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 #: 2880

Corresponding Measures:

De.2. Measure Title: Excess days in acute care (EDAC) after hospitalization for heart failure (HF)

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

De.3. Brief Description of Measure: The measure assesses days spent in acute care within 30 days of discharge from an inpatient hospitalization for HF to provide a patient-centered assessment of the post-discharge period. This measure is intended to capture the quality-of-care transitions provided to discharged patients who had a HF hospitalization by collectively measuring a set of adverse acute care outcomes that can occur post-discharge: emergency department (ED) visits, observation stays, and unplanned readmissions at any time during the 30 days post-discharge. In order to aggregate all three events, we measure each in terms of days. The Centers for Medicare & Medicaid Services (CMS) annually reports the measure for patients who are 65 years or older, are enrolled in Medicare Fee-For-Service (FFS) and are hospitalized in non-federal short-term acute care hospitals.

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. Safely transitioning patients from hospital to home requires a complex series of tasks which would be cumbersome to capture individually as process measures: timely and effective communication between providers, prevention of and response to complications, patient education about post-discharge care and self-management, timely follow-up, and more. Suboptimal transitions contribute to a variety of adverse events post-discharge, including ED evaluation, need for observation, and readmission.

Measures of unplanned readmission already exist, but there are no current NQF-endorsed measures for ED and observation stay utilization for this condition. It is thus difficult for providers and consumers to gain a complete picture of post-discharge outcomes. Moreover, separately reporting each of these outcomes encourages "gaming," such as re-categorizing readmission stays as observation stays to avoid a readmission outcome. By capturing a range of acute care events that are important to patients, we can produce a more complete picture of post-discharge outcomes that better informs consumers about care quality and incentivizes global improvement in transitional care.

S.4. Numerator Statement: The outcome for this measure is a count of the number of days the patient spends in acute care within 30 days of discharge from an eligible index admission for HF. We define days in acute care as days spent in an ED, admitted to an observation unit, or admitted as an unplanned readmission for any cause to a short-term acute care hospital, within 30 days from the date of discharge from the index HF hospitalization.

Additional details are provided in S.5 Numerator Details.

S.6. Denominator Statement: The target population for this measure is Medicare FFS beneficiaries aged 65 years and older hospitalized at non-Federal and VA acute care hospitals for HF.

The cohort includes admissions for patients discharged from the hospital with a principal diagnosis of HF (codes in the attached Data Dictionary) and with continuous 12 months Medicare enrollment prior to admission. CMS publicly reports this measure 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 measure excludes index hospitalizations that meet any of the following exclusion criteria:

1. Without at least 30 days of post-discharge enrollment in Medicare FFS
2. Discharged against medical advice
3. HF admissions within 30 days of discharge from a prior HF index admission
4. 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.

De.1. Measure Type: Outcome

S.17. Data Source: Claims, Other

S.20. Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Dec 09, 2016 **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 not formally paired with any measure; however, it is harmonized with a measure of hospital-level, all-cause, 30-day, risk-standardized readmission following heart failure hospitalization.

Preliminary Analysis: Maintenance 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 2015

- This is intended to be a measure of days spent in acute care within 30 days of discharge from an inpatient hospitalization for HF. Specifically, it is an outcome measure that is intended to capture the quality of care transitions provided to discharged patients who had a HF hospitalization by collectively measuring a set of adverse acute care outcomes that can occur post-discharge: emergency department (ED) visits, observation stays, and unplanned readmissions at any time during the 30 days post-discharge.
- The developer previously cited the incidence rate of heart failure (HF) approaches 10 per 1,000 in patients 65 years and older and continues to be one of the most common discharge diagnoses among the elderly; the prevalence of HF in the U.S. was estimated to be more than six million cases and is suspected to be the leading cause of death in people over age 65.
- Developer cited several studies supporting various care processes can influence post-discharge acute care utilization after a hospitalization for heart failure. Further the developer provided evidence suggesting that hospitals and health plans have been able to reduce readmission rates through more generalizable quality improvement initiatives, such as communication between providers, patient education, patient safety, and coordinated transitions to the outpatient environment contribute to patient outcomes.

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 referenced data indicating that the lifetime risk of HF is estimated at one-in-five at 40 years of age, and the prevalence in the aging U.S. population is expected to increase significantly in the next few decades.
- The developer cited a number of hospital/facility-level strategies have been shown to reduce re-hospitalization in the case of heart failure. Several of those strategies include:
    - Better care coordination at discharge reduced the likelihood of a readmission: discharge summaries that were transmitted to any outpatient clinician were associated with lower odds of readmission, and discharge summaries that included elements related to transitions of care were also associated with lower odds of readmission.

- In addition, a meta-analysis found that interventions such as patient education and patient education combined with other interventions were the most beneficial and interventions that included one or more interventions were 1.4 to 6.8 times less likely to be readmitted.

Questions for the Committee:

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

Guidance from the Evidence Algorithm

Box 1: The measure assesses a healthcare outcome → Box 2: The developer has provided empirical data that there is a relationship between the measured outcome and at least one healthcare outcome → 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. Data indicates variable performance and room for performance improvement.

