Secretary for Planning and Evaluation (ASPE) on Social Risk Factors and Performance in Medicare's Value-based Purchasing program (ASPE, 2020). We believe that while the current testing may not have produced results that would indicate incorporation of the two social risk factors included in testing, this measure is currently used both for public reporting and value-based purchasing. A primary limitation of the ASPE report was that none of the recommendations adequately addressed whether it was appropriate to adjust for social risk factors in the same measure used for more than one accountability purpose, which is the case here. This discrepancy, along with the fact that the additional analysis using the American Community Survey is not yet released, must be addressed prior to any reliance on the recommendations within this report. We also note that the measure developer chose to include social risk factors in two measures (#2888 and #3597) under review; we ask that this inconsistency be considered and rectified.

In addition, we question whether the measure continues to be useful to distinguish hospital performance and drive improvements based on the distribution of a hospital's performance scores. We raise this question because only 110 hospitals performed better than the national rate, and 149 hospitals were worse (as noted in section 2b4). The discussion on improvement (as noted in section 4b1 of the measure submission form) found only an increase of 0.1 absolute percentage points between July 2016-June 2017 and July 2018-June 2019 in this measure.

The AMA requests that the Standing Committee evaluate whether the measure continues to meet the measure evaluation criteria required for endorsement.

Reference:

Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health Human Services. Second Report to Congress on Social Risk Factors and Performance in Medicare's Value-Based Purchasing Program. 2020. https://aspe.hhs.gov/social-risk-factors-and-medicares-value-based-purchasing-programs

- Comment by: Federation of American Hospitals

The Federation of American Hospitals (FAH) appreciates the opportunity to comment on Measure #330, Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization. The FAH is concerned that even though the median reliability score was 0.57 for hospitals with at least 25 cases, reliability ranged from 0.14 to 0.96 and that the intraclass correlation coefficients (ICC) was 0.587. The FAH believes that the developer must increase the minimum sample size to a higher number to produce a minimum reliability threshold of sufficient magnitude (e.g., 0.7 or higher) and an ICC of 0.6 or higher.

In addition, the FAH is very concerned to see that the measure developer's rationale to not include social risk factors in the risk adjustment model was in part based on the recommendations from the report to Congress by Assistant Secretary for Planning and Evaluation (ASPE) on Social Risk Factors and Performance in Medicare's Value-based Purchasing program released in March of last year (ASPE, 2020). A fundament flaw within the ASPE report was the lack of any recommendation addressing how a single measure with multiple accountability uses should address inclusion of social risk factors as is the case with this measure, which is both publicly reported and included in the Hospital Value-Based Purchasing program. Regardless of whether the testing of social risk factors produced results that were sufficiently significant, the FAH believes that no developer should rely on the recommendations of this report until the question of how to handle multiple uses is addressed along with the additional analysis using the American Community Survey. We also note that the developer chose to include social risk factors in two measures (#2888 and #3597) under review and we ask that this inconsistency be considered.

Lastly, the FAH is concerned that there is insufficient variation in performance across hospitals and limited opportunities for improvement to support this measure's continued use in accountability programs. Specifically, the performance scores reported in 2b4. Identification of Statistically Significant and Meaningful Difference in Performance are generally low with only 110 hospitals identified as better than the national rate and 149 are worse than the national rate. We base our concerns on these results along with the discussion on improvement in section 4b1 of the measure submission form where only an increase of 0.1 absolute percentage points between July 2016-June 2017 and July 2018-June 2019 was found.

As a result, the FAH requests that the Standing Committee carefully consider whether the measure as specified should continue to be endorsed.

Reference:

Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health & Human Services. Second Report to Congress on Social Risk Factors and Performance in Medicare's Value-

Based Purchasing Program. 2020. https://aspe.hhs.gov/social-risk-factors-and-medicares-value-based-purchasing-programs

- Comment by: Anonymous
  - I support this measure.
- Of the 1 NQF member who have submitted a support/non-support choice:
  - 0 support the measure
  - 1 does not support the measure

Combined Methods Panel Scientific Acceptability Evaluation

Scientific Acceptability: Preliminary Analysis Form
Measure Number: 0330
Measure Title: Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization Type of measure:
☐ **Process** ☐ **Process: Appropriate Use** ☐ **Structure** ☐ **Efficiency** ☐ **Cost/Resource Use**
☒ **Outcome** ☐ **Outcome: PRO-PM** ☐ **Outcome: Intermediate Clinical Outcome** ☐ **Composite**
Data Source:
☒ **Claims** ☐ **Electronic Health Data** ☐ **Electronic Health Records** ☐ **Management Data**
☐ **Assessment Data** ☐ **Paper Medical Records** ☐ **Instrument-Based Data** ☒ **Registry Data**
☒ **Enrollment Data** ☒ **Other**
Level of Analysis:
☐ **Clinician: Group/Practice** ☐ **Clinician: Individual** ☒ **Facility** ☐ **Health Plan**
☐ **Population: Community, County or City** ☐ **Population: Regional and State**
☐ **Integrated Delivery System** ☐ **Other**
Measure is:
☐ **New** ☒ **Previously endorsed (**NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.)

RELIABILITY: SPECIFICATIONS

1. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?** ☒ **Yes** ☐ **No**
   Submission document: "MIF_xxxx" document, items S.1-S.22
   NOTE*: NQF staff will conduct a separate, more technical, check of eCQM specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.*
2. **Briefly summarize any concerns about the measure specifications.**
   Panel Member #1 NONE
   Panel Member #8 No concerns with specs. Similar to other readmission measures submitted with appropriate changes for the specific disease.
   Panel Member #9 No concerns.

RELIABILITY: TESTING

Submission document: "MIF_xxxx" document for specifications, testing attachment questions 1.1-1.4 and section 2a2

3. **Reliability testing level** ☒ **Measure score** ☒ **Data element** ☐ **Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure** ☒ **Yes** ☐ **No**
5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical *VALIDITY* testing** of **patient-level data** conducted?
   ☐ **Yes** ☐ **No NA (Panel Member #1) X – Not Appliable (Panel Member #5)**
6. **Assess the method(s) used for reliability testing**
   Submission document: Testing attachment, section 2a2.2
7. **Assess the results of reliability testing**

Submission document: Testing attachment, section 2a2.3
Panel Member #1 Split sample test-retest statistic 0.587
Missing from the split sample analysis was an analysis that has proven useful in the past to assessing how reliably the measure identifies relative performance, specifically, the quintile-to-quintile cross tab of hospitals comparing the first sample to the second.
S/N: Median 0.57
S/N: 10th percentile 0.20
S/N: 90th percentile 0.84
Out of 4,642 hospitals in measure cohort 110 performed "better than the U.S. national rate," 3,454 performed "no different from the U.S. national rate," and 149 performed "worse than the U.S. national rate." 929 were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing.
The 10th-90th percentile RSRRs are 20.3-23.7, a 3.4 percentage point difference. Given the number of cases, this is clinically meaningful, but the limited range imposes a substantial burden to reliably measure relative performance.
The results of the testing suggest the measure cannot reliably differentiate performance. The split sample test statistic is low against the standard that would differentiate hospitals, the S/N median is low against a standard that would differentiate hospitals. (The Adams S/N analysis made a S/N of 0.7 its de facto standard.) This is also reflected in the small number of hospitals identified as performing better than or worse than the national rate, 259 out of 3,713 with 25 or more cases (less than 10%).
Panel Member #2 Using the Spearman-Brown correlation was 0.587, which indicate the extent of agreement between the two independent assessments of the RSRR for each hospital was 0.587.
The median reliability score from signal-to-noise reliability analysis was 0.57 implying moderate agreement.
The measure is based on hospitals with HF admissions 25 or more during the measurement period; however, I do not see the exact number of those hospitals.
Panel Member #3 The Split-sample analysis finds a Spearman-Brown correlation statistic of 0.59 and facility level reliability statistic has a median of 0.57 – seems modest at best.
Panel Member #4 Using a minimum case volume of 25 had moderate reliability
Panel Member #6 Split-Sample Reliability: A total of 1,286,352 admissions were included in the analysis based on 3 years of data. After randomly splitting the sample into two halves, there were 642,047 admissions from 4,593 hospitals in one half and 644,305 admissions from 4,642 hospitals in the other half. The ICC was calculated for hospitals with 25 admissions or more. Using the Spearman-Brown prediction formula, the agreement between the two independent assessments of the RSRR for each hospital was 0.587.
Signal-to-Noise: The signal-to-noise reliability score was calculated for hospitals with at least 25 admissions. The median reliability score was 0.57, ranging from 0.14 to 0.96. The 25th and 75th percentiles were 0.31 and 0.75, respectively. The median reliability score demonstrates moderate agreement.
Panel Member #8 The results with split-sample testing on 1,286,352 admissions over 3 years were an intra-class coefficient of 0.587, which is moderate reliability.
The results of the signal-to-noise, calculated on all admissions were 0.57, which is moderate reliability.
Panel Member #9 Measure score reliability demonstrated moderate agreement based on the Landis Koch standards. Signal to noise reliability also demonstrated moderate reliability.

8. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? NOTE: If multiple methods used, at least one must be appropriate.
   Submission document: Testing attachment, section 2a2.2
   ☒ **Yes**
   ☐ **No**
   ☐ **Not applicable** (score-level testing was not performed)
9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?
   Submission document: Testing attachment, section 2a2.2

☒ **Yes**

☐ **No**

☒ **Not applicable** (data element testing was not performed)

10. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and **all** testing results):

☐ **High** (NOTE: Can be HIGH **only if** score-level testing has been conducted)

☒ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has **not** been conducted)

☒ **Low** (NOTE: Should rate **LOW** if you believe specifications are NOT precise, unambiguous, and complete or if testing methods/results are not adequate)

☐ **Insufficient** (NOTE: Should rate **INSUFFICIENT** if you believe you do not have the information you need to make a rating decision)

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**

Panel Member #1 The results of the testing suggest the measure cannot reliably differentiate performance. The split sample test statistic is low against the standard that would differentiate hospitals, the S/N median is low against a standard that would differentiate hospitals. (The Adams S/N analysis made a S/N of 0.7 its de facto standard.) This is also reflected in the small number of hospitals identified as performing better than or worse than the national rate, less than 10% of the sample, basically the tails of the distribution, with no evidence that these hospitals would be consistently in the tails across years or split sample.

Panel Member #2 See 7 above for my justification.

Panel Member #3 I would rate this measure between moderate and low - users of the tool need to be aware that reliability may be an issue, particularly for lower volume facilities.

Panel Member #4 This submission demonstrates integrity in the determination of case volumes for moderate reliability.

Panel Member #5 Reliability testing was adequate. Very low inter-decile distribution of Provider performance may be a problem.

Panel Member #6 The scores of both the split sample and signal-to-noise reliability were both well below .70 (mid-point of substantial agreement) at 0.59 and 0.57, indicating reliable results only slightly more than half the time. The developers indicate this represents "moderate agreement" at best. While I think these scores are too low for measures used in public reporting and value-based payment, there is not yet a threshold cut-off set by the SMP or NQF guidance to allow us to reject a measure with scores below some more generally acceptable threshold such as .7 or .8. I believe we should be setting higher standards for these measures given their importance in determining which hospitals receive penalties or reduced payments based on these measure scores. The developers note that reliability of measures used to define complex constructs such as clinical severity or patient comorbidities is significantly lower than for simpler constructs such as patient weight.

Panel Member #8 By two standardized testing methods applied to the entire population of admissions, the result meet the criteria of moderate reliability.

Panel Member #9 Would rate high if reliability results were higher

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

12. **Please describe any concerns you have with measure exclusions.**

Submission document: Testing attachment, section 2b2.

Panel Member #1 NONE

Panel Member #2 Table 3 documenting the numbers (%) of patients dropped due to each of the exclusion criteria is for hospitals with 25 or more qualifying admissions. Why is the latter (hospitals with 25 or more qualifying admissions) not an exclusion criterion?

Panel Member #4 None

Panel Member #6 None.

Panel Member #9 No concerns

Panel Member #8 Exclusions were those common to Medicare claims data measures with the largest contributor being a HF admission within 30 days of a prior HF index admission, about 8% of the data sample.

13. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.**

Submission document: Testing attachment, section 2b4.

Panel Member #1 The 10th-90th percentile RSRRs are 20.3-23.7, a 3.4 percentage point difference. Given the number of cases, this is clinically meaningful, but the narrow range imposes a burden to reliably measure relative performance.

Panel Member #2 None

Panel Member #3 None

Panel Member #4 None

Panel Member #6 The developers report a median odds ratio of 1.15 which they say suggests a meaningful increase in the risk of readmission if a patient is admitted with HF at a higher risk hospital compared to a lower risk hospital. A ratio indicates that a patient has a 15% increase in the odds of a readmission at higher risk performance hospital compared to a lower risk hospital, indicating the impact of quality on the outcome rate.

They add that the variation in rates and number of performance outliers suggests there remain differences in the quality of care received across hospitals for HF. However, the distribution of rates is not shown, such as differences in rates across deciles or quartiles. I would like to see these data to better understand whether the measure can truly identify "meaningful differences" in performance between hospitals.

Panel Member #8 The methodology for RSRR's involves calculating an observed/expected ratio multiplied by the national unadjusted rate. Predicted is determined using a hospital-specific intercept which is added to the sum of the estimated regression coefficients multiplied by the patient characteristics. Results are then summer over all patients attributed to a hospital to get a predicted value. Expected is derived the same way, using a common intercept for all hospitals.

It is concluded that if the RSRR's interval estimate contains the national rate, then the hospital is "no different" than the national rate. If the RSRR's interval does not contain the national rate, then the hospital is either "better than" or "worse than" the national rate.

Also utilized is the median odds ratio which represents the increase in the odds of a readmission at a given high-risk hospital compared to a lower risk hospital.

This resulted in an median odds ratio of 1.15, which means that the median increase was 15% from the model.

Of 4,642 hospitals, 110 performed "better than" the national rate, 3,454 performed "no different from" the national rate, 149 were "worse than the national rate, and 929 were classified as having too small of numbers to tell reliably.

Thus, the data was able to discriminate differences in performance.

Panel Member #9 No concerns

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**

Submission document: Testing attachment, section 2b5.

Panel Member #2 N/A

Panel Member #3 NA

Panel Member #4 None

Panel Member #6 They are using both Medicare and VA data. I suspect VA data is quite different, but I did not see any comparisons or testing to see if the coefficients or model might be different in that population. The original model in fact was based on CA all payer data, which I am also not convinced is fully representative of the entire nation, CA is quite different. So, I still have concerns about how the model was developed originally and what further testing has been done as the model is recalibrated on annual basis.

Panel Member #8 Not applicable.

15. **Please describe any concerns you have regarding missing data.**
Submission document: Testing attachment, section 2b6.
Panel Member #1 N/A
Panel Member #3 No information presented – authors said it was not applicable b/c the measure is claims based.
Panel Member #4 None
Panel Member #6 None.
Panel Member #8 There was no missing data.
16. **Risk Adjustment**
16a. **Risk-adjustment method** ☐ **None** ☒ **Statistical model** (**Panel Member #6** 37 risk factors) ☐ **Stratification**
16b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**
☒ Yes ☐ No ☒ Not applicable
16c. **Social risk adjustment:**
16c.1 Are social risk factors included in risk model? ☒ Yes ☒ No ☐ Not applicable
Panel Member #5 ZIP code level—Area Deprivation Index (ADI) from Census data (2009-2013)
16c.2 Conceptual rationale for social risk factors included? ☒ Yes ☐ No
16c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☒ Yes ☐ No
16d. **Risk adjustment summary:**
16d.1 All of the risk-adjustment variables present at the start of care? ☒ Yes ☐ No
16d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐ Yes ☐ No - **Panel Member #1** N/A
16d.3 Is the risk adjustment approach appropriately developed and assessed? ☒ Yes ☐ No
16d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☒ Yes ☐ No
16d.5. Appropriate risk-adjustment strategy included in the measure? ☒ Yes ☐ No
Panel Member #5 See previous comments
16e**. Assess the risk-adjustment approach**
Panel Member #1 Risk adjustment approach is standard CMS HCC model. SRF's while differentiated across patients and to some extent across hospitals, only explain a small portion of variance and correlation of scores based on risk adjustment model with and without SRFs is approximately 0.985.
Panel Member #2 The overall C-statistic of the model which is in the lower sixties suggests that the model does fit the data well.
Panel Member #3 The logistic regression model fit is modest (0.60). The addition of the risk deciles and the test for overfitting help build a case for modeling strategy.
Panel Member #4 Social risk factors are well conceptualized.
Panel Member #5 Risk adjustment was generally adequate, though not exceptional. Presentation of "Social Risk Proportion" (section 1b.4) was confusing.
Panel Member #6 Developers use a two-stage approach, first identifying the comorbidity or clinical status risk factors that were most important in predicting the outcome, then considering the potential addition of social risk factors. It is not clear why social risk factors were included only in 2nd phase indicating lesser importance to predicting readmissions. They first selected all chronic conditions (CCs) deemed relevant to the Medicare population and to the readmission outcome. Final variable selection was accomplished using a modified stepwise logistic regression based on 1,000 bootstrap samples. A logistic stepwise regression including all candidate variables was run on each sample, and they evaluated the percentage of times a candidate variable was significant at $p<0.01$ level in the models. They included not only variables that exceeded a "predetermined cutoff" (not stated?), but specific variables with particular clinical relevance were "forced" into the model regardless of percent of times significant in the models to ensure appropriate