Current review in 2021

Distribution of HF EDAC across the three most recent **three-year reporting periods** across hospitals with **at least 25 admissions**

- (July 1, 2016-June 30, 2019) | (July 1, 2015-June 30, 2018) | (July 1, 2014-June 30, 2017). Range of performance **most recent reporting period (2016-2019)**: -59.7 to 154.4 EDAC per 100 admissions and median EDAC 2.3 EDAC per 100 admissions.
- Data across all three reporting periods: Year 2016 – 2019 | Year 2015 – 2018 | Year 2014 – 2017
  - Number of Hospitals: 3713 | 3643 | 3690
  - Number of Admissions: 1,275,344 | 1,188,842 | 1,159,275
  - Mean (SD): 4.3 (24.9) | 4.3 (24.8) | 4.5 (25.2)
  - Range (Min to Max): -59.7 to 154.4 | -66.6 to 143.2 | -65 to 147.8
  - Minimum: -59.7 | -66.6 | -65.0
  - 10th percentile: -25.4 | -25.7 | -25.2
  - 50th percentile: 2.3 | 2.4 | 2.4
  - 70th percentile: 14.8 | 15.2 | 14.6
  - 90th percentile: 36.3 | 36.1 | 37.5
  - Maximum: 154.4 | 143.2 | 147.8

Distribution of HF EDAC across the three most recent three-year reporting periods across **all hospitals**:

Range of performance **most recent reporting period (2016-2019)**: -70.1 to 259 EDAC per 100 admissions and median EDAC 0.3 EDAC per 100 admissions.

- Data across all three reporting periods: Year 2016 – 2019 | Year 2015 – 2018 | Year 2014 – 2017
  - Number of Hospitals: 4642 | 4534 | 4577
  - Number of Admissions: 1286352 | 1199343 | 1169795
  - Mean (SD): 3.3 (25.5) | 3.3 (25.3) | 3.4 (25.1)

- Range (Min to Max): -70.1 to 259 | -66.6 to 340.3 | -65 to 147.8
- Minimum: -70.1 | -66.6 | -65.0
- 10th percentile: -25.6 | -25.5 | -25.5
- 50th percentile: -0.3 | 0.1 | 0.0
- 70th percentile: 13.0 | 13.4 | 12.9
- 90th percentile: 35.6 | 35.4 | 36.6
- Maximum: 259.0 | 340.3 | 147.8

Additionally, developer cites studies that indicate that the care for these patients is highly variable, and gaps exist in the quality of hospital care, particularly in the transition to outpatient care.

Disparities

Distribution of 30-day HF EDAC by Proportion of Dual Eligible Patients (July 2016 through June 2019) across hospitals (**with at least 25 cases**) by proportion of patients with social risk (as dual-eligible status and Agency for Healthcare Research and Quality [AHRQ] Socioeconomic Status [SES] Index).

- Quartile: hospitals in the first quartile for the proportion patients with of dual-eligible status | hospitals in the fourth (highest) quartile for the proportion patients with of dual-eligible status
- Social Risk Proportion (%): q1: (0-10.2%) | q4: (24.5-100%)
- Number of Hospitals: 928 | 928
  - Maximum: 148.8 | 154.4
  - 90th percentile: 31.0 | 45.5
  - 75th percentile: 13.8 | 28.4
  - Median: -0.7 | 6.6
  - 25th percentile: -14.6 | -11.2
  - 10th percentile: -26.9 | -24.1
  - Minimum: -54.0 | -59.7

Distribution of 30-day HF EDAC by Proportion of Patients with AHRQ SES Index Scores (July 2016 through June 2019) across hospitals (with at least 25 cases) by the facilities' proportion of patients in lower and upper social risk quartiles

- Social Risk Proportion (%): q1:(0-8.5%) | q4: (35.1-100%)
- # Of Hospitals: 921 | 921
  - Maximum: 114.3 | 130.4
  - 90th percentile: 24.9 | 42.6
  - 75th percentile: 11.8 | 24.4
  - Median: -3.6 | 6.6
  - 25th percentile: -17.1 | -9.6
  - 10th percentile: -29.5 | -22.5
  - Minimum: -56.1 | -59.7

Questions for the Committee:

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

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

Committee Pre-evaluation Comments:

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

1a. Evidence to Support Measure Focus: 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."

- Responding to the question posed by NQF staff, incorporating data from their logic model, if feasible, could change results.
- Readmissions/ED visits/Observation days are negative outcomes to care.
- No concerns
- Developer cited data estimating the lifetime risk of HF is one-in-five at age 40. Also cited articles that showed lower readmissions from better care coordination strategies and patient education. I am not aware of any new studies that would impact this measure.
- evidence relates directly outcome being measured

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?

- A measure specifically built to look at the performance of underserved populations might be revealing.
- As measured, from below expected to above expected, there is a wide variation in performance
- No concerns
- Yes. Developer presented data of the distribution of HF EDAC for the three most recent three-year reporting period across hospital with at least 25 admissions and across all hospitals. They also cited studies showing variability in quality care especially during the transition to outpatient care. Disparities: Dual Eligible and AHRQ SES score used to show variability in care.
- current performance data on the measure was provided; gap demonstrated and disparities identified

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

2c. For composite measures: empirical analysis support composite approach

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.

Composite measures only:

2d. Empirical analysis to support composite construction. Empirical analysis should demonstrate that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct.

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

Evaluators: NQF Scientific Methods Panel Subgroup 1

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: H-0; M-8; L-1; I-0 (Pass)
- Validity: H-0; M-7; L-0; I-1 (Pass)

Reliability

- The SMP passed the measure on reliability with Moderate rating (H-0; M-8; L-1; I-0).
- The developer conducted reliability testing at the measure score-level. The developer estimated measure reliability using intra-class correlation coefficient (ICC). This statistic can be used to assess the correlation and agreement between measurements.
- The developer used a split-sample 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.
- Using the split-sample approach, the developer reported ICC ranges from 0.456 for hospitals with at least two admissions to 0.698 for hospitals with at least 300 admissions. For hospitals with at least 25 admissions, the ICC was 0.527.
- The SMP did not raise any major concerns with reliability and passed the measure on this criterion.

Validity

- The SMP passed the measure on validity with a Moderate rating (H-0; M-7; L-0; I-1).
- The developer conducted face validity and empirical validity testing at the measure score level.
- Face validity was assessed in consultation with national guidelines for publicly report outcomes measures and using external stakeholder focus workgroup study and survey-based information provided by the 16-member technical expert panel.
    - The developer reports that of the 16 TEP members convened, 12 provided survey responses.
    - 11 of 12 (91.7%) TEP members convened by the developer strongly, moderately, or somewhat agreed with the statement: "The risk-standardized acute care days obtained from the measures as specified can be used to distinguish between better and worse quality hospitals."
- Construct validity was assessed as the relationships between the HF EDAC measure score and the risk standardized readmission rate (RSRR) group scores, the overall hospital rating scores, and the HF readmission measure.
- The developer posited a negative relationship between the HF EDAC scores, star-rating readmission score group and star-rating summary scores and hypothesized that HF EDAC and HF readmission measure scores would show a strong, positive correlation.
    - Correlation with Hospital Star Rating readmission group score: -0.418 ($p<.0001$),
    - Correlation with Overall Hospital Star Rating summary score: -0.371 ($p<.0001$),
    - Correlation with HF Readmission Measure: 0.574 ($p<.0001$). The data suggests a positive relationship between the HF EDAC scores and the HF readmission measure scores; the lower HF EDAC scores are more likely to have lower HF RSRRs.
- The SMP expressed concern with the choice of variables for the construct validity, noting the potential for endogeneity due to the overlapping readmission events between the EDAC measure and the Star Ratings measures, as the same readmission events are included in both measures.
    - The developer provided updated testing results by removing the comparator measure from the Star Rating Readmission Group score before analyzing the correlation and by removing the entire Readmission Group score.
    - The developer noted that a moderate correlation remains (r = -0.349 versus - 0.399) in the expected direction between the EDAC measure and Star Ratings, even after removing the overlapping measure. The developer also found a moderate association (r = -0.457 versus -0.579), albeit weaker, between the EDAC measure and Star Ratings after removing the entire Readmission Group from Star Ratings.

Risk Adjustment

- The developer used a statistical risk model with 37 risk factors.
- The developers derived a parsimonious risk adjustment model by using logistic regression with a stepwise backward elimination process using repeated in 1,000 bootstrap samples from the entire population via random selection with replacement.
- They retained candidate variables demonstrating a positive association with readmission at p-value <0.01 in each of the 1,000 repeated samples.
- Two social risk factors were tested and found to be statistically significant (i.e., Dual-eligible status and AHRQ SES index). The developer also performed a decomposition analysis. In this analysis, the clinical risk factors have a larger patient-level effect compared to their hospital-level effects. In contrast, both the low AHRQ SES variable and the dual eligible variable have a larger hospital-level effect compared to the patient-level effect. Based on these analyses, the developer did not adjust this measure for either dual eligibility or the AHRQ SES Index

- The c-statistic for risk-adjustment model is 0.59 and the R-squared value is 0.027. The developer indicates that a c-statistic of 0.59 demonstrates fair model discrimination, suggesting a moderate ability to distinguish patients with high risk from low risk of having at least one excess day in acute care.
- Concerning the risk adjustment model, the SMP debated whether updates to the risk adjustment model should include creating both training and validation datasets, as some SMP members argued that this is standard in evaluating model performance for risk adjustment models. The SMP noted that this occurred during the initial development of the risk adjustment model for NQF #2880, but it was not done for the updates to the current model.
- A SMP member also argued since the model now uses ICD-10 codes instead of ICD-9 codes, it should be treated as a new model. The developer confirmed that for ongoing model performance, they do not have a development (training) and validation set, as they are not reselecting risk variables every year, they simply recalculate the beta coefficients of the same risk variables.
- Some SMP members were not concerned by the lack of a validation dataset for the updated model since the risk variables have not changed. The SMP agreed that there is not consensus on this issue in the evaluation guidance and that it should be discussed at a future SMP advisory meeting.

Meaningful differences

- The developer provided a distribution of performance for 4,642 hospitals in the measure cohort. 447 facilities EDAC were fewer days than average; 2,467 were reported as average; 799 had EDAC more days than average. 929 facilities were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing.

Questions for the Committee regarding reliability:

- *Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?*
- *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.)?*
- *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 concerns.
- Claims based measure. Readmission, ED and observation status clearly defined.
- No concerns
- Split sample reliability (test-retest) used. >=25 admissions (80% of hospitals) ICC=0.527. >=100 admissions (53.7% of hospitals) ICC=0.632. No concerns.
- no concerns

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

- It concerns me that not more of the SMP rated the reliability high.
- No concerns
- No. SMP rated Moderate reliability
- no concerns

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

- The comparator measures are of concern.
- Issue of correlation with measures of readmission raised and addressed through rerunning correlations without including readmissions measures or components. Effects modest but in right direction.
- No concerns
- No.
- no concerns

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?

- Concerned that social risk factors end up being dropped from the model.
- Risk adjustment model is 2 part model. C-stat on first part (any days) is moderate. Prediction of days in second part, conditional on any days, is poor. Not clear that measured variation in days, rather than readmission is enhanced by this measure.
- C-stat on any use moderate. Ability of Poisson regression to estimate days given any days is low. Not clear that this is a more reliable measure than the readmission measure.
- No concerns
- No risk adjustment included by developer. Dual eligible status and AHRQ SES index variables had a large hospital effect compared to patient level effect
- no concerns

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.

- All data elements are in defined fields in a combination of electronic sources.
- This measure uses administrative claims and enrollment data and as such, offers no data collection burden to hospitals or providers.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

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?