risk adjustment for HF (e.g., end of life, frailty variables such as pressure ulcers, cancers, stroke, CKD). This resulted in a final risk adjustment model with 37 variables.
As a second stage, the developers assessed the relationship between two social risk factors (SFRs) and the outcome and examined the incremental effect in the multivariable model. They used dual-eligible status and the AHRQ SES index as the two SFRs. They assessed the relationship between the SRF variables with the outcome and the incremental effect in a multivariable model (i.e., the extent to which the addition of any one of these variables improved model performance or changed hospital results). As an additional step, they assessed whether there was a "contextual effect" at the hospital level to assure the impact of the SFR on the outcome was not primarily due to differences in hospitals. They used decomposition analysis to assess the independent effects of the SRF variables at the patient level and the hospital level.
The clinical variables as noted were not all statistically significant.
The SFR variables however showed disparities in readmission rates; 2020 observed rate for dual eligible patients was 24.8% (compared to 21.3% for non-duals), and for patients with low AHRQ SES scores 23.7% (compared to 21.5% for high SES patients). They also evaluated the incremental effect of SRF variables on the risk adjustment model and found effect size (Odds Ratios) of 1.07 and 1.07 when added independently into the model. NOTE that these ORs are similar to effect of MANY of the clinical factors included in the model. They found the C-statistic was relatively unchanged with addition of any of the SRF variables (constant at 0.61, which is not necessarily good fit to start with). I would argue that the independent addition of many of the clinical variables included in the model that had ORs closer to 1.00 would also not change the C-statistic.
Finally, they found the addition of SRF variables had little effect on hospital rates. The median absolute change in hospitals' RSRRs when adding a dual eligibility indicator was 0.023% (interquartile range [IQR] -0.017% – 0.028%) with a correlation coefficient between RSRRs for each hospital with and without dual eligibility added of 0.999. The median absolute change in hospitals' RSRRs when adding a low AHRQ SES Index score indicator to the model was 0.141% (IQR -0.132% – 0.147%) 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 0.980.
The contextual effect analysis of patient level vs. hospital level effects of the SRFs showed both the patient-level and hospital-level dual eligibility, as well as low AHRQ SES Index effects, were significantly associated with HF readmission in the decomposition analysis. They claim that "*the significance of the hospital-level effects indicates that if dual eligibility or low AHRQ SES Index variables were 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."* I would argue that this is not necessarily true; if appropriate risk adjustment methods are used, the adjustment can be made without losing the hospital effect which differentiates high performing vs. low performing hospitals. In fact, while the SES effect was greater on hospital level (not surprising as AHRQ index is community-based measure), the dual effect was greater at the patient level.
Based on results and recent recommendation in ASPE 2020 report to Congress recommending quality measures are NOT adjusted for SFRs, CMS chose not to include the 2 SRFs in the final model. Based on evidence presented, I am not convinced this decision was correct or consistent with their logic for inclusion of clinical risk factors.
Panel Member #8 An expert panel examined the Condition Categories dataset and eliminated those variables that did not make clinical sense to include in this measure. The remaining variables were then subject to a stepwise logistic regression analysis to determine those that were most significantly associated with the outcome. Finally, the clinical team reviewed those and provided a cut-off limit to acceptance. This result in 37 variables included in the model.
Social risk variables were included in the analysis but did not contribute significantly to the model's effectiveness in determining the RSRR.
A c-statistic of 0.61 was obtained, indicating fair model discrimination. Risk decile plots (of predicted risk) were also utilized and showed that the higher deciles had higher observed outcomes (higher readmission rates).

Panel Member #9 Very thorough analysis using Dual-Eligibility as a proxy measure for SES and AHRQ's SES Indicators

***For cost/resource use measures ONLY*:**

17. **Are the specifications in alignment with the stated measure intent?**

☐ **Yes** ☐ **Somewhat** ☐ **No (If "Somewhat" or "No", please explain)**

18. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**

VALIDITY: TESTING

19. **Validity testing level:** ☒ **Measure score** ☒ **Data element** ☐ **Both**

20. **Method of establishing validity of the measure score:**

☒ **Face validity**

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

☐ **N/A (score-level testing not conducted)**

21. **Assess the method(s) for establishing validity**

Submission document: Testing attachment, section 2b2.2

Panel Member #1 Correlation with STAR ratings readmission score
Correlation with STAR ratings overall score
Correlation with AMI EDAC scores

Panel Member #2 Empirical validity testing was conducted in this submission. Empirical validity was conducted by correlating the HF RSRR with (i) Hospital Star Rating readmission group score (hypothesized correlation: negative), (ii) Overall Hospital Star Rating (hypothesized correlation: negative) and (iii) HF Excess Days in Acute Care (EDAC) (hypothesized correlation: positive).

Panel Member #3 The measure developers focus on validity using three measures from hospitals compare. Seems appropriate but limited.

Panel Member #4 Developer examined correlations with the Start rating readmission scores and summary score.

Panel Member #6 Developers identified the measure's correlation with other measures that target the same domain of quality, including the Hospital Star Rating readmission group score measure, the Overall Hospital Star Rating, and the HF Excess Days in Acute Care (EDAC) measure which is a broader measure including readmission, emergency room visits and observation room stays within 30 days of index HF admission.

They also conducted a test of validity using a similar model based on medical records. They developed a measure cohort with the medical record data using the inclusion/exclusion criteria and risk-adjustment strategy consistent with the claims-based administrative measure but using chart-based risk adjusters, such as blood pressure, not available in the claims data. They matched a sample of 64,329 of the same patients in the administrative data and compared the output of the two measures. They used the area under the receiver operating characteristic (ROC) curve for the two models, comparing the predictive ability in readmission rates in the lowest predicted decile and the highest predicted decile. They then estimated hospital-level RSRRs using the corresponding hierarchical logistic regression administrative and medical record models for the linked patient sample and examined the linear relationship between the two sets of estimates using regression techniques and weighting by the total number of cases in each hospital.

Panel Member #8 There were three bases utilized for validity testing, all correlation with other measures – the Hospital Star Rating Readmission Group, the Overall Hospital Star Rating, and HF Excess Days in Acute Care (EDAC)

Panel Member #9 Interesting method of using the Star Ratings given the controversial methodology of calculating the overall Star Ratings.

22. **Assess the results(s) for establishing validity**

Submission document: Testing attachment, section 2b2.3

Panel Member #1 Correlation with STAR ratings readmission score -.585
Correlation with STAR ratings overall score: -.378
Correlation with AMI EDAC scores: 0.574

Levels of correlation are sufficient, although the confidence intervals for RSRR across quartiles of STAR ratings substantially overlap.
EDAC measure includes costs associated with readmission, which is directly measured by measure under consideration. So high correlation is to be expected. Would like to see correlation with EDAC measure excluding readmission.
Panel Member #2 The correlation between HF RSRRs and Star-Rating readmissions score is -0.585, which suggests that hospitals with lower HF RSRRs are more likely to have higher Star-Rating readmission scores. The correlation between HF RSRRs and Star-Rating summary score is -0.378, which suggests that hospitals with lower HF RSRRs are more likely to have higher Star-Rating summary scores.
The correlation between HF RSRRs and HF EDAC scores is 0.574, which suggests that hospitals with lower HF RSRRs are more likely to have lower HF EDAC scores.
Panel Member #3 The strongest correlation was with the overall start rating readmissions domain (-0.58), which I guess is to be expected. This whole section is making me think about readmissions as an outcome because it is a mix of expected and unexpected or negative events. The validation testing does not really help us tease out or further understand the portion of this measure that is true outcome.
Panel Member #4 The degree of consensus was moderate to low.
Panel Member #5 The measure has been widely used by many researchers in published articles.
Panel Member #6 Correlation between HF RSRRs and Star-Rating readmissions scores was -0.585, indicating hospitals with lower readmission rates were more likely to have higher Star-Rating readmission scores as expected. The correlation between HF RSRRs and the Overall Star-Rating summary score was -0.378, which suggests that hospitals with lower HF RSRRs are more likely to have higher Star-Rating summary scores as expected. The correlation of the latter was hypothesized to be lower at the Overall Star Rating is influenced by many other measures. The correlation between HF RSRRs and HF EDAC scores was 0.574, which suggests that hospitals with lower HF RSRRs are more likely to have lower HF EDAC scores as expected.
The performance of the administrative and medical record models was similar. The ROC curve statistics 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% to 38%, a range of 23%. For the medical record model, the corresponding range was 16% to 34%, a range of 18%. The correlation coefficient of the standardized rates from the administrative and medical record models was 0.97.
These results support measure score validity, and the ROC results comparing the 2 models based on 2 different data sources show high validity based on similar fit.
Panel Member #8 Correlations with the above were box plots showing -0.585 for the Star-Rating Group Score, -0.378 for the Overall Star Ratings, and +0.574 for the HF EDAC Scores.
In addition, ROC curves were generated for medical record validation of the administrative claims data. AUC's were 0.61 and 0.58.
Panel Member #9 Moderate correlation of HF Readmissions with Star Rating scores was demonstrated. Stronger correlation for HF Excess Days in Acute Care (EDAC)

23. **Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?**

Submission document: Testing attachment, section 2b1.

☒ **Yes - Panel Member #1** Yes but see note in 22.

☐ **No**

☐ **Not applicable** (score-level testing was not performed)

24. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?** *NOTE that data element validation from the literature is acceptable.*

Submission document: *Testing attachment, section 2b1.*

☒ **Yes - Panel Member #1** We have previously accepted the argument that CMS auditing of data for payment was an acceptable measure of data element accuracy.

☐ **No**

☒ **Not applicable** (data element testing was not performed)

25. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.**

☒ **High** (NOTE: Can be HIGH only if score-level testing has been conducted)

☒ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has NOT been conducted)

☒ **Low** (NOTE: Should rate LOW if you believe that there **are** threats to validity and/or relevant threats to validity were **not assessed OR** if testing methods/results are not adequate)

☐ **Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the score level and the data element level **is required**; if not conducted, should rate as INSUFFICIENT.)

26. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**

Panel Member #1 Measure has face validity and reasonable correlation with other similar measures.

Panel Member #2 My score is based on 16e, 21 and 22 above.

Panel Member #3 Modest to low criterion related validity and the authors only test on aspect of validity.

Panel Member #4 A demonstration of an implicit quality construct is the lowest level of empirical validity testing or validity maturity level 0. To demonstrate a moderate level, the developer must show an empirical association between the explicit quality construct and the material outcome. The EDAC correlation might be informative if ED visits and observation bed stays were assessed separately from readmissions with an associated behavioral hypothesis.

Panel Member #5 Developer demonstrated an effort to risk adjust measure to create valid measure score. Very low inter-decile distribution across Providers may be a problem in discriminating among Providers. I have a personal bias against using the RSRR approach (described in S.14) comparing the "predicted" to the "expected" Provider rates because both values are dependent upon the quality (power and specificity) of the regression models. However, the RSRR methodology has been deemed acceptable by SMP by consensus and I will abide by that decision.

Panel Member #6 The validity test results show good validity of the model. The ability to develop a similar model using medical record data vs. claims data, and one that includes data elements not available in claims (such as lab values) indicate strong validity of the model

Panel Member #8 Correlation with other scores of quality were good, either positively or negatively correlated. To the extent that the other measures are good measures, this one tends to correlate. It was nice to see medical record validation for claims data.

Panel Member #9 The QM Stewards demonstrated moderate validity using a unique correlation between the HF RSRR and the Star Ratings. Strong correlation with the EDAC score was a better comparison for supporting validity.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

27. **What is the level of certainty or confidence that the empirical analysis demonstrates that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct?**

☐ **High**

☐ **Moderate**

☐ **Low**

☐ **Insufficient**

28. **Briefly explain rationale for rating of EMPIRICAL ANALYSES TO SUPPORT COMPOSITE CONSTRUCTION**

ADDITIONAL RECOMMENDATIONS

29. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**

Panel Member #9 See comments related to evaluating the risk adjustment findings related to social risk factors.

Developer Submission

NQF #: 0330

Corresponding Measures:

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

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

De.3. Brief Description of Measure: The measure estimates a hospital-level risk-standardized readmission rate (RSRR) for patients discharged from the hospital with a principal diagnosis of heart failure (HF). Readmission is defined as unplanned readmission for any cause within 30 days of the discharge date for the index admission. Readmissions are classified as planned and unplanned by applying the planned readmission algorithm. The target population is patients age 65 and over. The Centers for Medicare & Medicaid Services (CMS) annually reports the measure for patients who are 65 years or older and are enrolled in fee-for-service (FFS) Medicare and hospitalized in non-federal hospitals or are patients hospitalized in Veterans Health Administration (VA) facilities.

1b.1. Developer Rationale: The goal of this measure is to improve patient outcomes. 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. Readmissions following HF are influenced by 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; several studies have demonstrated that appropriate, timely, and high-quality treatment can 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.

By providing patients, physicians, hospitals, and policy makers with information about hospital risk-standardized readmission rates following hospitalization for HF, HF readmission is a priority area for outcomes measure development. 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 readmissions. We define readmissions as any inpatient acute care admission, with the exception of certain planned readmissions, within 30 days from the date of discharge from an index admission with a principal discharge diagnosis of HF in patients 65and older. 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.

Additional details are provided in S.5 Numerator Details.

S.6. Denominator Statement: The cohort includes admissions for patients aged 65years 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 is publicly reported by CMS for those patients 65 years and older who are Medicare FFS or VA beneficiaries admitted to non-federal or VA hospitals, respectively.

Additional details are provided in S.7 Denominator Details

S.8. Denominator Exclusions: The 30-day HF readmission measure excludes index admissions for patients:

1. Without at least 30 days of post-discharge enrollment in Medicare FFS (in the case of patients who are not VA beneficiaries);
2. Discharged against medical advice (AMA);
3. Admitted within 30 days of a prior index admission for HF; and
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.

De.1. Measure Type: Outcome

S.17. Data Source: Claims, Enrollment Data, Other

S.20. Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: May 15, 2008 **Most Recent Endorsement Date:** Dec 09, 2016

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, all-cause, 30-day, risk-standardized mortality (RSMR) following HF hospitalization.

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

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

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

NQF_evidence_HFreadmission_Fall2020_final_7.22.20.docx

1a.1 *For Maintenance of Endorsement:* Is there new evidence about the measure since the last update/submission?

Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

Yes

1a. Evidence (sub criterion 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:

Date of Submission: **11/2/2020**

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

Outcome

☒ Outcome: 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization

☐ Patient-reported outcome (PRO):

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

☐ Intermediate clinical outcome (*e.g., lab value*):

☐ Process:

☐ Appropriate use measure:

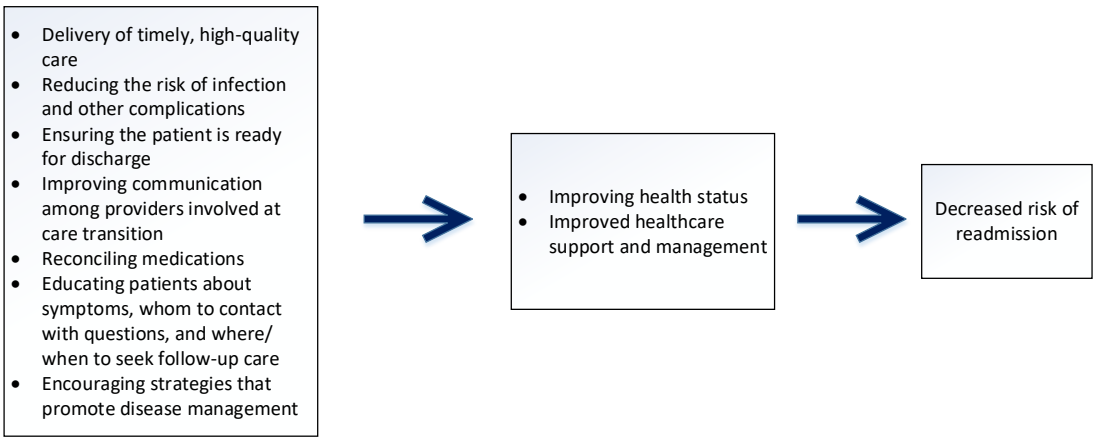
☐ Structure:

☐ Composite:

1a.2 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

The goal of this measure is to improve patient outcomes by providing patients, physicians, and hospitals with information about hospital-level, risk-standardized readmission rates following 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; and 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.

Figure 1: HF Logic Model



1a.3 Value and Meaningfulness: IF this measure is derived from patient report, provide evidence that the target population values the measured ***outcome, process, or structure*** and finds it meaningful. (Describe how and from whom their input was obtained.)

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

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

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES - Provide empirical data demonstrating the relationship between the outcome (or PRO) to at least one healthcare structure, process, intervention, or service.

The incidence rate of heart failure (HF) approaches 10 per 1,000 in patients 65 years and older (NHLBI et al., 2007), and continues to be one of the most common discharge diagnoses among the elderly (Jessup and Brozena et al., 2003). Prevalence of HF in the U.S. is estimated to be more than 6 million cases (Mozaffarian et al., 2015, Lloyd-Jones et al., 2009; Jackson et al., 2018; Benjamin et al., 2020), and is suspected to be the leading cause of death in people over age 65 (Hines et al., 2014). The lifetime risk of HF is estimated at 1 in 5 at 40 years of age, and the prevalence in the aging US population is expected to increase by 46% by 2030 (Heidenreich 2013). Total direct medical costs of HF were estimated at $30.7 billion in 2012 and are projected to increase by approximately 127% to $69.7 billion by 2030 (Jackson et al., 2018; Heidenreich et al., 2013).