- Perhaps other social risk factors which are not administrative data should be included. All administrative data measure makes the measure feasible.
- claims based measure. no issues.
- No concerns
- No concerns. No data collection burden to hospitals are providers since electronic sources using administrative claims and enrollment data
- no concerns

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 evaluates 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: Care Compare, Centers for Medicare and Medicaid Services (CMS)
- Payment Program: CMS Hospital Inpatient Quality Reporting Program (IQR)

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 notes that each hospital generally 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 public reporting websites in the summer of each calendar year.
- The developer adds that since the measure is risk-standardized using data from all hospitals, hospitals cannot independently calculate their score. However, CMS Hospital-Specific Reports with details about every patient from their facility that was included in the measure calculation.
- Lastly, the developer states that Hospital-Specific Reports (HSRs) also provide hospitals with more detailed benchmarks with which to gauge their performance relative to peer hospitals and interpret their results, including comorbidity frequencies for their patients relative to other hospitals in their state and the country.
- Hospitals have access to other resources that are updated in the Spring of each year and are publicly reported. These include HSR guides, tutorial videos, FAQs, annual updates and specification reports, SAS code, measure fact sheets, and other resources.
- Accountable entities and other stakeholders can submit questions via an online portal, in which experts on measure specifications and/or implementation respond to those inquiries.
- Additionally, the developer routinely scans the literature for articles describing research related to this measure.
- The developer has received feedback and inquiries related to the overlap of EDAC and other readmissions measures, interpretation of measure results and performance categories, specifications, performance period, etc.
- Every year, the feedback and literature are considered by technical and clinical experts. The developer states that any issues that warrant additional analytic work due to potential changes in the measure specifications are addressed as a part of annual measure reevaluation.
- Based on feedback received, the developer states that they "revised the methodology used to count the number of observation-stay days in the EDAC outcome. The use of both physician and facility claims (and use of the claim with the longer duration when both claims are present) was changed to use of physician claims only in cases when a facility claim is not available. This change, however, had minimal impact on measure results."

***Questions for the Committee**:*

- *How have (or can) the performance results be used to further the goal of high-quality, efficient healthcare?*
- *How has the measure been vetted in real-world settings by those being measured or others?*

Preliminary rating for Use: ☒ Pass ☐ No Pass

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

4b. Usability evaluates 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

To compare performance on the HF EDAC measures across performance periods, the developer provided the distribution of measure scores for hospitals with at least 25 admissions for Medicare FFS admission only. Note, the developer removed VA admissions as they only became part of the cohort during the most recent

reporting period (2016-2019). The developer reports improvement over the past three reporting periods (from right to left, 2014-2017, 2015-2018, and 2016-2019) in measure scores across most of the distribution, from the 30th percentile through the 80th percentile.

- Periods: YEAR 2016-2019 | YEAR 2015-2018 | YEAR 2014-2017
- Number of Hospitals: | 3586 | 3643 | 3690
- Number of Admissions: 1219779 | 1188842 | 1159275
- Mean (SD): 4.2(24.8) | 4.3(24.8) | 4.5(25.2)
- Range (Min to Max): -59.7 to 154.4 | -66.6 to 143.2 | -65 to 147.8
- Minimum: -59.7 | -66.6 | -65.0
- 10th percentile: -25.4 | -25.7 | -25.2
- 20th percentile: -16.6 | -17.2 | -16.5
- 30th percentile: -10.1 | -9.8 | -9.8
- 40th percentile: -3.6 | -3.4 | -3.6
- 50th percentile: 2.3 | 2.4 | 2.4
- 60th percentile: 8.3 | 8.4 | 8.4
- 70th percentile: 14.8 | 15.2 | 14.6
- 80th percentile: 23.9 | 24.1 | 23.9
- 90th percentile: 36.1 | 36.1 | 37.5

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

- No unexpected findings were identified by the developer.

Potential harms

- The developer states that they did not identify any unintended consequences during measure development or model testing, but they are committed to monitoring this measure's use and assessing potential unintended consequences over time, such as the inappropriate shifting of care, increased patient morbidity and mortality, and other negative unintended consequences for patients.

***Questions for the Committee**:*

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

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

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

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?

- Yes to above. Some hospitals have used their data to improve care transitions. Feedback from hospitals is sought.
- Feedback mechanism described in documentation. Usable by hospital, perhaps (would like some direct commentary from hospitals on how they use the reports). Potential patient use of measure is minimal.
- No concerns
- This measure is publicly reported on Care Compare, CMS. Payment program: CMS Hospital inpatient quality reporting Program. Hospitals receive measure results in April/May each year through CMS QualityNet website and the results are reported on CMS websites in the summer. The developer received feedback related to overlap of EDAC and other readmissions measures. In response to feedback, the developer revised the methodology used to count the number of observation-stay days in the EDAC outcome. This change had minimal impact on measure results
- performance results disclosed and available and feedback was provided

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.

- The measure has shown impact. No unintended consequences reported.
- No obvious harms.
- No concerns
- Improvement in measure scores over the past three reporting periods. No unintended negative consequences identified.
- measure developer describes how the performance results could be used to further the goal of high-quality, efficient healthcare

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
- 0230: Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute myocardial infarction (AMI) hospitalization
- 0330: Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization
- 0505: Hospital 30-day all-cause risk-standardized readmission rate (RSRR) following acute myocardial infarction (AMI) hospitalization.
- 0506: Hospital 30-day, All-Cause, Risk-Standardized Readmission Rate (RSRR) Following Pneumonia Hospitalization
- 1551: Hospital-level 30-day risk-standardized readmission rate (RSRR) following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA)
- 1789: Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)
- 1891: Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization
- 2515: Hospital 30-day, all-cause, unplanned, risk-standardized readmission rate (RSRR) following coronary artery bypass graft (CABG) surgery
- 2881: Excess days in acute care (EDAC) after hospitalization for acute myocardial infarction (AMI)
- 2882: Excess days in acute care (EDAC) after hospitalization for pneumonia

Harmonization

The developer developed the measure in the Medicare Fee-for-Service (FFS) population and completely harmonized the cohort definition and risk-adjustment strategy with those of the existing CMS 30-day HF readmission measure.