Clinical experience suggests that the care for these patients is highly variable, and studies indicate there are gaps in the quality of hospital care—particularly in the transition to outpatient care (Albert 2009, Jha 2005; Patel et al., 2018). Moreover, there is substantial inter-hospital variation in the risk of readmission that is not clearly explained by differences in case mix (Lahewala et al., 2018; Roshanghalb et al., 2019). 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; and patient safety and coordinated transitions to the outpatient environment all contribute to patient outcomes but are difficult to measure by individual process measures.

The HF risk-standardized readmission rate 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.

The diagram above shows 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% (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, 1999; van Walraven et al., 2002; Weiss et al., 2010; Krumholz et al., 2002; Bradley et al., 2013; Radhakrishnan et al., 2018). Perhaps 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. As a patient advocate, they were responsible for patient education, follow-up, medication reconciliation, and preparing individualized discharge instructions sent to the patient's primary care provider (Jack et al., 2009; Patel et al., 2018). 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 (Jack et al., 2009). 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 (Nelson et al., 2000). Another study found that transitional care models prioritizing effective collaboration across providers/facilities through follow-up calls, patient tracking through medical charts, and team communication within and across facilities/providers, may reduce readmissions after AMIs and other conditions (Radhakrishnan et al., 2018).

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 (Fisher et al., 1994). Therefore, this measure does not risk adjust for such hospital characteristics.

References:

Albert NM, Yancy CW, Liang L, et al. Use of aldosterone antagonists in heart failure. *JAMA.* 2009;302(15):1658-1665.

Benjamin EJ, Muntner P, Alonso A, et al. Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association [published correction appears in Circulation. 2020 Jan 14;141(2):e33]. Circulation. 2019;139(10):e56–e528. doi:10.1161/CIR.0000000000000659.

Bradley EH, Curry L, Horwitz LI, et al. Hospital strategies associated with 30-day readmission rates for patients with heart failure. *Circ Cardiovasc Qual Outcomes*. 2013;6(4):444-450. doi:10.1161/CIRCOUTCOMES.111.000101.

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.

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.

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.

Heidenreich PA, Albert NM, Allen LA, et al. Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. Circ Heart Fail. 2013;6(3):606–619. doi:10.1161/HHF.0b013e318291329a.

Hines AL, Barrett ML, Jiang HJ, Steiner CA. Conditions with the largest number of adult hospital readmissions by payer, 2011 Rockville, MD: Agency for Healthcare Research and Quality; 2014. [updated 2014 Apr; accessed 2016 Apr 1]. Available from: https://www.hcup-us.ahrq.gov/reports/statbriefs/sb172-Conditions-Readmissions-Payer.jsp

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.

Jackson SL, Tong X, King RJ, Loustalot F, Hong Y, Ritchey MD. National Burden of Heart Failure Events in the United States, 2006 to 2014. *Circ Heart Fail.* 2018;11(12):e004873. doi:10.1161/CIRCHEARTFAILURE.117.004873.

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.

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

Jha AK, Li Z, Orav EJ, Epstein AM. Care in U.S. Hospitals — The Hospital Quality Alliance Program. *New England Journal of Medicine*. 2005;353(3):265-274.

Kao, D.P., J. Lindenfeld, D. Macaulay, H.G. Birnbaum, J.L. Jarvis, U.S. Desai, and R.L. Page, 2nd, Impact of a Telehealth and Care Management Program on All-Cause Mortality and Healthcare Utilization in Patients with Heart Failure. Telemed J E Health, 2016. 22(1): p. 2-11.

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.

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.

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.

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.

Lahewala S, Arora S, Tripathi B, et al. Heart failure: Same-hospital vs. different-hospital readmission outcomes [published correction appears in Int J Cardiol. 2020 Jun 15;309:100]. *Int J Cardiol.* 2019;278:186-191. doi:10.1016/j.ijcard.2018.12.043.

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.

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

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.

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.

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.

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.

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.

Patel PH, Dickerson KW. Impact of the Implementation of Project Re-Engineered Discharge for Heart Failure patients at a Veterans Affairs Hospital at the Central Arkansas Veterans Healthcare System. *Hosp Pharm*. 2018;53(4):266-271. doi:10.1177/0018578717749925.

Radhakrishnan K, Jones TL, Weems D, Knight TW, Rice WH. Seamless Transitions: Achieving Patient Safety Through Communication and Collaboration. *J Patient Saf.* 2018;14(1):e3-e5.

Roshanghalb A, Mazzali C, Lettieri E. Multi-level models for heart failure patients' 30-day mortality and readmission rates: the relation between patient and hospital factors in administrative data. *BMC Health Serv Res*. 2019;19(1):1012. Published 2019 Dec 30. doi:10.1186/s12913-019-4818-2.

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.

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.

1a.3. SYSTEMATIC REVIEW(SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES, INCLUDING THOSE THAT ARE INSTRUMENT-BASED) If the evidence is not based

on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

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

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

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

*cell intentionally left blank

_____

1a.4 OTHER SOURCE OF EVIDENCE

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

N/A

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

N/A

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

N/A

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

N/A

1b. Performance Gap

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

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

1b.1. Briefly explain the rationale for this measure *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

If a COMPOSITE *(e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.*

The goal of this measure is to improve patient outcomes. 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. Readmissions following HF are influenced by 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; several studies have demonstrated that appropriate, timely, and high-quality treatment can 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.

By providing patients, physicians, hospitals, and policy makers with information about hospital risk-standardized readmission rates following hospitalization for HF, HF readmission is a priority area for outcomes measure development. 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 maintenance of endorsement****. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

Variation in readmission rates indicates opportunity for improvement. We conducted analyses using data from July 1, 2016 to June 30, 2019 Medicare claims and VA administrative data (n= 1,286,352 admissions from 4,642 hospitals).

The three-year hospital-level risk standardized readmission rates (RSRRs) have a mean of 22.0% and range from 16.7-31.2% in the study cohort. As shown below, the median risk-standardized rate is 21.9% (20th and 70th percentiles are 21.0% and 22.4%, respectively).

Distribution of Hospital Heart Failure RSRRs over Different Time Periods (All Hospitals)

Results for each data year Characteristic//07/2016-06/2017//07/2017-06/2018//07/2018-06/2019//07/2016-06/2019

Number of Hospitals//4537//4516//4483//4642

Number of Admissions//420222//434269//431861//1286352

Mean(SD)//21.9(1.1)//22(1.1)//22(1.1)//22(1.4)

Range(Min-Max)//17.3-28.8//17.1-29.6//17.1-28.3//16.7-31.2

Minimum//17.3//17.1//17.1//16.7

10th percentile//20.7//20.8//20.8//20.3

20th percentile//21.2//21.3//21.3//21.0

30th percentile//21.4//21.6//21.6//21.4

40th percentile//21.7//21.8//21.7//21.6

50th percentile//21.8//21.9//21.9//21.9

60th percentile//22.0//22.1//22.1//22.1

70th percentile//22.2//22.4//22.3//22.4

80th percentile//22.5//22.7//22.6//22.9

90th percentile//23.1//23.3//23.2//23.7

Maximum//28.8//29.6//28.3//31.2

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 maintenance of endorsement****. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

Distribution of 30-day HF RSRRs by Proportion of Dual Eligible Patients:

Data Source: Medicare FFS claims, VA claims and Medical Beneficiary Summary File (MBSF) data

Dates of Data: July 2016 through June 2019

Variation in RSRRs across hospitals (with at least 25 cases) by proportion of patients with social risk//

Description of Social Risk Variable//Dual Eligibility

Quartile//Q1//Q4

Social Risk Proportion (%)//(0-8.42)//(35.14-100)

# of Hospitals//927//928

100%Max//27.0//31.2

90%//23.4//24.3

75%//22.4//23.2

50%//21.6//22.3

25%//20.7//21.5

10%//19.8//20.7

0%Min//16.7//18.4

Distribution of 30-day HF RSRRs by Proportion of Patients with AHRQ SES Index Scores:

Data Source: Medicare FFS claims, VA claims and The American Community Survey (2013-2017) data

Dates of Data: July 2016 through June 2019

Variation in RSRRs across hospitals (with at least 25 cases) by proportion of patients in lower and upper social risk quartiles//

Description of Social Risk Variable //AHRQ SES Index

Quartile//Q1//Q4

Social Risk Proportion (%)//(0-10.22)//(24.46-100)

# of Hospitals//928//928

100%Max//31.2//28.8

90%//23.8//24.3

75%//22.6//23.1

50%//21.7//22.3

25%//20.7//21.5

10%//19.7//20.9

0%Min//16.7//17.9

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

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 sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

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

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

Cardiovascular, Cardiovascular : Congestive Heart Failure

De.6. Non-Condition Specific*(check all the areas that apply):*

Care Coordination, Care Coordination : Readmissions, Care Coordination : Transitions of Care, Safety

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

Elderly, Populations at Risk

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

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: NQF_datadictionary_HFreadmission_Fall2020_final_7.22.20.xlsx

S.2c. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

No, this is not an instrument-based measure **Attachment:**

S.2d. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

Not an instrument-based measure

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

No

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

Updates consisted of updating the specifications to include new and modified ICD-10 CM/PCS codes.

S.4. Numerator Statement *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*

IF an OUTCOME MEASURE*, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).*

The outcome for this measure is 30-day readmissions. We define readmissions as any inpatient acute care admission, with the exception of certain planned readmissions, within 30 days from the date of discharge from an index admission with a principal discharge diagnosis of HF in patients 65and older. 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.

Additional details are provided in S.5 Numerator Details.

S.5. Numerator Details *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection,*

specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)

IF an OUTCOME MEASURE, *describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).*

The 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 using Medicare claims and VA administrative 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 HF readmission measure, CMS used the Planned Readmission Algorithm without modifications.

The planned readmission algorithm and associated code tables are attached in data field S.2b (Data Dictionary or Code Table).

S.6. Denominator Statement *(Brief, narrative description of the target population being measured)*

The cohort includes admissions for patients aged 65years 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 is publicly reported by CMS for those patients 65 years and older who are Medicare FFS or VA beneficiaries admitted to non-federal or VA hospitals, respectively.

Additional details are provided in S.7 Denominator Details

S.7. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

IF an OUTCOME MEASURE*, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).*

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

1. Principal discharge diagnosis of HF;
2. Enrolled in Medicare fee-for-service (FFS) Part A and Part B for the 12 months prior to the date of admission, and enrolled in Part A during the index admission, or those who are VA beneficiaries;
3. Aged 65 or over;
4. Discharged alive from a non-federal short-term acute care hospital or VA hospital; and,
5. Not transferred to another acute care facility.

S.8. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*

The 30-day HF readmission measure excludes index admissions for patients:

1. Without at least 30 days of post-discharge enrollment in Medicare FFS (in the case of patients who are not VA beneficiaries);
2. Discharged against medical advice (AMA);
3. Admitted within 30 days of a prior index admission for HF; and
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.

S.9. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

The HF readmission measure excludes index admissions for patients:

1. Without at least 30 days of post-discharge enrollment in Medicare FFS (in the case of patients who are not VA beneficiaries), which is identified with enrollment data from the Medicare Enrollment Database.

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

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

Rationale: Providers did not have the opportunity to deliver full care and prepare the patient for discharge.

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

Rationale: Additional HF admissions within 30 days are excluded as index admissions because they are part of the outcome. A single admission does not count as both an index admission and a readmission for another index admission.

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, which are identified by the corresponding codes included in claims data (codes can be found in attached Data Dictionary).

Rationale: Patients with these procedures are a clinically distinct group with a different risk of the readmission outcome.

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

N/A

S.11. Risk Adjustment Type (Select type. Provide specifications for risk stratification in measure testing attachment)

Statistical risk model

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Lower score

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

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 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 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" readmissions to the number of "expected" readmissions, multiplied by the national unadjusted readmission rate. For each hospital, the numerator of the ratio ("predicted") 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 ("expected") is the number of readmissions expected on the basis of 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, or better quality, and a higher ratio indicates higher-than-expected readmission, 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 effect is added to the sum of the estimated regression coefficients multiplied by the patient characteristics. The results are log transformed and summed over all patients attributed to a hospital to get a predicted value. The "expected" number of 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 log transformed and summed over all patients in the hospital to get an expected value. To assess hospital performance for each reporting period, we re-estimate the model coefficients using the years of data in that period.

This calculation transforms the ratio of predicted over expected into a rate that is compared to the national observed readmission rate. The hierarchical logistic regression models are described fully in the original methodology report (Krumholz et al., 2005).

References:

1. Normand S-LT, Shahian DM. 2007. Statistical and Clinical Aspects of Hospital Outcomes Profiling. Stat Sci 22(2): 206-226.
2. Krumholz H, Normand S, Galusha D, et al. Risk-Adjustment Models for HF and HF 30-Day Readmission Methodology. 2005.

S.15. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*

IF an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.

N/A. This measure is not based on a sample or survey.

S.16. Survey/Patient-reported data *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*

Specify calculation of response rates to be reported with performance measure results.

N/A

S.17. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*

If other, please describe in S.18.

Claims, Enrollment Data, Other

S.18. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument (e.g., name of database, clinical registry, collection instrument, etc., and describe how data are collected.)*

IF instrument-based, identify the specific instrument(s) and standard methods, modes, and languages of administration.

Data sources for the Medicare FFS measure:

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

Medicare Enrollment Database (EDB): This database contains Medicare beneficiary demographic, benefit/coverage, and vital status information. This data source was used to obtain information on several inclusion/exclusion indicators such as Medicare status on admission as well as vital status. These data have previously been shown to accurately reflect patient vital status (Fleming et al., 1992). The Master Beneficiary Summary File (MBSF) is an annually created file derived from the EDB that contains enrollment information for all Medicare beneficiaries including dual eligible status. Years 2016-2019 were used.

Veterans Health Administration (VA) Data: This data source contains data for VA inpatient and outpatient services including: inpatient hospital care, outpatient hospital services, skilled nursing facility care, some home health agency services, as well as inpatient and outpatient physician data for the 12 months prior to and including each index admission. Unlike Medicare FFS patients, VA patients are not required to have been enrolled in Part A and Part B Medicare for the 12 months prior to the date of admission.

The American Community Survey (2013-2017): We used the American Community Survey (2013-2017) to derive an updated AHRQ SES index score at the patient nine-digit zip code level for use in studying the association between our measure and SRFs.

References

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.19. Data Source or Collection Instrument *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

No data collection instrument provided.

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

Facility

S.21. Care Setting *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*

Inpatient/Hospital

If other:

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

N/A

2. Validity – See attached Measure Testing Submission Form

NQF_testing_HFreadmission_Fall2020_final_11.02.20-637419006540287026.docx

2.1 *For maintenance of endorsement*

Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

Yes

2.2 *For maintenance of endorsement*

Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

Yes

2.3 *For maintenance of endorsement*

Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes social risk factors is not prohibited at present. Please update sections 1.8, 2a2, 2b1,2b4.3 and 2b5 in the Testing attachment and S.140 and S.11 in the online submission form. NOTE: These sections must be updated even if social risk factors are not included in the risk-adjustment strategy. You MUST use the most current version of the Testing Attachment (v7.1) -- older versions of the form will not have all required questions.

Yes - Updated information is included.

Measure Testing (subcriteria 2a2, 2b1-2b6)

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: **11/3/2020**

Type of Measure:

Measure	Measure (continued)
☒ Outcome (*including PRO-PM*)	☐ Composite – *STOP – use composite testing form*
☐ Intermediate Clinical Outcome	☐ Cost/resource
☐ Process *(including Appropriate Use)*	☐ Efficiency
☐ Structure	*

*cell intentionally left blank

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.17*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ claims	☒ claims
☐ registry	☐ registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒ other: Medicare Enrollment Data, VHA Administrative Data	☒ other: Census Data/American Community Survey, VHA Administrative Data, Master Beneficiary Summary File, Cooperative Cardiovascular Project (CCP)

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 data used for testing included Medicare Parts A and B claims as well as the Medicare Enrollment Database (EDB). Additionally, census as well as enrollment data were used to assess socioeconomic factors (dual eligible variable obtained through enrollment data; Agency for Healthcare Research and Quality [AHRQ] socioeconomic status [SES] index obtained through census data). Veterans' Health Administration (VHA) data are also included in the testing dataset. 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 for testing 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.20*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other:	☐ other:

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, short-term acute care inpatient US hospitals (including territories) with Medicare fee-for-service (FFS) beneficiaries aged 65 years or over are included. In addition, for the testing data presented, VHA hospitals and their 65 years and older patients are included in the measure. 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 in Table 1.

Measure Development and Testing

For measure development, we used Medicare administrative claims data (2004). The dataset also included administrative data on each patient for the 12 months prior to the index admission and the 30 days following it. The dataset contained inpatient and facility outpatient claims and Medicare enrollment database (EDB) data. We randomly split the data (2004) into two equal samples: **the Development Dataset** and **Internal Validation Dataset.**

Measure Testing

For analytical updates for this measure, we used three-years of Medicare administrative claims data (July 2016 – June 2019). The dataset also included administrative data on each patient for the 12 months prior to the index admission and the 30 days following it. The dataset contained inpatient and facility outpatient claims and Medicare enrollment database (EDB) data. The dataset also included administrative data from the VHA as these hospitals are currently publicly reported for this measure.