The developer also noted the key differences/characteristics between EDAC measures and readmissions measures:

- EDAC measures are based on the count of excess days spent in acute care whereas the readmission measures focus on the dichotomous presence of any readmission within the 30 days past discharge.
- In addition to readmission, the EDAC measure also counts observation stays and ED visits as acute care time.
- This difference in the outcome measure imposes differences on the statistical modeling and reporting format. The interpretations of the measures are also based on relative differences in excess days in acute care based on variations in case mix. There are no differences in data collection burden.

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 specificationns that are not harmonized? Are there any additional steps needed for the measures to be harmonized?

- harmonized with existing CMS 30-day HF readmission measure.
- Correlation with HF readmission measure is about 0.6. Not clear which measure should be preferred. That would require more analysis than provided.
- No concerns
- Harmonized with the existing CMS 30 day HF Readmission measure. Developer noted the difference in focus of these two measures. HF EDAC measures counted excess days spent in acute care and CMS focuses on any readmission within 30 days after discharge.
- no concerns

Combined Methods Panel Scientific Acceptability Evaluation

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 2880

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

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 3: I am confused about the impact of excluding HF admissions within 30 days of a priori HF admission.

   Panel Member 4: No concerns.

   Panel Member 6: No major concerns.

   Panel Member 8: Developers incorporate exposure time and therefore account for the potential of death as a competing variable. Likewise, they exclude patients who are readmitted for transplant as a "planned" readmission. Not excluded are patients readmitted for VAD implantation. This appears to a small but definable defect.

   Panel Member 9: None

RELIABILITY: TESTING

Type of measure:

☐ **Process** ☐ **Process: Appropriate Use** ☐ **Structure** ☐ **Efficiency** ☐ **Cost/Resource Use**

☐ **Outcome** ☒ **Outcome: PRO-PM** ☐ **Outcome: Intermediate Clinical Outcome** ☐ **Composite**

Data Source:

☒ **Claims** ☐ **eMeasure (HQMF) implemented in EHRs** ☐ **Abstracted from Electronic Health Records**
☐ **Abstracted from Paper Medical Records** ☐ **Instrument-Based Data** ☐ **Registry**
☒ **Enrollment Data** ☒ **Other (please specify):**
Panel Member 2: Census Data/American Community Survey, VHA Administrative Data, Medicare Enrollment Data (including Master Beneficiary Summary File
Panel Member 4: Census Data/American Community Survey, VHA Administrative Data, Medicare Enrollment Data (including the Master Beneficiary Summary File)
Panel Member 8: Medicare Enrollment Data (including the Master Beneficiary Summary File), VHA Administrative Data

Level of Analysis:

☐ **Group/Practice** ☐ **Individual Clinician** ☒ **Hospital/facility/agency** ☐ **Health Plan**
☐ **Population: Regional, State, Community, County or City** ☐ **Accountable Care Organization**
☐ **Integrated Delivery System** ☐ **Other (please specify)**

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**

   NOTE: "level of analysis" reflects which entity is being assessed or held accountable by the 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**

6. **Assess the method(s) used for reliability testing**

   Submission document: Testing attachment, section 2a2.2

   Panel Member 1: Split-sample intraclass correlation coefficient

   Panel Member 2: Split half testing with Spearman-Brown adjustment; no concerns.

   Panel Member 3: The developer used split-sample reliability without replacement (creating 2 non-overlapping samples) to assess reliability comparing ICC's for hospitals with varying numbers of admissions. Although still within NQF guidance, it would have been more instructive to have produced splines with standard error bars from their modeling of EDAC days.

   Panel Member 4: The testing approach seems reasonable regarding reliability: split sample reliability.

   Panel Member 5: split-sample reliability testing

   Panel Member 6: No major concerns.

   Panel Member 7: Split-sample reliability, estimated using the ICC (2,1) and then adjusted to the full sample using the Spearman-Brown prophecy formula.

   Panel Member 8: random non overlapping split sample interclass correlation coefficient.

Panel Member 9: The method used was appropriate for testing measure score reliability in this context.

7. **Assess the results of reliability testing**

   Submission document: Testing attachment, section 2a2.3

   Panel Member 1: Modest ICC values with <300 discharges. The real-world challenge is that the minority of hospitals account for the majority of discharges, and in hospitals with at least 300 discharges, ICC = 0.7. That said, as Medicare Advantage grows, the "typical" ICC will continue to deteriorate.

   Panel Member 3: The split sample ICC's range from .465-.698, with roughly 2/3 of the hospitals (those with ≤50 admissions) having ICC values <.60. Only hospitals with ≥300 admissions had an ICC value of ~.70.

   Panel Member 4: The test result is modest for 3 of the 6 scenarios of hospital admission counts. The 3 largest hospital counts (where the smallest count is. >=100 admits) ranged from 0.632 to 0.698. For 2 of the 3 smallest hospital counts (where the range is >=25 to >=50 admits) was 0.527 & 0.573. My summary here removes the smallest hospital count (>=2) as CMS' typical minimum threshold is 25 cases.

   Panel Member 5: Split-sample reliability testing only achieves the threshold value of 0.7 for hospitals with >= 300 admissions (1500 hospitals). For hospitals with >=25 admissions, the ICC is 0.53.

   Panel Member 6: No major concerns.

   Panel Member 7: ICC=0.527 at current volume threshold for reporting, although it would rise to 0.632 at a plausible increase in the threshold from 25 to 100.

   Panel Member 8: 0.465 for hospitals with >/=2 readmissions to 0.68 for hospitals with >/=300 readmissions. However, 0.527 for hospitals with >/= 25 readmissions, which accounted for 99% of hospitals

   Panel Member 9: Reliability is acceptable, if a minimum sample size of at least 25 is maintained. The ICC value at that level is .527 - not great, and maybe not acceptable in the future, but passable at this cycle.

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**

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: Majority of hospitals have low reliability, due to sample size (number of discharges). Problem will only worsen.

Panel Member 3: The ICC's are in the moderate range for split sample reliability.

Panel Member 4: The test result is modest for 3 of the 6 scenarios of hospital admission counts. The 3 largest hospital counts (where the smallest count is. >=100 admits) ranged from 0.632 to 0.698. For 2 of the 3 smallest hospital counts (where the range is >=25 to >=50 admits) ranged from 0.527 & 0.573. My summary here removes the smallest hospital count (>=2) as CMS' typical minimum threshold is 25 cases.

Panel Member 5: The measure reliability of 0.53 is below 0.7 threshold that we are now considering, but above 0.4 that we have accepted in the past.

Panel Member 6: No major concerns.

Panel Member 7: Overall split sample ICC of 0.527 is acceptable by current standards.

Panel Member 8: reasonable results with appropriate methodology for testing

Panel Member 9: ICC of .527 at a minimum sample size of 25 per hospital.

VALIDITY: TESTING

12. **Validity testing level (check all that apply):**

    ☒ **Measure score** ☐ **Data element** ☐ **Both**

13. **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**

    ☐ **No**

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

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

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

    ☒ **Face validity**

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

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

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

    Submission document: Testing attachment, section 2b1.

    ☒ **Yes**

    ☒ **No**

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

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

    Submission document: Testing attachment, section 2b2.2

    Panel Member 1: 1. Face validity assessed by committee of experts 2. Empirical validity assessed by correlation with star ratings and readmission measures

    Panel Member 2: No concerns

    Panel Member 3: Face validity was assessed using survey-based information provided by the 16 member TEP assembled to address this and other measures. Construct validity was assessed as the relationships between the HF EDAC measure score and the risk standardized readmission group scores, the overall

hospital rating scores and the HF readmission measure. The developer posited a negative relationship between the HF EDAC scores, star-rating readmission score group and star-rating summary scores. They also hypothesized a positive relationship between the HF EDAC scores and the HF RSRR scores.

Panel Member 4: The validity testing method was appropriate for measure score testing. Additionally, other testing was conducted that was not specific to measure score, but the testing was adequate for validity testing. Regarding measure score testing, the measure developer examined the relationship between this EDAC measure & the following: -Hospital Star Rating readmission group score -Overall Hospital Star Rating summary score -Heart Failure (HF) Readmission Measure Regarding other testing, the measure developer conducted the following: -lit. review regarding the validity of claims-based measures [p12] -face validity assessed by external groups

Panel Member 6: No major concerns.

Panel Member 7: Construct validation was performed using the Hospital Star Rating readmission group score, the Hospital Star Rating summary score, and the HF Readmission rate. All three are intrinsically correlated measures, because readmissions are the most important driver of EDAC. The validity analysis only shows that the measure is correlated with itself.

Panel Member 8: Face validity: consensus of expert panel--appropriate Empirical testing: comparison with three other metrics, Hospital Star Rating for Readmission, Hospital Star Rating Overall and HF Readmission

Panel Member 9: Face validity is fine, but this is a renewal measure so empirical validity testing is required now. The developer chose to correlate the EDAC measure (which includes readmissions) with other measures that also either include readmissions or are JUST readmissions, in the same clinical population. The same readmissions are apparently being counted on both sides of the correlations. Something has to be done to pull out the automatic, by definition, level of correlation from whatever remains.

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

Submission document: Testing attachment, section 2b2.3

Panel Member 1: >80% of experts moderately or strongly agreed with the validity of the measure. Star ratings were negatively correlated with measure, readmission was relatively strongly positively correlated.

Panel Member 2: No concerns

Panel Member 3: Although the correlations between the HF EDAC measure and validity variables are statistically significant in the expected direction, there is concern for endogeneity of the HF EDAC measures, with the validity variables. That is, there is an apparent inclusion of readmission for HF (not only LOS for HF of readmission) and the validation variables potentially inflated (due to non-independence) empirical results.

Panel Member 4: Regarding measure score testing, the correlation with the EDAC measure and: - CMS readmits group rating was moderate to strong at -0.418 [p18] - CMS overall star rating was moderate at -0.371 [p19] -CMS HF readmits rating was moderate to strong at 0.574 [p18] Regarding the face validity testing, the survey results of the group were in the desired direction, e.g. 83% of the group "strongly agreed" or "moderately agreed" to the statement the EDAC measure "can be used to distinguish between better and worse quality hospitals."

Panel Member 5: "performed empiric validity testing by examining the correlation of this measure with:

- Correlation between HF EDAC Scores and Star Rating Readmission Group Scores -0.418 (p<.0001),
- Correlation between HF EDAC Scores and Overall Hospital Star Rating Scores -0.371 (p<.0001)
- Correlation between HF EDAC Scores and HF Readmission Measure Scores 0.574 (p<.0001),

These results show evidence of empiric validity."