Table 1. Dataset Descriptions

Dataset	Applicable Section in the Testing Attachment	Description of Dataset
Development and Validation Datasets (Medicare Fee-For-Service Administrative Claims Data)	Section 2b3 Risk Adjustment/Stratification **2b3.6. Statistical Risk Model Discrimination Statistics** **2b3.7. Statistical Risk Model Calibration Statistics**	Entire Cohort: Dates of Data: 2004 Number of admissions = 1,469,277 Number of measured hospitals: 4,730 This cohort was randomly split for initial model testing. First half of split sample - Number of Admissions: 283,919 - Number of Measured Hospitals: 4,669 Second half of split sample - Number of Admissions: 283,528 - Number of Measured Hospitals: 4,680
Testing Dataset (Medicare Fee-For-Service Administrative Claims Data and VA Administrative data (July 1, 2016 – June 30, 2019)	Section 2a2 Reliability Testing Section 2b1 Validity Testing Section 2b2 Testing of Measure Exclusion Section 2b3 Risk Adjustment/Stratification **2b3.6. Statistical Risk Model Discrimination Statistics** Section 2b4 Meaningful Differences	Dates of Data: July 2016 – June 2019 Number of admissions = 1,286,352 Patient Descriptive Characteristics: mean age = 80.5 years; % male = 48.4 Number of measured hospitals: 4,642

Dataset	Applicable Section in the Testing Attachment	Description of Dataset
The American Community Survey (ACS)	Section 2b3: Risk adjustment/Stratification for Outcome or Resource Use Measures	Dates of Data: 2013-2017 We used the AHRQ SES index score derived from the American Community Survey (2013-2017) to study the association between the 30-day readmission outcome and SRFs. The AHRQ SES index score is based on beneficiary 9-digit zip code level of residence and incorporates 7 census variables found in the American Community Survey.
Master Beneficiary Summary File (MBSF)	Section 2b3: Risk adjustment/Stratification for Outcome or Resource Use Measures	Dates of Data: July 2016 – June 2019 We used dual eligible status (for Medicare and Medicaid) derived from the MBSF to study the association between the 30-day measure outcome and dual-eligible status.

1.8 What were the social risk factors that were available and analyzed? For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

We selected social risk factor (SRF) variables to analyze after reviewing the literature and examining available national data sources. We sought to find variables that are reliably captured for all patients in this measure. There is a large body of literature linking various SRFs to poorer health status and higher readmissions over a lifetime. Income, education, and occupation are the most commonly examined SRFs studied. The causal pathways for SRF variable selection are described below in Section 2b3.3a. Unfortunately, these variables are not available at the patient level for this measure, therefore proxy measures of income, education level and economic status were selected.

The SRF variables used for analysis were:

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

Following guidance from the Assistant Secretary for Planning and Evaluation (ASPE) and a body of literature demonstrating differential health care and health outcomes among dual eligible patients, we identified dual eligibility as a key variable (ASPE 2016, ASPE 2020). We recognize that Medicare-Medicaid dual eligibility has limitations as a proxy for patients' income or assets because it does not provide a range of results and is only a dichotomous outcome. However, the threshold for over 65-year-old Medicare patients is valuable, as it takes into account both income and assets and is consistently applied across states for the older population. We acknowledge that it is important to test a wider variety of SRFs including key variables such as education and

poverty level; therefore, we also tested a validated composite based on census data linked to as small a geographic unit as possible.

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

Finally, The AHRQ SES index score is a well-validated variable that describes the average SES of people living in small defined geographic areas (Bonito et al., 2008). Its value as a proxy for patient-level information is dependent on having the most granular-level data with respect to communities that patients live in. We considered the area deprivation index (ADI) among many other potential indicators when we initially evaluated the impact of SRF indicators. We ultimately did not include the ADI at the time, partly due to the fact that the coefficients used to derive ADI had not been updated for many years. Recently, the coefficients for ADI have been updated and therefore we compared the ADI with the AHRQ SES Index and found them to be highly correlated. In this submission, we present analyses using the census block level, the most granular level possible using American Community Survey (ACS) data. A census block group is a geographical unit used by the US Census Bureau which is between the census tract and the census block. It is the smallest geographical unit for which the bureau publishes sample data. The target size for block groups is 1,500 and they typically have a population of 600 to 3,000 people. We used 2013-2017 ACS data and mapped patients' 9-digit ZIP codes via vendor software to the census block group level. Given the variation in cost of living across the country, the median income and median property value components of the AHRQ SES Index were adjusted by regional price parity values published by the Bureau of Economic Analysis (BEA). This provides a better marker of low SES neighborhoods in high expense geographic areas. We then calculated an AHRQ SES Index score for census block groups that can be linked to 9-digit ZIP codes. We used the percentage of patients with an AHRQ SES index score equal to or below 42.7 to define the lowest quartile of the AHRQ SES Index.

References:

Adler NE, Newman K. Socioeconomic disparities in health: pathways and policies. Health affairs (Project Hope). 2002; 21(2):60-76.

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.

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.

Chen HF, Nevola A, Bird TM, et al. Understanding factors associated with readmission disparities among Delta region, Delta state, and other hospitals. Am J Manag Care May 2018. 25(5): e150-156.

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

Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation (ASPE). Second Report to Congress: Social Risk Factors and Performance in Medicare's Value-based Purchasing Programs. 2020; https://aspe.hhs.gov/system/files/pdf/263676/Social-Risk-in-Medicare%E2%80%99s-VBP-2nd-Report.pdf. Accessed July 2, 2020.

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.

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

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

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.

Kaplan CM, Thompson MP, Waters TM. How Have 30-Day Readmission Penalties Affected Racial Disparities in Readmissions?: an Analysis from 2007 to 2014 in Five US States. J Gen Intern Med. 2019;34(6):878-883.

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

Mackenbach JP, Cavelaars AE, Kunst AE, Groenhof F. Socioeconomic inequalities in cardiovascular disease mortality; an international study. European heart journal. 2000; 21(14):1141-1151.

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

Tonne C, Schwartz J, Mittleman M, Melly S, Suh H, Goldberg R. Long-term survival after acute myocardial infarction is lower in more deprived neighborhoods. Circulation. Jun 14 2005; 111(23):3063-3070.

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*)

Measure Score Reliability

We performed two types of reliability testing. First, we estimated the overall measure score reliability by calculating the intra-class correlation coefficient (ICC) using a split sample (i.e. test-retest) method. Second, we estimated the facility-level reliability (signal-to-noise reliability).

Split-Sample Reliability

The reliability of a measurement is the degree to which repeated measurements of the same entity agree with each other. For measures of hospital performance, the measured entity is naturally the hospital, and reliability is the extent to which repeated measurements of the same hospital give similar results. Accordingly, our approach to assessing reliability is to consider the extent to which assessments of a hospital using different but randomly selected subsets of patients produce 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, and then measured again using a second random subset exclusive of the first, and the agreement of the two resulting performance measures compared across hospitals (Rousson, Gasser, and Seifert, 2002).

For split-sample reliability of the measure in the aged 65 years and older population, we randomly sampled half of patients within each hospital for a three year period, 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 (Shrout & Fleiss, 1979), and assessed the values according to conventional standards (Landis & Koch, 1977). Specifically, we used a combined 2016-2019 sample, randomly split it into two approximately equal subsets of patients, and calculated the RSRR for each hospital for each sample. The agreement of the two RSRRs was quantified for hospitals in each sample using the intra-class correlation as defined by ICC (2,1). (Shrout & Fleiss, 1979).

Using two non-overlapping random samples provides a conservative 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 formula to estimate the reliability of the measure if the whole cohort were used, based on an estimate from half the cohort.

Signal-to-Noise

We estimated the signal to noise reliability (facility-level reliability), which is the reliability with which individual units (hospitals) are measured. While test re-test reliability is the most relevant metric from the perspective of overall measure reliability, it is also meaningful to consider the separate notion of "unit" reliability; that is, the reliability with which individual units (here, hospitals) are measured. The reliability of any one facility's measure score will vary depending on the number of patients admitted for HF. Facilities with more volume (i.e., with more patients) will tend to have more reliable scores, while facilities with less volume will tend to have less reliable scores. Therefore, we used the formula presented by Adams and colleagues (2010) to calculate facility-level reliability.

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

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

Additional Information

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 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; we seek to avoid variables which do not meet this standard.

In addition, CMS uses several hospital auditing programs to assess overall claims code accuracy and 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.

Furthermore, we assessed the variation in the frequency of the variables over time: detailed information is presented in the measure's 2020 Condition-Specific Measure Updates and Specifications Report cited below.

References

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

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

DeBuhr J, McDowell, K, Grady J, et al., 2020 Condition-Specific Measure Updates and Specifications Report Hospital-Level 30-Day Risk-Standardized Readmission Measures - Available at: https://www.qualitynet.org/inpatient/measures/readmission/methodology

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, 420-3428.

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*)

Measure Score Reliability Results

Split-Sample Reliability

In total, 1,286,352 admissions were included in the analysis, using 3 years of data. After randomly splitting the sample into two halves, there were 642,047 admissions from 4,593 hospitals in one half and 644,305 admissions from 4,642 hospitals in the other half. As a metric of agreement, we calculated the ICC for hospitals with 25 admissions or more. Using the Spearman-Brown prediction formula, the agreement between the two independent assessments of the RSRR for each hospital was 0.587.

Signal-to-Noise

We calculated the signal-to-noise reliability score for each hospital with at least 25 admissions* (see Table 2 below). The median reliability score was 0.57, ranging from 0.14 to 0.96. The 25th and 75th percentiles were 0.31 and 0.75, respectively. The median reliability score demonstrates moderate agreement.

Table 2. Signal-to-noise reliability distribution for HF readmission

Mean	Std. Dev.	Min	5th Percentile	10th Percentile	25th Percentile	Median	75th Percentile	90th Percentile	95th Percentile	Max
0.54	0.24	0.14	0.16	0.20	0.31	0.57	0.75	0.84	0.87	0.96

*Hospital scores are calculated for all hospitals (including those that have fewer than 25 admissions) but only publicly reported for those that have at least 25 admissions to ensure hospital results are reliable.

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?*)

Measure Score Reliability Results

The split-sample reliability score of 0.587, discussed in the previous section, represents the lower bound of estimate of the true measure reliability.

Using the approach used by Adams et. al. and Yu et al., we obtained the median signal-to-noise reliability score of 0.57, which demonstrates moderate agreement.

Our interpretation of the results is based on the standards established by Landis and Koch (1977):

< 0 – Less than chance agreement;

0 – 0.2 Slight agreement;

0.21 – 0.39 Fair agreement;

0.4 – 0.59 Moderate agreement;

0.6 – 0.79 Substantial agreement;

0.8 – 0.99 Almost Perfect agreement; and

1 Perfect agreement

Taken together, these results indicate that there is moderate reliability in the measure score.

In the absence of empirically supported standards, our position is that 'acceptability' depends on context. For simple concepts or constructs, such as a patient's weight, the expectation is that the test-retest reliability of a measure of that construct should be quite high. However, for complex constructs, such as clinical severity, patient comorbidity, or symptom profiles used to identify a condition or clinical state, reliability of measures used to define these constructs is quite a bit lower.

References:

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

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

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

2b1. VALIDITY TESTING

2b1.1. 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*) **NOTE**: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

2b1.2. 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)*

Empirical Validity

Stewards of NQF-endorsed measures going through the re-endorsement process are required to demonstrate external validity testing at the time of maintenance review, or, if this is not possible, to justify the use of face validity only. To meet this requirement for the HF readmission measure, we identified and assessed the measure's correlation with other measures that target the same domain of quality (e.g. complications, safety, or post-procedure utilization) for the same or similar populations. The goal was to identify if better performance in this measure was related to better performance on other relevant structural or outcomes measures. After literature review and consultations with measures experts in the field, there were very few measures identified that assess the same domains of quality. Given that challenge, we selected the following to use for validity testing.

1. Hospital Star Rating readmission group score: CMS's Overall Hospital Star Rating assesses hospitals' overall performance (expressed on Hospital Compare graphically, as stars) based on a weighted average of group scores from different domains of quality (mortality, readmissions, safety, patient experience, imaging, effectiveness of care, and timeliness of care). The readmission group is comprised of the readmission measures that are publicly reported on Hospital Compare. The readmission group score is derived from a latent-variable model that identifies an underlying quality trait for that group. For the validity testing presented in this testing form, we used readmission group scores from 4,642 Medicare FFS hospitals from July 2019. The full methodology for the Overall Hospital Star Rating can be found at https://www.qualitynet.org/inpatient/public-reporting/overall-ratings/resources.
2. Overall Hospital Star Rating: CMS's Overall Hospital Star Rating assesses hospitals' overall performance (expressed on Hospital Compare graphically, as stars) based on a weighted average of "group scores" from different domains of quality (mortality, readmissions, safety, patient experience, imaging, effectiveness of care, and timeliness of care). Each group is comprised of measures that are reported on Hospital Compare. Group scores for each individual group are derived from latent-variable models that identify an underlying quality trait for each group. Group scores are combined into an overall hospital score using fixed weights; overall hospital scores are then clustered, using k-means clustering, into five groups and are assigned one to five stars (the hospital's Star Rating). For the validity testing presented in this testing form, we used hospital's Star Ratings from 4,642 Medicare FFS hospitals from July 2019. The full methodology for the Overall Hospital Star Rating can be found at https://www.qualitynet.org/inpatient/public-reporting/overall-ratings/resources.
3. HF Excess Days in Acute Care (EDAC): The HF EDAC measure calculates the time spent for unplanned readmissions, observation stays, and emergency department visits for any reason in the 30 days following an index admission for HF. The EDAC measure presents a comprehensive picture of acute care utilization and the burden of these events on patients. The HF EDAC measure complements the HF readmission measure because it provides information on a broader range of unplanned acute care utilization following hospitalization. The EDAC measures expand on the readmission measures by including not only readmissions, but also ED visits and observation stays, to present a more comprehensive picture of acute care utilization. Moreover, by measuring days spent in acute care for any of these visits, the EDAC measures capture the burden of these events on patients. The full methodology for the HF EDAC measure can be found at https://www.qualitynet.org/inpatient/measures/edac/methodology.

We examined the relationship of performance in the HF readmission measure scores (RSRRs) with each of these external measures of hospital quality. For the external measures, the comparison was against performance within quartiles of the readmission group score or the EDAC score, or in the case of Star Ratings, to the Star Rating category (1-5 Stars). We predicted the HF readmission scores would be more strongly associated with the Hospital Star Rating readmission group score than the Overall Star Ratings scores, with lower RSRRs associated with better Star Ratings. With EDAC, we assume that lower RSRRs will be strongly associated with lower EDAC rates.

In addition to providing empirical evidence, we have found multiple sources that support that readmissions can represent a signal of hospital quality. Readmissions have been shown to be associated with low hospital quality. Hospitals that have adopted strategies to improve care processes such as discharge planning, patient education, and transitions of care, tend to perform better on these measures (e.g. Borza et al., 2019; Cyriac et al., 2016; Jack et al., 2009; Curry et al., 2011; Bradley et al., 2013; Koehler et al., 2009; Harrison et al., 2011; Hernandez et al., 2010; Kao et al., 2016; Radhakrishnan et al. 2018; Leppin et al., 2014; Patel et al., 2018; Ohar et al., 2018; Wright et al., 2019).

References

Borza T, Oerline MK, Skolarus TA, et al. Association Between Hospital Participation in Medicare Shared Savings Program Accountable Care Organizations and Readmission Following Major Surgery. *Ann Surg*. 2019;269(5):873-878. doi:10.1097/SLA.0000000000002737.

Bradley EH, Curry L, Horwitz LI, et al. Hospital strategies associated with 30-day readmission rates for patients with heart failure. *Circ Cardiovasc Qual Outcomes*. 2013;6(4):444-450. doi:10.1161/CIRCOUTCOMES.111.000101

Curry LA, Spatz E, Cherlin E, et al. What distinguishes top-performing hospitals in acute myocardial infarction mortality rates? A qualitative study. *Ann Intern Med*. 2011;154(6):384-390. doi:10.7326/0003-4819-154-6-201103150-00003

Cyriac, James MD; Garson, Leslie MD; Schwarzkopf, Ran MD; Ahn, Kyle MD; Rinehart, Joseph MD; Vakharia, Shermeen MD, MBA; Cannesson, Maxime MD, PhD; Kain, Zeev MD, MBA. Total Joint Replacement Perioperative Surgical Home Program: 2-Year Follow-Up, Anesthesia & Analgesia: July 2016 - Volume 123 - Issue 1 - p 51-62 doi: 10.1213/ANE.0000000000001308.

Harrison PL, Hara PA, Pope JE, Young MC, Rula EY. The impact of postdischarge telephonic follow-up on hospital readmissions. *Popul Health Manag*. 2011;14(1):27-32. doi:10.1089/pop.2009.0076

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. doi:10.1001/jama.2010.533

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.

Kao, D.P., J. Lindenfeld, D. Macaulay, H.G. Birnbaum, J.L. Jarvis, U.S. Desai, and R.L. Page, 2nd, Impact of a Telehealth and Care Management Program on All-Cause Mortality and Healthcare Utilization in Patients with Heart Failure. Telemed J E Health, 2016. 22(1): p. 2-11.

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. *J Hosp Med*. 2009;4(4):211-218. doi:10.1002/jhm.427

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.

Ohar JA, Loh CH, Lenoir KM, Wells BJ, Peters SP. A comprehensive care plan that reduces readmissions after acute exacerbations of COPD. *Respir Med.* 2018;141:20-25.

Patel PH, Dickerson KW. Impact of the Implementation of Project Re-Engineered Discharge for Heart Failure patients at a Veterans Affairs Hospital at the Central Arkansas Veterans Healthcare System. *Hosp Pharm*. 2018;53(4):266-271. doi:10.1177/0018578717749925.

Radhakrishnan K, Jones TL, Weems D, Knight TW, Rice WH. Seamless Transitions: Achieving Patient Safety Through Communication and Collaboration. *J Patient Saf.* 2018;14(1):e3-e5.

Wright EA, Graham JH, Maeng D, et al. Reductions in 30-day readmission, mortality, and costs with inpatient-to-community pharmacist follow-up. *J Am Pharm Assoc (2003).* 2019;59(2):178-186.

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. 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, comparing the predictive ability in 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:

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.

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.

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.

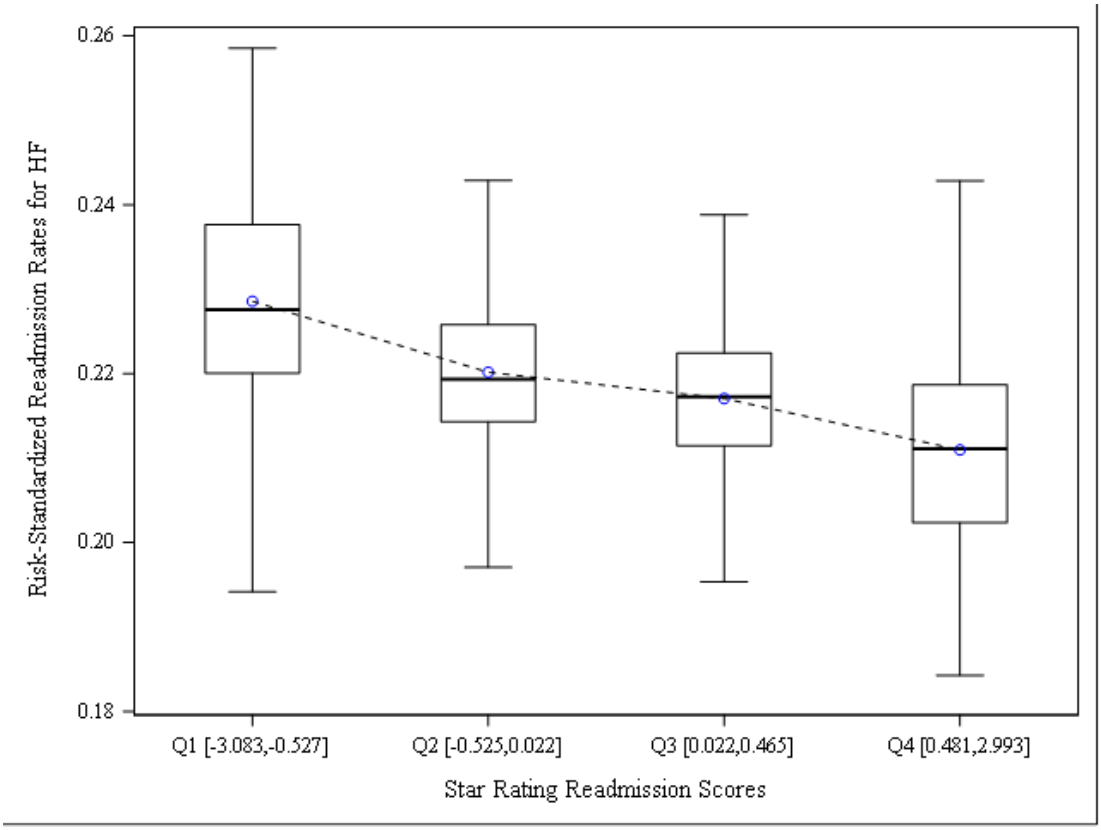
Shahian DM, He X, O'Brien S, et al. Development of a Clinical Registry-Based 30-Day Readmission Measure for Coronary Artery Bypass Grafting Surgery. Circulation 2014; DOI: 0.1161/CIRCULATIONAHA.113.007541. Published online before print June 10, 2014.

2b1.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Correlation between HF RSRRs and Star-Ratings Readmission Group Scores

Figure 1 shows the box-whisker plots of the HF readmission measure RSRRs within each quartile of Star-Rating readmission scores. The blue circles represent the mean RSRRs of Star-Rating readmission score quartiles. The correlation between HF RSRRs and Star-Rating readmissions score is -0.585, which suggests that hospitals with lower HF RSRRs are more likely to have higher Star-Rating readmission scores.

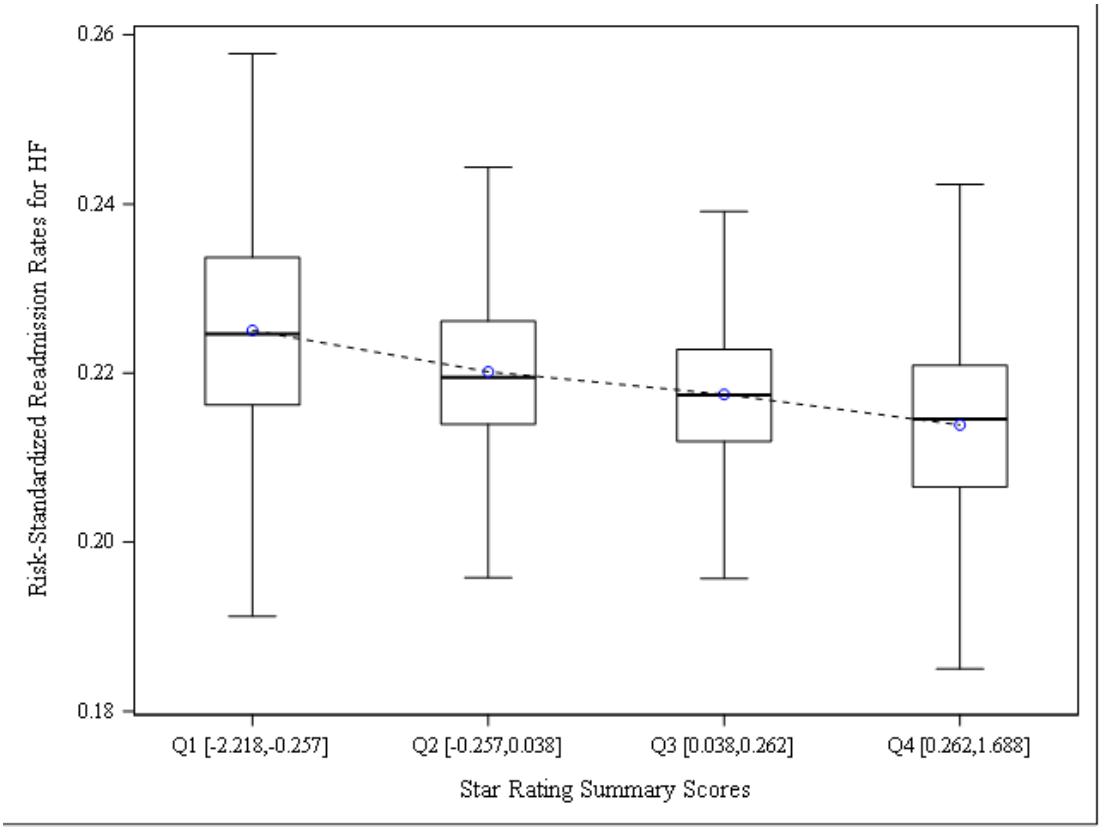
Figure 1 Box-whisker plots of the HF readmission measure RSRRs within each quartile of Star-Rating readmission scores



Correlation between HF RSRRs and Overall Star-Ratings Scores

Figure 2 shows the box-whisker plots of the HF readmission measure RSRRs within each quartile of Star-Rating summary scores. The blue circles represent the mean RSRRs of Star-Rating summary score quartiles. The correlation between HF RSRRs and Star-Rating summary score is -0.378, which suggests that hospitals with lower HF RSRRs are more likely to have higher Star-Rating summary scores.

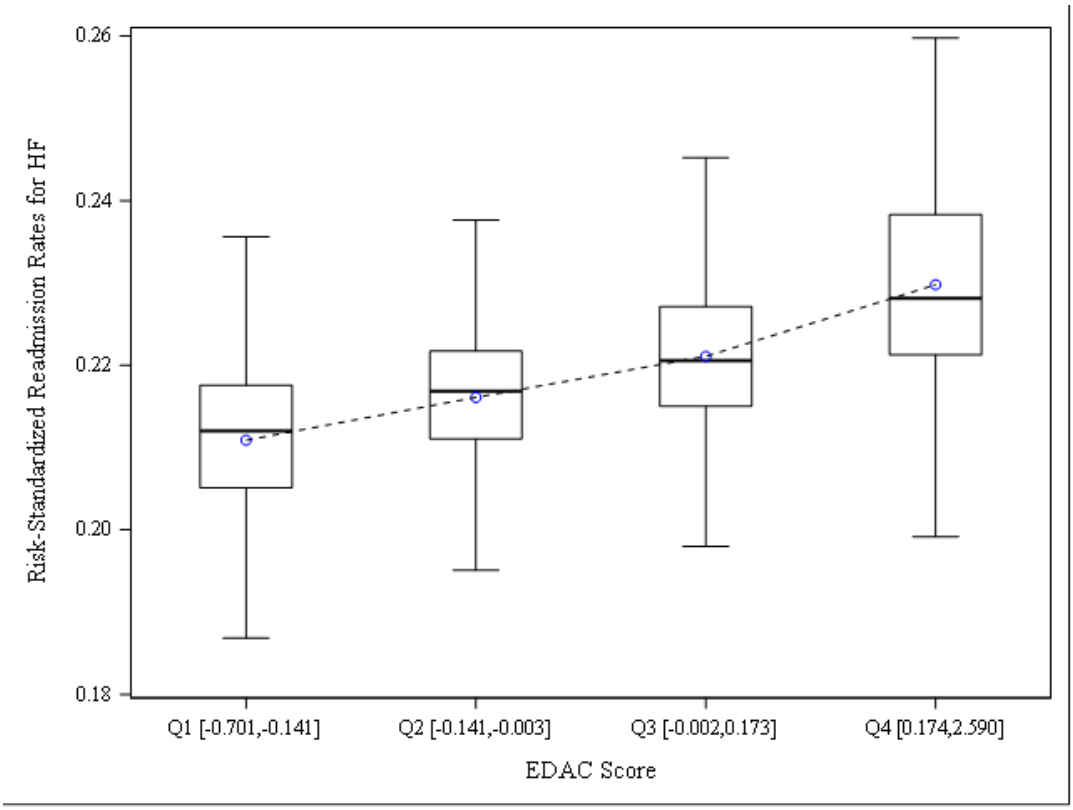
Figure 2 Box-whisker plots of the HF readmission measure RSRRs within each quartile of Star-Rating summary scores



Correlation between HF RSRRs and HF EDAC Scores

Figure 3 shows the Box-whisker plots of the HF readmission measure RSRRs and the HF EDAC scores. The blue circles represent the mean HF EDAC score quartiles. The correlation between HF RSRRs and HF EDAC scores is 0.574, which suggests that hospitals with lower HF RSRRs are more likely to have lower HF EDAC scores.

Figure 3 Box-whisker plots of the HF readmission measure RSRRs and the HF EDAC scores



Medical Record Validation Results

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.

2b1.4. 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

This validation approach compares the 30-day HF readmission measure results against the Star Ratings readmission and summary scores as well as the HF EDAC measure. Figure 1 and 2 Box Plots shown above demonstrate an observed trend of lower risk-standardized readmissions with higher Star Ratings and Figure 3 Box Plot results demonstrate an observed trend of higher readmission rates with higher excess days in acute care scores, which supports measure score validity. The correlation coefficients associated with the HF EDAC and star ratings readmission domain scores and the HF readmission measure indicate strong associations. A more moderate associated is seen with the overall star ratings score, which is to be expected given that all measures are calculated by complex statistical models. Overall, the results above show that the trend and direction of this association is in line with what would be expected.

Medical Record Validation

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.

2b2. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b3*

2b2.1. 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 (**Testing Dataset**). These exclusions are consistent with similar NQF-endorsed outcome measures. Rationales for the exclusions are detailed in data field S.9 (Denominator Exclusions).

2b2.2. 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 the **Testing Dataset (Table 3)**, below is the distribution of exclusions among hospitals with 25 or more admissions:

Table 3

Exclusion	N	%	Distribution across hospitals (N=3,766): Min, 25th, 50th, 75th, 99th percentile, Max
1. Discharged against medical advice (AMA)	8,592	0.60	(0.00, 0.00, 0.29, 0.83, 9.56)
2. Without at least 30 days post-discharge enrollment in FFS Medicare for index admission	9,569	0.67	(0.00, 0.00, 0.49, 0.98, 9.38)
3. HF admission within 30 days of a prior HF index admission	112,210	7.90	(0.00, 5.60, 7.27, 8.86, 19.2)
4. Patients 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 are excluded	4,426	0.31	(0.00, 0.00, 0.00, 0.11, 8.78)

2b2.3. 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 0.60% 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 when they leave AMA. Given that a very small percentage of patients are being excluded, it is unlikely this 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 0.67% of all index admissions excluded from the initial index 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.

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 7.9% of all index admissions excluded from the initial index cohort.

Exclusion 4 (patients with LVAD, history of LVAD, transplant, history of transplant) accounts for 0.31% 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 is excluded, this exclusion is unlikely to affect measure score.

2b3. 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 2b4.

2b3.1. 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 risk categories**

☐ **Other,**

2b3.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

See risk model specifications in Section 2b3.4a and the attached data dictionary.

2b3.2. If an outcome or resource use component measure is *not risk adjusted or stratified*, provide *rationale and analyses* to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

N/A. This measure is risk adjusted.

2b3.3a. Describe the conceptual/clinical *and* statistical methods and criteria used to select patient factors (clinical factors or social risk 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*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

Selecting Risk Variables

Our goal in selecting risk factors for adjustment was to develop parsimonious models that included clinically relevant variables strongly associated with the risk of readmission in the 30 days following an index admission. We used a two stage approach, first identifying the comorbidity or clinical status risk factors that were most important in predicting the outcome, then considering the potential addition of social risk factors.

The original measure was developed with ICD-9. When ICD-10 became effective in 2015, we transitioned the measure to use ICD-10 codes as well. 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 used to define this measure during development. A code set is attached in field S.2b. (Data Dictionary).

For risk model development, we started with Condition Categories (CCs) which are part of CMS's Hierarchical Condition Categories (HCCs). The current HCC system groups the 70,000+ ICD-10-CM and 17,000+ ICD-9-CM codes into larger clinically coherent groups (201 CCs) that are used in models to predict mortality or other outcomes (Pope et al. 2001; 2011). The HCC system groups ICD-codes into larger groups that are used in models to predict medical care utilization, mortality, or other related measures.

To select candidate variables, a team of clinicians reviewed all CCs and excluded those that were not relevant to the Medicare population or that were not clinically relevant to the readmission outcome (for example, attention deficit disorder, female infertility). All potentially clinically relevant CCs were included as candidate variables and, consistent with CMS's other claims-based readmission measures, some of those CCs were then combined into clinically coherent CC groupings.

To inform final variable selection, a modified approach to stepwise logistic regression was performed. The Development Sample was used to create 1,000 "bootstrap" samples. For each sample, we ran a logistic stepwise regression that included the candidate variables. The results (not shown in this report) were summarized to show the percentage of times that each of the candidate variables was significantly associated with readmission ($p<0.01$) in each of the 1,000 repeated samples (for example, 90 percent would mean that the candidate variable was selected as significant at $p<0.01$ in 90 percent of the times). We also assessed the direction and magnitude of the regression coefficients.

The clinical team reviewed these results and decided to retain risk adjustment variables above a predetermined cutoff because they demonstrated a strong and stable association with risk of readmission and were clinically relevant. Additionally, specific variables with particular clinical relevance to the risk of readmission were forced into the model (regardless of percent selection) to ensure appropriate risk adjustment for HF. These included variables representing markers for end of life/frailty, such as:

Markers for end of life/frailty:

- Decubitus Ulcer or Chronic Skin Ulcer (CC 157-CC 161)
- Metastatic and Other Severe Cancers (CC 9-CC 14)
- Hemiplegia, Paraplegia, Paralysis, Functional Disability (CC 70-CC 74, CC 103, CC 104, CC 189-CC 190)
- Stroke (CC 99-CC 100)
- Chronic Kidney Disease, Stage 5 (CC 136)

This resulted in a final risk-adjustment model that included 37 variables.

We weigh SRF adjustment using a comprehensive approach that evaluates the following:

- Well-supported conceptual model for influence of SRFs on measure outcome (detailed below);
- Feasibility of testing meaningful SRFs in available data (section 1.8); and
- Empiric testing of SRFs (section 2b3.4b).

Below, we summarize the findings of the literature review and conceptual pathways by which social risk factors may influence risk of the outcome, as well as the statistical methods for SRF empiric testing. Our conceptualization of the pathways by which patients' social risk factors affect the outcome is informed by the literature cited below and IMPACT Act–funded work by the National Academy of Science, Engineering and Medicine (NASEM) and the Department of Health and Human Services Assistant Secretary for Policy and Evaluation (ASPE).

Causal Pathways for Social Risk Variable Selection

Although some recent literature evaluates the relationship between patient SRFs and the readmission outcome, few studies directly address causal pathways or examine the role of the hospital in these pathways (see, for example, Chang et al 2007; Kim et al., 2007; Lindenauer et al., 2013; Trivedi et al., 2014; Buntin et al., 2017; Hamadi et al., 2019; Lahewala et al., 2018; Kosar et al., 2020). 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 social risk factors that have been examined in the literature can be categorized into three domains: (1) patient-level variables, (2) neighborhood/community-level variables, and (3) hospital-level variables.