Panel Member 6: No major concerns.

Panel Member 7: Correlations are adequate - for example, r=0.574 between EDAC and readmissions - but the test is very easy to pass given that readmissions drive EDAC. Process-outcome correlations or pre-post analyses of intervention effects are strongly preferred.

Panel Member 8: Empirical testing is not appropriate--correlation with other measures of readmission is expected. This does not provide validity that measure is measuring quality in the care of patients with congestive heart failure. Given the high mortality and morbidity of these patient group, correlation with more clinically meaningful criteria such as mortality over the year following index admission would be much more "valid" assessment of the ability of the measure to capture quality of patient care

Panel Member 9: The reported correlations are not bad, but as noted above, the results are not compelling because some level of correlation among the measures used is not only expected, but automatic.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

Submission document: Testing attachment, section 2b2.

Panel Member 1: Medicare Advantage Why are those <65 excluded? There are bona fide Medicare FFS beneficiaries in this group, too, even if disabled or with ESRD.

Panel Member 4: No concerns with the exclusions.

Panel Member 6: No major concerns.

Panel Member 7: Exclusions are appropriate.

Panel Member 9: None

19. **Risk Adjustment**

Submission Document: "xxxx_measure testing form", section 2b3

19a. **Risk-adjustment method**

☐ None ☒ Statistical model ☐ Stratification

☐ Other method assessing risk factors (please specify)

19b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☐ Yes ☐ No ☒ Not applicable

19c. **Social risk adjustment:**

19c.1 Are social risk factors included in risk model? ☒ Yes ☒ No ☐ Not applicable

19c.2 Conceptual rationale for social risk factors included? ☒ Yes ☒ No

19c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☒ Yes ☒ No

19d. **Risk adjustment summary:**

19d.1 All of the risk-adjustment variables present at the start of care? ☒ Yes ☒ No

19d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☒ Yes ☐ No

19d.3 Is the risk adjustment approach appropriately developed and assessed? ☒ Yes ☐ No

19d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☒ Yes ☒ No

19d.5. Appropriate risk-adjustment strategy included in the measure? ☒ Yes ☒ No

19e**. Assess the risk-adjustment approach**

Panel Member 1: Two-part model (any days, number of days), with adjustments for age, sex, and comorbid conditions

Panel Member 2: The methodological approach was overall excellent. With respect to model calibration, all 3 EDAC measures exhibited a similar pattern of over-estimation of risk in the highest decile and under-estimation of risk in the lowest decile. I am not sure whether the apparent under-/over-estimation is large enough to cause bias or whether there is a simple adjustment that could improve the fit in the extreme deciles.

Panel Member 3: Despite positive evidence that both patient and hospital level SDH and clinical risk factor variables were statistically significant for both logistic and Poisson models, because "median changes in adjusted vs. unadjusted scores were small" and anticipating "unintended consequences of adjustment," the developer opted not to risk adjust empirical results.

Panel Member 4: The risk adjustment methods are appropriate for the given measure. No issues with the measures, but issues with the results, which follow: The c-statistics noted in response to 2b3.6 are 0.587 & 0.59. While there are no hardlines as to a high, moderate & low c-statistic, from everything I've seen at, or below, 0.6 is generally unacceptable. The R squared result of 0.027 is low. The response in 2b3.10 essentially confirms this in the statement: "0.027 indicates that patients' clinical risk factors can explain 2.7% of the variation in the numbers of excess days in acute care" Having said the above, the risk decile plot (fig. 5) show acceptable performance in each decile.

Panel Member 6: No major concerns.

Panel Member 7: The models perform poorly, with c=0.59 from the first stage logistic model and R2=0.027 from the second stage Poisson model. Two tested social risk factors are statistically significant in the Poisson model, but do not appear to meaningfully affect hospital performance estimates.

Panel Member 8: 1) c-statistic for the model is 0.59 which is very poor and not really adequate for a publicly reported metric. It appears as though model was developed with 2010-2013 and retested and recalibrate with more recent data. Perhaps re-development of the model with more recent data may yield model with better discrimination 2) it is not clear that the model accounts for clustering of events within individual patients--a given patient who is readmitted is likely to be one who is subsequently more likely to be readmitted. One patient with multiple readmissions is not necessarily the same quality as multiple patients with a single readmission and it is not clear that the model can distinguish between the two 3) Decision not to include social risk factors is discussed extensively and rejected out of concern that they are more closely related to hospital than patient effects. This may be appropriate, but there is concern because they do demonstrate a small mean change of o.5 in EDAC--however, the change is not likely to impact the mean or median as much as it might impact hospitals at the edges--and it is the tails that hospitals will be identified as outliers. Therefore, more meaningful analysis would be net reclassification index, especially of hospital which the current model identifies as outliers.

Panel Member 9: Some social factors might have been included, but the empirical evidence is reasonably strong that exclusion does not affect the end result. It should be noted that the same approach to "does it make a difference" would also exclude just about all of the clinical variables on the same logic and analytic approach, but this is an issue that is not unique to this measure.

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

    Submission document: Testing attachment, section 2b4.

    Panel Member 3: Without adequate risk adjustment, and given the elimination of hospitals considered "number of cases too small" for reliable results, it is difficult to assess the empirical results provided.

    Panel Member 4: No concerns as there is a fair degree of variation expressed when testing at the 95% confidence interval. Specifically, of hospitals with 25 or more cases, 12% had "fewer days" and 21% had "more days".

    Panel Member 5: none

    Panel Member 6: No major concerns.