Patient-level variables describe characteristics of individual patients, and include patient's income or education level (Eapen et al., 2015). Neighborhood/community-level variables use information from sources such as the American Community Survey as either a proxy for individual patient-level data or to measure environmental factors. Studies using these variables use one-dimensional measures such as median household income or composite measures such as the AHRQ-validated SES index score (Blum et al., 2014). Some of these variables may include the local availability of clinical providers (Herrin et al., 2015; Herrin et al., 2016). 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 et al., 2013; Jha et al., 2013).

The conceptual relationship, or potential causal pathways by which these possible social 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. **Patients with social risk factors may have worse health at the time of hospital admission**. Patients who have lower income/education/literacy or unstable housing may have a worse general health status and may present for their hospitalization or procedure with a greater severity of underlying illness. These social risk factors, which are characterized by patient-level or neighborhood/community-level (as proxy for patient-level) variables, may contribute to worse health status at admission due to competing priorities (restrictions based on job), lack of access to care (geographic, cultural, or financial), or lack of health insurance. Given that these risk factors all lead to worse general health status, this causal pathway should be largely accounted for by current clinical risk-adjustment.
2. **Patients with social risk factors often receive care at lower quality hospitals**. Patients of lower income, lower education, or unstable housing have inequitable access to high quality facilities, in part because such facilities are less likely to be found in geographic areas with large populations of poor patients. Thus, patients with low income are more likely to be seen in lower quality hospitals, which can explain increased risk of readmission following hospitalization.
3. **Patients with social risk factors may receive differential care within a hospital**. The third major pathway by which social risk factors may contribute to readmission risk is that patients may not receive equivalent care within a facility. For example, patients with social risk factors such as lower education may require differentiated care (e.g. provision of lower literacy information – that they do not receive).
4. **Patients with social risk factors may experience worse health outcomes beyond the control of the health care system.** Some social risk factors, such as income or wealth, may affect the likelihood of readmissions 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 financial priorities which don't allow for adequate recuperation or access to needed treatments, or a lack of access to care outside of the hospital.

Although we analytically aim to separate these pathways to the extent possible, we acknowledge that risk factors often act on multiple pathways, and, as such, individual pathways can be complex to distinguish analytically. Further, some social risk factors, despite having a strong conceptual relationship with worse outcomes, may not have statistically meaningful effects on the risk model. They also have different implications on the decision to risk adjust or not.

Based on this model and the considerations outlined in section 1.8 – namely, that the AHRQ SES index and dual eligibility variables aim to capture the SRFs that are likely to influence these pathways (income, education, housing, and community factors) - the following social risk variables were considered for risk-adjustment:

- Dual eligible status
- AHRQ SES index

Statistical Methods

We assessed the relationship between the SRF variables with the outcome and examined the incremental effect in a multivariable model. For this measure, we also examined the extent to which the addition of any one of these variables improved model performance or changed hospital results.

One concern with including SRFs 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). Identifying the relative contribution of the hospital level is important in considering whether a factor should be included in risk adjustment; if an effect is primarily a hospital-level effect, adjusting for it is equivalent to adjusting for differences in hospital quality. Thus, as an

additional step, we assessed whether there was a "contextual effect" at the hospital level. To do this, we performed a decomposition analysis to assess the independent effects of the SRF variables at the patient level and the hospital level. If, for example, the elevated risk of readmission for patients of low SES were largely 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 were 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 modeled the SRF variables as follows, let X_{ij} be a binary indicator of the SRF status of the i^{th} patient at the j^{th} hospital, and X_j the percent of patients at hospital j with $X_{ij} = 1$. Then we added both $X_{ij} \equiv X_{patient}$ and $X_j \equiv X_{hospital}$ to the model. 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 variable, $X_{hospital}$, 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, whether one effect dominates the other, 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 on the readmission rate of an average patient; and 2) a patient's SES 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 in the model is binary whereas the hospital's proportion of low SES patients is continuous. Therefore, in order to quantitatively compare the relative size of the patient and hospital effects, we calculated a range of predicted probabilities of readmission based on the fitted model.

Specifically, to estimate an average hospital effect, we calculated the predicted probabilities for the following scenarios: (1) Assuming all patients do not have the risk factor ($X_{ij} = 0$) and hospital level risk factor is at 5% percentile (P5) of all hospital values; (2) Assuming all patients do not have the risk factor and hospital level risk factor is at 95% percentile (P95); (3) Assuming all patients have the risk factor ($X_{ij} =1$) and hospital level risk factor is at 5% percentile (P5); (4) Assuming all patients have the risk factor and hospital level risk factor is at 95% percentile (P95). The average hospital effect is estimated by ((2)-(1) + (4)-(3))/2 (P95-P5). Then, to estimate an average patient effect, we first calculated the predicted probabilities by assuming patient-level risk factor equal to 0 or 1 at different hospital risk factor percentiles (0%, 5%, 10%, 25%, 50%, 75%, 90%, 95%, and 100%). Then at each of those percentiles, we could obtain the difference of predicted probabilities between all patients not having the risk factor and then all patients having the risk factor. We calculated the average of those differences in predicted probabilities ('delta') as the patient effect.

In summary, the difference in predicted probabilities at the 95th and 5th percentiles (P95-P5) estimates the hospital-level effect of the SRF on readmission. The difference in predicted probabilities when all patients have and do not have the SES risk factor (delta) estimates the patient-level effect of the SES risk factor on readmission. The hospital-level effect is greater than the patient-level effect when P95-P5 is greater than delta. We used P95 and P5 rather than the maximum (P100) and minimum (P0) to avoid outlier values.

We also performed the same analysis for several clinical covariates to contrast the relative contributions of patient- and hospital-level effects of clinical variables to the relative contributions for the SRFs.

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.

Buntin MB, Ayanian JZ. Social Risk Factors and Equity in Medicare Payment. *New England Journal of Medicine.* 2017;376(6):507-510.

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.

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

Chen HF, Nevola A, Bird TM, et al. Understanding factors associated with readmission disparities among Delta region, Delta state, and other hospitals. Am J Manag Care May 2018.24(5): e150-156.

Committee on Accounting for Socioeconomic Status in Medicare Payment Programs; Board on Population Health and Public Health Practice; Board on Health Care Services; Institute of Medicine; National Academies of Sciences, Engineering, and Medicine. Accounting for Social Risk Factors in Medicare Payment: Identifying Social Risk Factors. Washington (DC): National Academies Press (US); 2016 Jan 12. (https://www.ncbi.nlm.nih.gov/books/NBK338754/doi:10.17226/21858)

Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation. Report to Congress: Social Risk Factors and Performance under Medicare's Value-based Payment Programs. December 21, 2016. (https://aspe.hhs.gov/pdf-report/report-congress-social-risk-factors-and-performance-under-medicares-value-based-purchasing-programs).

Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation (ASPE). Second Report to Congress: Social Risk Factors and Performance in Medicare's Value-based Purchasing Programs. 2020; https://aspe.hhs.gov/system/files/pdf/263676/Social-Risk-in-Medicare%E2%80%99s-VBP-2nd-Report.pdf. Accessed July 2, 2020.

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, 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.

Hamadi H, Moody L, Apatu E, Vossos H, Tafili A, Spaulding A. Impact of hospitals' Referral Region racial and ethnic diversity on 30-day readmission rates of older adults. J Community Hosp Intern Med Perspect. 2019;9(3):181-188.

Herrin J, Kenward K, Joshi MS, Audet AM, Hines SJ. Assessing Community Quality of Health Care. Health Serv Res. 2016 Feb;51(1):98-116. doi: 10.1111/1475-6773.12322. Epub 2015 Jun 11. PMID: 26096649; PMCID: PMC4722214.

Herrin J, St Andre J, Kenward K, Joshi MS, Audet AM, Hines SC. Community factors and hospital readmission rates. Health Serv Res. 2015 Feb;50(1):20-39. doi: 10.1111/1475-6773.12177. Epub 2014 Apr 9. PMID: 24712374; PMCID: PMC4319869.

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.

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

Kosar CM, Loomer L, Ferdows NB, Trivedi AN, Panagiotou OA, Rahman M. Assessment of Rural-Urban Differences in Postacute Care Utilization and Outcomes Among Older US Adults. *JAMA Netw Open*. 2020;3(1):e1918738. Published 2020 Jan 3. doi:10.1001/jamanetworkopen.2019.18738.

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.

Lahewala S, Arora S, Tripathi B, et al. Heart failure: Same-hospital vs. different-hospital readmission outcomes [published correction appears in Int J Cardiol. 2020 Jun 15;309:100]. *Int J Cardiol*. 2019;278:186-191. doi:10.1016/j.ijcard.2018.12.043.

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.

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

Pope GC, Ellis RP, Ash AS, et al. Diagnostic cost group hierarchical condition category models for Medicare risk adjustment. Final Report to the Health Care Financing Administration under Contract Number 500-95-048. 2000; http://www.cms.hhs.gov/Reports/downloads/pope_2000_2.pdf. Accessed February 25, 2020.

Pope GC, Kautter J, Ingber MJ, et al. Evaluation of the CMS-HCC Risk Adjustment Model: Final Report. 2011; https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/downloads/evaluation_risk_adj_model_2011.pdf. Accessed February 25, 2020.

Roshanghalb A, Mazzali C, Lettieri E. Multi-level models for heart failure patients' 30-day mortality and readmission rates: the relation between patient and hospital factors in administrative data. *BMC Health Serv Res*. 2019;19(1):1012. Published 2019 Dec 30. doi:10.1186/s12913-019-4818-2.

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.

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:

☒ **Published literature**

☒ **Internal data analysis**

☐ **Other (please describe)**

2b3.4a. What were the statistical results of the analyses used to select risk factors?

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

Table 4. Adjusted Odds Ratios (ORs) and 95% Confidence Intervals (Cis) for the HF Readmission Hierarchical Logistic Regression Model over Different Time Periods in Testing Dataset

Variable	07/2016-06/2017 OR (95% CI)	07/2017-06/2018 OR (95% CI)	07/2018-06/2019 OR (95% CI)	07/2016-06/2019 OR (95% CI)
Age minus 65 (years above 65, continuous)	1.00 (0.99-1.00)	1.00 (0.99-1.00)	1.00 (0.99-1.00)	1.00 (1.00-1.00)
Male	1.01 (0.99-1.02)	1.00 (0.99-1.02)	0.99 (0.97-1.00)	1.00 (0.99-1.01)
History of coronary artery bypass graft (CABG) surgery	1.02 (1.01-1.04)	1.04 (1.02-1.06)	1.02 (1.00-1.04)	1.03 (1.02-1.04)
Metastatic cancer and acute leukemia (CC 8)	1.12 (1.07-1.17)	1.16 (1.11-1.22)	1.20 (1.15-1.25)	1.16 (1.13-1.19)
Cancer (CC 9-14)	1.03 (1.01-1.05)	1.04 (1.02-1.06)	1.03 (1.01-1.05)	1.03 (1.02-1.04)
Diabetes mellitus (DM) or DM complications (CC 17-19, 122-123)	1.07 (1.05-1.09)	1.07 (1.05-1.09)	1.08 (1.06-1.09)	1.07 (1.06-1.08)
Protein-calorie malnutrition (CC 21)	1.09 (1.06-1.11)	1.10 (1.07-1.12)	1.09 (1.06-1.11)	1.09 (1.07-1.10)
Other significant endocrine and metabolic disorders; disorders of fluid/electrolyte/acid-base balance (CC 23-24)	1.12 (1.10-1.14)	1.13 (1.11-1.15)	1.12 (1.10-1.14)	1.12 (1.11-1.13)
Liver or biliary disease (CC 27-32)	1.09 (1.07-1.12)	1.09 (1.06-1.11)	1.10 (1.08-1.13)	1.09 (1.08-1.11)
Peptic ulcer, hemorrhage, other specified gastrointestinal disorders (CC 36)	1.06 (1.04-1.08)	1.05 (1.03-1.07)	1.06 (1.03-1.08)	1.06 (1.04-1.07)
Other gastrointestinal disorders (CC 38)	1.05 (1.03-1.07)	1.08 (1.06-1.09)	1.05 (1.03-1.07)	1.06 (1.05-1.07)
Severe hematological disorders (CC 46)	1.22 (1.17-1.28)	1.14 (1.09-1.19)	1.18 (1.12-1.24)	1.18 (1.15-1.21)
Iron deficiency or other/unspecified anemias and blood disease (CC 49)	1.16 (1.14-1.18)	1.14 (1.12-1.16)	1.17 (1.15-1.19)	1.16 (1.14-1.17)
Dementia or other specified brain disorders (CC 51-53)	1.00 (0.98-1.02)	1.01 (0.99-1.03)	1.00 (0.98-1.02)	1.00 (0.99-1.01)
Drug/alcohol abuse/dependence/psychosis (CC 54-56)	1.11 (1.09-1.13)	1.13 (1.11-1.15)	1.11 (1.09-1.13)	1.12 (1.10-1.13)
Major psychiatric disorders (CC 57-59)	1.04 (1.01-1.07)	1.02 (1.00-1.05)	1.03 (1.01-1.06)	1.03 (1.01-1.04)
Depression (CC 61)	1.01 (0.99-1.03)	1.02 (1.00-1.04)	1.01 (0.99-1.03)	1.01 (1.00-1.03)
Other psychiatric disorders (CC 63)	1.07 (1.05-1.09)	1.07 (1.05-1.09)	1.07 (1.05-1.09)	1.07 (1.06-1.08)

Variable	07/2016-06/2017 OR (95% CI)	07/2017-06/2018 OR (95% CI)	07/2018-06/2019 OR (95% CI)	07/2016-06/2019 OR (95% CI)
Hemiplegia, paraplegia, paralysis, functional disability (CC 70-74, 103-104, 189-190)	1.08 (1.05-1.11)	1.06 (1.03-1.09)	1.07 (1.04-1.10)	1.07 (1.05-1.08)
Cardio-respiratory failure and shock (CC 84 plus ICD-10-CM codes R09.01 and R09.02, for discharges on or after October 1, 2015; CC 84 plus ICD-9-CM codes 799.01 and 799.02, for discharges prior to October 1, 2015)	1.12 (1.10-1.14)	1.11 (1.09-1.13)	1.11 (1.09-1.13)	1.12 (1.10-1.13)
Congestive heart failure (CC 85)	1.11 (1.09-1.14)	1.10 (1.08-1.13)	1.11 (1.08-1.13)	1.11 (1.09-1.12)
Acute coronary syndrome (CC 86-87)	1.11 (1.08-1.13)	1.10 (1.08-1.13)	1.11 (1.09-1.13)	1.11 (1.09-1.12)
Coronary atherosclerosis or angina (CC 88-89)	1.05 (1.03-1.07)	1.06 (1.04-1.08)	1.06 (1.04-1.08)	1.05 (1.04-1.07)
Valvular and rheumatic heart disease (CC 91)	1.07 (1.06-1.09)	1.07 (1.05-1.08)	1.07 (1.05-1.08)	1.07 (1.06-1.08)
Specified arrhythmias and other heart rhythm disorders (CC 96-97)	1.06 (1.04-1.08)	1.05 (1.03-1.07)	1.06 (1.05-1.08)	1.06 (1.05-1.07)
Other and unspecified heart disease (CC 98)	1.04 (1.03-1.06)	1.03 (1.01-1.04)	1.03 (1.01-1.05)	1.04 (1.03-1.05)
Stroke (CC 99-100)	1.00 (0.97-1.02)	1.01 (0.98-1.04)	1.03 (1.01-1.06)	1.01 (1.00-1.03)
Vascular or circulatory disease (CC 106-109)	1.05 (1.03-1.07)	1.07 (1.05-1.09)	1.07 (1.06-1.09)	1.06 (1.05-1.07)
Chronic obstructive pulmonary disease (COPD) (CC 111)	1.14 (1.12-1.16)	1.14 (1.12-1.16)	1.16 (1.14-1.18)	1.14 (1.13-1.16)
Fibrosis of lung or other chronic lung disorders (CC 112)	1.04 (1.02-1.07)	1.08 (1.05-1.10)	1.08 (1.05-1.10)	1.07 (1.05-1.08)
Asthma (CC 113)	1.06 (1.03-1.08)	1.04 (1.02-1.06)	1.02 (1.00-1.05)	1.04 (1.03-1.05)
Pneumonia (CC 114-116)	1.08 (1.06-1.10)	1.09 (1.07-1.11)	1.09 (1.08-1.11)	1.09 (1.08-1.10)
Dialysis status (CC 134)	1.14 (1.11-1.18)	1.16 (1.12-1.20)	1.14 (1.10-1.17)	1.14 (1.12-1.16)
Renal failure (CC 135-140)	1.24 (1.21-1.26)	1.25 (1.23-1.28)	1.24 (1.22-1.26)	1.24 (1.23-1.26)
Nephritis (CC 141)	1.04 (1.02-1.07)	1.02 (1.00-1.04)	1.06 (1.03-1.08)	1.04 (1.03-1.06)

Variable	07/2016-06/2017 OR (95% CI)	07/2017-06/2018 OR (95% CI)	07/2018-06/2019 OR (95% CI)	07/2016-06/2019 OR (95% CI)
Other urinary tract disorders (CC 145)	1.06 (1.04-1.08)	1.06 (1.04-1.08)	1.05 (1.03-1.07)	1.06 (1.05-1.07)
Decubitus ulcer or chronic skin ulcer (CC 157-161)	1.08 (1.06-1.10)	1.10 (1.08-1.12)	1.08 (1.05-1.10)	1.08 (1.07-1.10)

2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk 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.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

Throughout this section, we present new SRF testing results based on the current testing dataset (2020); in addition, we show prior analyses included in the 2016 endorsement maintenance forms for comparison purposes.

Variation in prevalence of the factor across measured entities in 2020 and 2016 (Table 5)

SRFs	2020 Prevalence % (IQR)	2016 Prevalence % (IQR)
Dual	**16.0% (9.60-25.1%)**	**17.5% (11.2-25.2%)**
AHRQ Low SES	**19.0% (6.50-36.1%)**	**19.0% (7.80-36.8%)**

The prevalence of social risk factors in the HF cohort varies across measured entities in 2020. The median percentage of dual eligible patients was 16.0% (9.60-25.1%) and 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 (lowest quartile) was 19.0% (6.50-36.1%) in 2020. These results are consistent with the 2016 results presented above.

Comparison of observed readmission rates in patients with and without social risk in 2020 and 2016 (Table 6)

SRFs	2020 Observed Rate	2016 Observed Rate
Dual (vs. Non-Dual)	24.8% (vs. 21.3%)	25.5% (vs. 21.9%)
AHRQ Low SES (vs. SES score above 42.7)	23.7% (vs. 21.5%)	24.3% (vs. 21.8%)

The patient-level observed HF readmission rates are higher for dual-eligible patients (24.8%) compared with 21.3% for non-dual patients in 2020. Similarly, the readmission rate for patients with an AHRQ SES index score equal to or below 42.7 was 23.7% compared with 21.5% for patients with an AHRQ SES index score above 42.7 in 2020. For both SRF variables, patient-level readmission rates have declined among all characteristic groups of patients.

Incremental effect of SRF variables in a multivariable model in 2020 and 2016

We examined the strength and significance of the SRF variables in the context of a multivariable model. When we include 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. In 2020, dual eligibility and the AHRQ SES index have effect sizes (odds ratios) of 1.07 and 1.07 when added independently to the model. Furthermore, the effect size of each variable is attenuated (1.06 and 1.06 for dual and SES) when both are added to the model.

We also find that the C-statistic is essentially unchanged with the addition of any of these variables into the model (Table 7).

Table 7

HF Readmission Models	2020 C-Statistic	2016 C-Statistic
Base Model: risk-adjusted model using the original clinical risk variables selected for the 2020 CMS public report of the HF readmission measure	**0.61**	**0.61**
Base Model plus AHRQ Low SES based on beneficiary residential 9-digit ZIP codes (SES9) as a social risk variable	**0.61**	**0.61**
Base Model plus dual eligibility (dual) as a social risk variable	**0.61**	**0.61**
Base Model plus SES9 and dual as social risk variables	**0.61**	**--**

Furthermore, we find that the addition of any of these variables into the hierarchical 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 was 0.023% (interquartile range [IQR] -0.017% – 0.028%) with a correlation coefficient between RSRRs for each hospital with and without dual eligibility added of 0.999. The median absolute change in hospitals' RSRRs when adding a low AHRQ SES Index score indicator to the model was 0.141% (IQR -0.132% – 0.147%) 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 0.980.

Contextual Effect Analysis

As described in 2b3.3a, we performed a decomposition analysis in 2020 and 2016 for each SRF 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 tables/figures below.

Both the patient-level and hospital-level dual eligibility, as well as low AHRQ SES Index effects, were significantly associated with HF readmission in the decomposition analysis. **The significance of the hospital-level effects indicates that if dual eligibility or low AHRQ SES Index variables were 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.**

To assess the relative contributions of the patient- and hospital-level effects, we calculated a range of predicted probabilities of readmission for the SRF variables and clinical covariates (comorbidities), as described in section 2b3.3a. The results are presented in the figures and table below (table of predicted probabilities for SES variables).

For SRF variables, the hospital-level effect (P95-P5) is considerably greater than the patient-level effect (delta) (Figures 4 and 5; predicted probabilities for SRF variables). For clinical variables, the patient-level effect (delta) is greater than the hospital-level effect (P95-P5) for renal failure (Figures 4 and 5; predicted probabilities for clinical variables). The hospital-level effect (P95-P5) is slightly greater than the patient-level effect (delta) for COPD (Figures 4 and 5; predicted probabilities for clinical variables). There is a consistent pattern demonstrating that SRFs have a much greater hospital-level effect than patient-level effect. The clinical variables (with the exception of COPD) had the opposite pattern, with a greater effect at the patient-level than

at the hospital-level for renal failure. In sum, including SRFs 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 SRFs might be related to poor outcomes. First, we find that although unadjusted rates of readmission are higher for patients of low SES, the addition of SRFs 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 SRFs. 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 SRFs in the decomposition analyses for 2020 and 2016 (Figures 4 and 5 below) 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. Direct adjustment for patient SRFs would essentially "over adjust" the measure, that is to say, it would be adjusting for an endogenous factor 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 SRFs have a distinct causal pathway in their impact on readmission risk.

Table 7. Parameter Estimates for Hospital-Level and Patient-Level in 2020 and 2016 from Decomposition Analysis

Parameter	2020 Estimate (standard error), p-value	2016 Estimate (standard error), p-value
Low SES census block group (AHRQ SES index linked to 9-digit ZIP – Adjusted for Cost of Living) – Patient Level	0.026 (0.006), <0.0001	0.055 (0.006), <0.0001
Low SES census block group (AHRQ SES index linked to 9-digit ZIP – Adjusted for Cost of Living) – Hospital Level	0.321 (0.020), <0.0001	0.246 (0.019), <0.0001
Dual-Eligible – Patient Level	0.410 (0.006), <0.0001	0.065 (0.006), <0.0001
Dual-Eligible – Hospital Level	0.316 (0.026), <0.0001	0.396 (0.036), <0.0001

Figure 4. Decomposition Analysis for 2020, HF Readmission

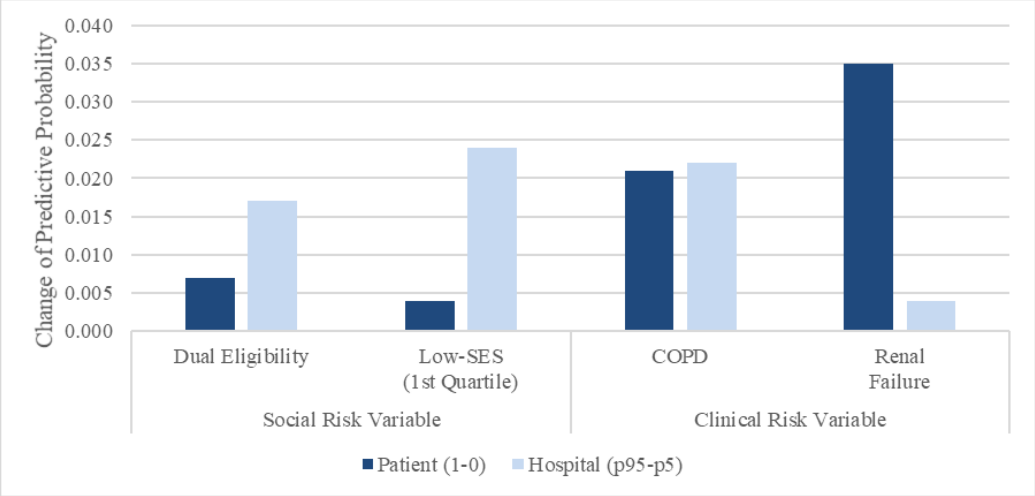
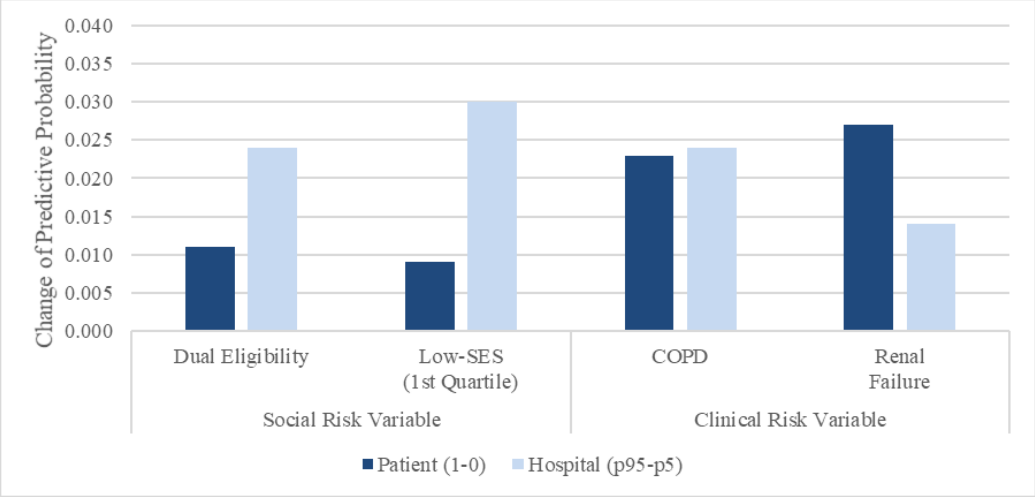


Figure 5. Decomposition Analysis for 2016, HF Readmission



Summary

We found wide variation in the prevalence of the two SRFs we examined, and we found that both 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 social risk. More importantly, we found that for both factors, there was a considerably greater hospital-level effect, compared with the patient-level effect, indicating that any patient-level adjustment alone would also adjust for quality differences between hospitals. Based on comprehensive investigation over the last five years, ASPE has also recommended that quality measures are not adjusted for SRFs (ASPE 2020). Given these empiric findings, ASPE's latest recommendations, and the fact that this is a hospital quality measure, CMS chose to not include these two SRFs in the final risk model at this time.

References:

Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation (ASPE). Second Report to Congress: Social Risk Factors and Performance in Medicare's Value-based Purchasing Programs. 2020; https://aspe.hhs.gov/system/files/pdf/263676/Social-Risk-in-Medicare%E2%80%99s-VBP-2nd-Report.pdf. Accessed July 2, 2020.

2b3.5. 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 <u>2b3.9</u>

Approach to assessing model performance

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, or the 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 **the development dataset** 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.

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

Development and Validation Dataset:

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)

Results for the Testing Cohort

C-statistic = 0.61

Predictive ability (lowest decile %, highest decile %): (12.5, 35.7)

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

2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

For the measure cohort, the results are summarized below:

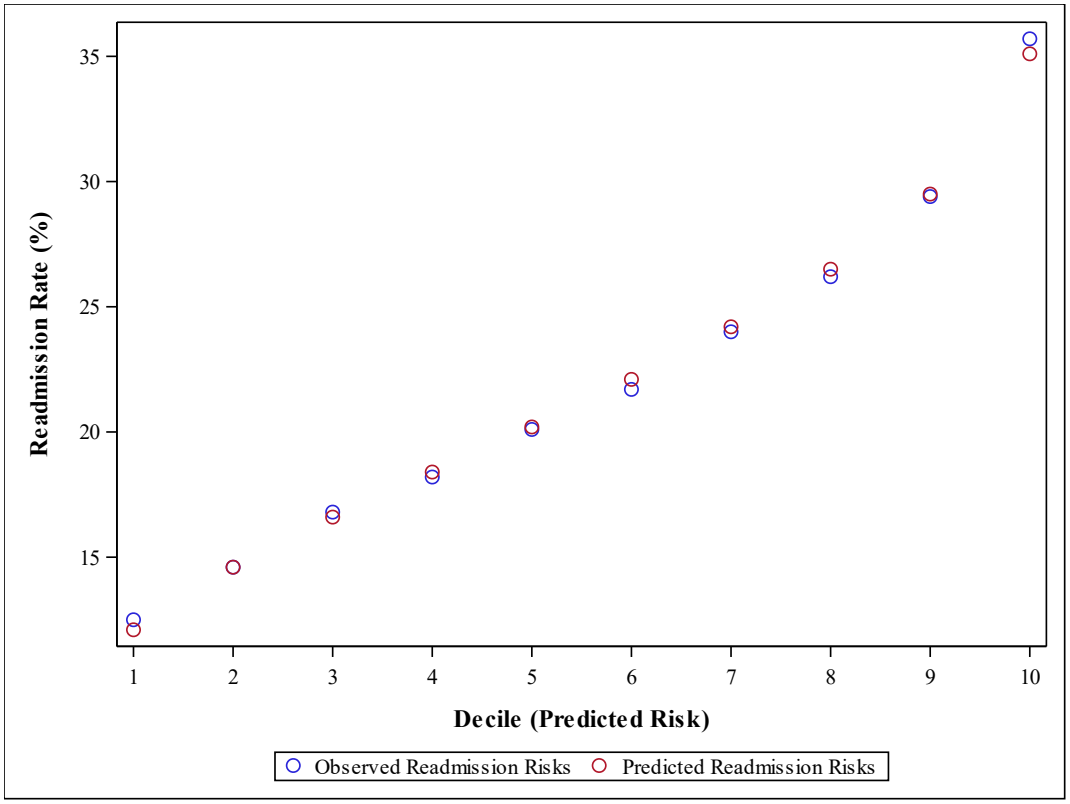
Development sample Calibration: (0, 1)

Validation sample Calibration: (-0.02, 1.01)

2b3.8. 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 2016 – June 2019 (Testing Dataset).

Figure 6



2b3.9. Results of Risk Stratification Analysis:

N/A

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

Discrimination Statistics

The C-statistic of 0.61 indicates fair model discrimination. The model indicated a wide range between the lowest decile and highest decile, indicating the ability to distinguish high-risk subjects from low-risk subjects.

Calibration Statistics

Over-fitting (Calibration γ0, γ1)

If the γ0 in the validation samples are substantially far from zero and the γ1 is substantially far from 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) and is comparable to other readmission outcome measures.

2b3.11. 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

2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b4.1. 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 measure score is hospital-specific risk-standardized readmission rates. These rates are obtained as the ratio of predicted to expected readmissions, multiplied by the national unadjusted rate. The "predicted" number of readmissions (the numerator) is calculated using the coefficients estimated by regressing the risk factors and the hospital-specific intercept on the risk of readmissions. The estimated hospital-specific intercept is added to the sum of the estimated regression coefficients multiplied by the patient characteristics. The results are then 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 then transformed and summed over all patients in the hospital to get an expected value. To assess hospital performance for each reporting period, we re-estimated the model coefficients using the years of data in that period.

We characterize the degree of variability by:

Reporting the distribution of RSRRs:

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.

Providing the median odds ratio (MOR) (Merlo et al, 2006). The median odds ratio represents the median increase in the odds of a readmission within 30 days of a HF admission date on a single patient if the admission occurred at a higher risk hospital compared to a lower risk hospital. MOR quantifies the between-hospital variance in terms of odds ratio; it is comparable to the fixed effects odds ratio.

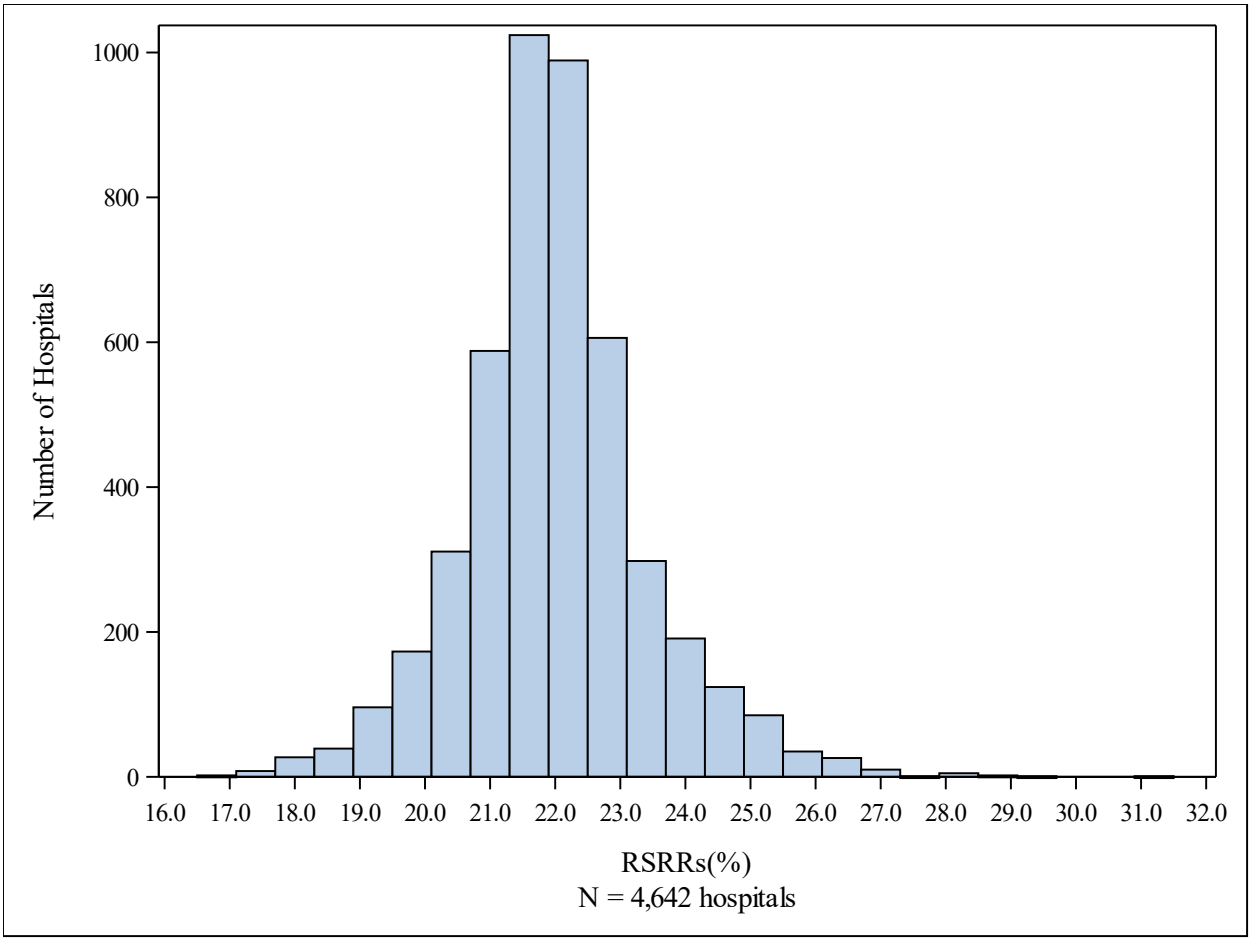
Reference:

Merlo J, Chaix B, Ohlsson H, Beckman A, Johnell K, Hjerpe P, Råstam L, Larsen K. (2006) A brief conceptual tutorial of multilevel analysis in social epidemiology: Using measures of clustering in multilevel logistic regression to investigate contextual phenomena. J Epidemiol Community Health, 60(4):290-7.

2b4.2. 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.

Figure 7. Distribution (Histogram) Of Hospital-Level HF RSRRs



Out of 4,642 hospitals in measure cohort 110 performed "better than the U.S. national rate," 3,454 performed "no different from the U.S. national rate," and 149 performed "worse than the U.S. national rate." 929 were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing.

The median odds ratio was 1.15.

2b4.3. 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 median odds ratio suggests a meaningful increase in the risk of readmission if a patient is admitted with HF at a higher risk hospital compared to a lower risk hospital. A value of 1.15 indicates that a patient has a 15% higher odds of a readmission at higher risk performance hospital compared to a lower risk hospital, indicating the impact of quality on the outcome rate is substantial.

The variation in rates and number of performance outliers suggests there remain differences in the quality of care received across hospitals for HF. This evidence supports continued measurement to reduce the variation.

2b5. 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 social risk factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or 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 social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with***

more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.

2b5.1. 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

2b5.2. 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

2b5.3. 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

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. 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 non responders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

The HF readmission measure used claims-based data for development and testing. There was no missing data in the development and testing data.

2b6.2. 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

2b6.3. 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*) Update this field for ***maintenance of endorsement***.

ALL data elements are in defined fields in electronic claims

3b.2. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources. For ***maintenance of endorsement***, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

N/A

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

Attachment:

3c. Data Collection Strategy

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

3c.1. *Required for maintenance of endorsement.* Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

***IF instrument-based,* consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.**

This measure uses administrative claims and enrollment data and as such, offers no data collection burden to hospitals or providers.

3c.2. Describe any fees, licensing, or other requirements to use any aspect of the measure as specified *(e.g., value/code set, risk model, programming code, algorithm)*.

N/A

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current *and* Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
*	Public Reporting Hospital Compare https://www.medicare.gov/hospitalcompare/search.html Hospital Compare https://www.medicare.gov/hospitalcompare/search.html Payment Program Hospital Readmission Reduction (HRRP) Program https://www.qualitynet.org/inpatient/hrrp Hospital Readmission Reduction (HRRP) Program https://www.qualitynet.org/inpatient/hrrp

*cell intentionally left blank

4a1.1 For each CURRENT use, checked above (update for *maintenance of endorsement*), provide:

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

Public Reporting
Program Name, Sponsor: Hospital Compare, Centers for Medicare and Medicaid Services (CMS)
Purpose: Under Hospital Compare and other CMS public reporting websites, CMS collects quality data from hospitals, with the goal of driving quality improvement through measurement and transparency by publicly displaying data to help consumers make more informed decisions about their health care. It is also intended to encourage hospitals and clinicians to improve the quality and cost of inpatient care provided to all patients. The data collected are available to consumers and providers on the Hospital Compare website at: https://www.medicare.gov/hospitalcompare/search.html. Data for selected measures are also used for paying a portion of hospitals based on the quality and efficiency of care, including the Hospital Value-Based Purchasing Program, Hospital-Acquired Condition Reduction Program, and Hospital Readmissions Reduction Program.
Payment Program
Program Name, Sponsor: Hospital Readmission Reduction Program (HRRP), 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 non-IPPS cancer hospitals. Critical access hospitals, non-IPPS cancer hospitals, and hospitals located in U.S territories are not 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.

4a1.2. 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.

4a1.3. 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.

4a2.1.1. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.

The exact number of measured entities (acute care hospitals) varies with each new measurement period. For the period between 2016 – 2019, all non-federal short-term acute care hospitals (including Indian Health Service hospitals), critical access hospitals, and VA hospitals (4,642 hospitals) were included in the measure calculation. Only those hospitals with at least 25 HF admissions were included in public reporting.

Each hospital receives their measure results in April/May of each calendar year through CMS's QualityNet website. The results are then publicly reported on CMS's Hospital Compare website in July of each calendar year. Since the measure is risk standardized using data from all hospitals, hospitals cannot independently calculate their score.

However, CMS provides each hospital with several resources that aid in the interpretation of their results (described in detail below). These include Hospital-Specific Reports with details about every patient from their facility that was included in the measure calculation (for example, dates of admission and discharge, discharge diagnoses, outcome [died or not], transfer status, and facility transferred from). These reports facilitate quality improvement activities such as review of individual deaths and patterns of deaths; make visible to hospitals post-discharge outcomes that they may otherwise be unaware of; and allow hospitals to look for patterns that may inform quality improvement (QI) work (e.g. among patient transferred in from particular facilities). CMS also provides measure FAQs, webinars, and measure-specific question and answer inboxes for stakeholders to ask specific questions.

The Hospital-Specific Reports also provide hospitals with more detailed benchmarks with which to gauge their performance relative to peer hospitals and interpret their results, including comorbidity frequencies for their patients relative to other hospitals in their state and the country.

Additionally, the code used to process the claims data and calculate measure results is written in SAS (Cary, NC) and is provided each year to hospitals upon request.

4a2.1.2. Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

During the Spring of each year, hospitals have access to the following list of updated resources related to the measure which is provided directly or posted publicly for hospitals to use:

1. Hospital-Specific Reports (HSR): available for hospitals to download from QualityNet in April/May of each calendar year; includes information on the index admissions included in the measure calculation for each facility, detailed measure results, and state and national results.
2. HSR User Guide: available with the HSR and posted on QualityNet; provides instructions for interpreting the results and descriptions of each data field in the HSR.
3. Mock HSR: posted on QualityNet; provides real national results and simulated state and hospital results for stakeholders who do not receive an HSR.
4. HSR Tutorial Video: A brief animated video to help hospitals navigate their HSR and interpret the information provided.

5. Public Reporting Preview and Preview Help Guide: available for hospitals to view from QualityNet in Spring of each calendar year; includes measure results that will be publicly reported on CMS's public reporting websites.
6. Annual Updates and Specification Reports: posted in April/May of each calendar year on QualityNet with detailed measure specifications, descriptions of changes made to the measure specifications with rationale and impact analysis (when appropriate), updated risk variable frequencies and coefficients for the national cohort and updated national results for the new measurement period.
7. Frequently asked Questions (FAQs): includes general and measure-specific questions and responses, as well as infographics that explain complex components of the measure's methodology and are posted in April/May of each calendar year on QualityNet.
8. The SAS code used to calculate the measure with documentation describing what data files are used and how the SAS code works. This code and documentation are updated each year and are released upon request beginning in July of each year.
9. Measure Fact Sheets: provides a brief overview of measures, measure updates, and are posted in April/May of each calendar year on QualityNet.

During the summer of each year, the publicly reported measure results are posted on CMS's public reporting websites, a tool to find hospitals and compare their quality of care that CMS created in collaboration with organizations representing consumers, hospitals, doctors, employers, accrediting organizations, and other federal agencies. Measure results are updated in July of each calendar year.

4a2.2.1. Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.

Describe how feedback was obtained.

Questions and Answers (Q&A)

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

Literature Reviews

In addition, we routinely scan the literature for scholarly articles describing research related to this measure. We summarize new information obtained through these reviews every three years as a part of comprehensive reevaluation as mandated by the Measure Management System (MMS) Blueprint.

4a2.2.2. Summarize the feedback obtained from those being measured.

Summary of Questions or Comments from Hospitals submitted through the Q&A process:

For the HF readmission measure, we have received the following inquiries from hospitals since the completion of previous endorsement maintenance:

1. Requests for detailed measure specifications including CC-to-ICD-9 code crosswalks, and ICD-10 codes used to define the measure cohort or in the risk-adjustment model;
2. Requests for the SAS code and SAS packs used to calculate measure results;
3. Request for the national 30-day heart failure readmission rate;
4. Queries about how cohorts and outcomes are defined including whether there is overlap in the cohorts and outcomes assessed in the AMI, coronary artery bypass grafting, and stroke mortality measures;
5. Questions about how transfers are handled in the measure calculation;

6. Queries about how to calculate the measure and to interpret the statistical model including the interpretation of coefficients for risk variables;
7. Request for the Measure Calculation Package included in the Hospital Value-Based Program; and;
8. Requests for hospital-specific measure information, such as data included in the HSRs.

4a2.2.3. Summarize the feedback obtained from other users

Summary of Question and Comments from Other Stakeholders:

For the HF Readmission measure, feedback received from other stakeholders since the completion of previous endorsement maintenance:

1. Requests for detailed measure specifications including the narrative specifications for the measure, CC-to-ICD-9 code crosswalks, and ICD-10 codes used to define the measure cohort or in the risk-adjustment model;
2. Requests for the SAS code and SAS packs used to calculate measure results;
3. Request for the Measure Calculation Package included in the Hospital Value-Based Program;
4. Requests for clarification of how inclusion and exclusion criteria are applied;
5. Queries about how cohorts and outcomes are defined;
6. Questions about how transfers are handled in the measure calculation;
7. Queries about the implementation of the measure in CMS initiatives and payment programs; and
8. Requests about hospital-specific measure information including mock Hospital-Specific Reports.

Summary of Relevant Publications from the Literature Review:

Since the last endorsement cycle, we have reviewed more than 400 articles related to readmissions following HF admissions. Relevant articles shared key themes related to: spillover effects of the HF readmission measure on readmission rates for other conditions; considerations for additional risk adjustment variables, including social risk factors and other clinical comorbidities; potential unintended consequences of readmission measures on mortality outcomes; impact of not including Medicare Advantage patients in readmission measures; effectiveness of transitional care models on reducing readmissions; clinical differences between patients with reduced (HFrEF) versus preserved ejection fraction (HFpEF) and, the impact of potential strategies to avoid readmissions within the 30-day timeframe.

Researchers have conducted considerable investigation of potential unintended consequences since the implementation of the HF readmission measure. More specifically, the relationship between the implementation of the AMI, HF, and PN readmission measures in the Hospital Readmissions Reduction Program (HRRP) and subsequent trends in their respective mortality rates has been studied.

Some studies have argued that between 2006–2014, readmissions for HF decreased but post-discharge mortality increased, suggesting a potential unintended consequence that readmission measures may be incentivizing hospitals to not readily readmit patients with HF, and as a result, mortality rates increased (Khera et al., 2018; Wadhera et al. 2018; Meyer et al., 2018). However, the same studies have acknowledged that HF mortality was increasing prior to HRRP implementation and that factors unrelated to HRRP could have caused this trend — for example, the increasing use of do not resuscitate orders (DNRs) could lead to an increase in mortality rates. These findings suggest that the increase in mortality (which, again, preceded HRRP) is not a result of denying admission to people seeking acute care services. Of note, other studies have found no apparent increase in HF mortality (Dharmarajan et al., 2017; MedPAC, 2018; Stensland., 2019).

Given the importance of this potential issue on patient outcomes, CMS commissioned an independent group to investigate whether there have been increases in mortality rates after HRRP implementation. CMS found through this investigation that no sufficient evidence exists to suggest that mortality has increased because of the HRRP readmission measures. CMS is committed to continuing to monitor trends in same-condition readmission and mortality rates through annual measure reevaluation and surveillance tasks.

References:

Dharmarajan K, Wang Y, Lin Z, et al. Association of Changing Hospital Readmission Rates With Mortality Rates After Hospital Discharge. JAMA. 2017;318(3):270-278.

Khera R, Dharmarajan K, Wang Y, et al. Association of the Hospital Readmissions Reduction Program With Mortality During and After Hospitalization for Acute Myocardial Infarction, Heart Failure, and Pneumonia. JAMA Netw Open. 2018;1(5):e182777.

Medicare Payment Advisory Commission. Mandated report: The effects of the Hospital Readmissions Reduction Program. Washington, DC 07/18 2018.

Meyer N, Harhay MO, Small DS, et al. Temporal Trends in Incidence, Sepsis-Related Mortality, and Hospital-Based Acute Care After Sepsis. Crit Care Med. 2018;46(3):354-360.

Stensland J. MedPAC evaluation of Medicare's Hospital Readmission Reduction Program: Update. In:2019.

Wadhera RK, Joynt Maddox KE, Wasfy JH, Haneuse S, Shen C, Yeh RW. Association of the Hospital Readmissions Reduction Program With Mortality Among Medicare Beneficiaries Hospitalized for Heart Failure, Acute Myocardial Infarction, and Pneumonia. JAMA. 2018;320(24):2542-2552.

4a2.3. Describe how the feedback described in 4a2.2.1 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

Each year issues raised through the Q&A process or in the literature related to this measure are considered by measure and clinical experts. Any issues that warrant additional analytic work due to potential changes in the measure specifications are addressed as a part of annual measure reevaluation. If small changes are indicated after additional analytic work is complete, those changes are usually incorporated into the measure in the next measurement period. If the changes are substantial CMS may propose the changes through rulemaking and adopt the changes only after CMS received public comment on the changes and finalizes those changes in the IPPS or other rule. There were no questions or issues raised by stakeholders requiring additional analysis or changes to the measures since the last annual form submission in 2018. There have been no changes made to the measure since Maintenance of endorsement in 2016.

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.

4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)

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

The median HF 30-day, all-cause, RSRR for the HF readmission measure for the 3-year period between July 1, 2016 and June 30, 2019 was 21.9%. The median RSRR increased by 0.1 absolute percentage points from July 2016-June 2017 (median RSRR: 21.8%) to July 2018-June 2019 (median: RSRR: 21.9%).

4b2. 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).

4b2.1. Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.

N/A

4b2.2. Please explain any unexpected benefits from implementation of this measure.

Health services researchers have also explored potential spillover effects of the HF readmission measure's implementation and reductions in readmissions for non-targeted conditions. Several studies support positive spillover effects, as there has been systematic improvement in risk-standardized readmission rates for patients not included in HRRP measures (Carey et al., 2015; Angraal et al., 2018; Demiralp et al., 2018; Sukul et al., 2017; Myers et al., 2020).

References:

Angraal S, Khera R, Zhou S, et al. Trends in 30-Day Readmission Rates for Medicare and Non-Medicare Patients in the Era of the Affordable Care Act. Am J Med. 2018;131(11):1324-1331 e1314.

Carey K, Lin MY. Readmissions To New York Hospitals Fell For Three Target Conditions From 2008 To 2012, Consistent With Medicare Goals. Health Aff (Millwood). 2015;34(6):978-985.

Demiralp B, He F, Koenig L. Further Evidence on the System-Wide Effects of the Hospital Readmissions Reduction Program. Health Serv Res. 2018;53(3):1478-1497.

Myers LC, Faridi MK, Hasegawa K, Hanania NA, Camargo CA Jr. The Hospital Readmissions Reduction Program and Readmissions for Chronic Obstructive Pulmonary Disease, 2006-2015. Ann Am Thorac Soc. 2020;17(4):450-456. doi:10.1513/AnnalsATS.201909-672OC.

Sukul D, Sinha SS, Ryan AM, Sjoding MW, Hummel SL, Nallamothu BK. Patterns of Readmissions for Three Common Conditions Among Younger US Adults. Am J Med. 2017;130(10):1220 e1221-1220 e1216.

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

0505: Hospital 30-day all-cause risk-standardized readmission rate (RSRR) following acute myocardial infarction (AMI) hospitalization.

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

2879: Hybrid Hospital-Wide Readmission (HWR) Measure with Claims and Electronic Health Record Data

2880: Excess days in acute care (EDAC) after hospitalization for heart failure (HF)

2886: Risk-Standardized Acute Admission Rates for Patients with Heart Failure

2888: Accountable Care Organization Risk-Standardized Acute Hospital Admission Rate for Patients with Multiple Chronic Conditions

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

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified.

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):

Are the measure specifications harmonized to the extent possible?

Yes

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.

Available at measure-specific web page URL identified in S.1 **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

Co.2 Point of Contact: Helen, Dollar-Maples, Helen.Dollar-Maples@cms.hhs.gov, 410-786-7214-

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: Doris, Peter, Doris.peter@yale.edu

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.

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, 2018

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? 2020

Ad.6 Copyright statement: N/A

Ad.7 Disclaimers: N/A

Ad.8 Additional Information/Comments: N/A