Panel Member 7: none

Panel Member 8: See answer to #19 above

Panel Member 9: The measure can identify statistical outliers - no idea whether the differences observed are meaningful or not.

21. **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 4: No concerns

Panel Member 6: No major concerns.

Panel Member 7: none

Panel Member 8: Use of multiple data sources is merely to define eligible patients or SRFs and is appropriate

Panel Member 9: Not applicable

22. **Please describe any concerns you have regarding missing data.**

Submission document: Testing attachment, section 2b6.

Panel Member 3: None

Panel Member 4: No concerns. Measure developer states there was no missing data as claims were used for this measure.

Panel Member 6: No major concerns.

Panel Member 7: none

Panel Member 8: no concerns

Panel Member 9: None

For cost/resource use measures ONLY:

If not cost/resource use measure, please skip to question 25.

23. **Are the specifications in alignment with the stated measure intent?**
Consider these specific aspects of the measure specifications: attribution, cost categories, target population.

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

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

Panel Member 6: No major concerns.

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: Low model discrimination (c = 0.59)

Panel Member 4: Regarding measure score testing, the correlation with the EDAC measure and: - CMS readmits group rating were moderate to strong at -0.418 [p18] - CMS overall star rating were moderate at -0.371 [p19] -CMS HF readmits rating were moderate to strong at -0.574 [p18] Regarding the face validity testing, the survey results of the group were in the desired direction, e.g. 83% of the group "strongly agreed" or "moderately agreed" to the statement the EDAC measure "can be used to distinguish between better and worse quality hospitals." [p18]

Panel Member 5: Model validation was performed using the development data and not in a validation data set.

Panel Member 6: No major concerns.

Panel Member 8: Low c-statistic of risk model, inadequate testing of impact of leaving social risk factors out of the final model and inadequate accounting for clustering within patients.

Panel Member 9: The specific correlations used to establish empirical validity at the measure score level are not informative, as the same readmission events seem to be included on both sides of the correlations being calculated.

For composite measures ONLY

If not composite, please skip this section.

Submission documents: "xxxx_measure testing form", section 2c

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 6: No major concerns.

Panel Member 7: Choice of variables for construct validation analyses.

Panel Member 9: The developer may be able to do some additional analysis of measure score validity in time for the March meeting - this will be relevant for all of the EDAC measures.

Developer Submission

NQF #: 2880

Corresponding Measures:

De.2. Measure Title: Excess days in acute care (EDAC) after hospitalization for heart failure (HF)

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

De.3. Brief Description of Measure: The measure assesses days spent in acute care within 30 days of discharge from an inpatient hospitalization for HF to provide a patient-centered assessment of the post-discharge period. This measure is intended to capture the quality of care transitions provided to discharged patients who had a HF hospitalization by collectively measuring a set of adverse acute care outcomes that can occur post-discharge: emergency department (ED) visits, observation stays, and unplanned readmissions at any time during the 30 days post-discharge. In order to aggregate all three events, we measure each in terms of days. The Centers for Medicare & Medicaid Services (CMS) annually reports the measure for patients who are 65 years or older, are enrolled in Medicare Fee-For-Service (FFS), and are hospitalized in non-federal short-term acute care hospitals.

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. Safely transitioning patients from hospital to home requires a complex series of tasks which would be cumbersome to capture individually as process measures: timely and effective communication between providers, prevention of and response to complications, patient education about post-discharge care and self-management, timely follow-up, and more. Suboptimal transitions contribute to a variety of adverse events post-discharge, including ED evaluation, need for observation, and readmission.

Measures of unplanned readmission already exist, but there are no current NQF-endorsed measures for ED and observation stay utilization for this condition. It is thus difficult for providers and consumers to gain a complete picture of post-discharge outcomes. Moreover, separately reporting each of these outcomes encourages "gaming," such as re-categorizing readmission stays as observation stays to avoid a readmission outcome. By capturing a range of acute care events that are important to patients, we can produce a more complete picture of post-discharge outcomes that better informs consumers about care quality and incentivizes global improvement in transitional care.

S.4. Numerator Statement: The outcome for this measure is a count of the number of days the patient spends in acute care within 30 days of discharge from an eligible index admission for HF. We define days in acute care as days spent in an ED, admitted to an observation unit, or admitted as an unplanned readmission for any cause to a short-term acute care hospital, within 30 days from the date of discharge from the index HF hospitalization.

Additional details are provided in S.5 Numerator Details.

S.6. Denominator Statement: The target population for this measure is Medicare FFS beneficiaries aged 65 years and older hospitalized at non-Federal and VA acute care hospitals for HF.

The cohort includes admissions for patients discharged from the hospital with a principal diagnosis of HF (codes in the attached Data Dictionary) and with continuous 12 months Medicare enrollment prior to admission. CMS publicly reports this measure 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 measure excludes index hospitalizations that meet any of the following exclusion criteria:

1. Without at least 30 days of post-discharge enrollment in Medicare FFS
2. Discharged against medical advice

3. HF admissions within 30 days of discharge from a prior HF index admission

4. 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.

De.1. Measure Type: Outcome

S.17. Data Source: Claims, Other

S.20. Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Dec 09, 2016 **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 not formally paired with any measure; however, it is harmonized with a measure of hospital-level, all-cause, 30-day, risk-standardized readmission following heart failure hospitalization.

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_2880_HF_EDAC_Evidence_Spring2021_010521_FINAL-637541846122311336.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 (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 2880

Measure Title: Excess days in acute care (EDAC) after hospitalization for heart failure

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

Date of Submission: **4/5/2021**

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

Outcome

☒ Outcome: Excess days in acute care (EDAC) after hospitalization for heart failure–

☐ 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:

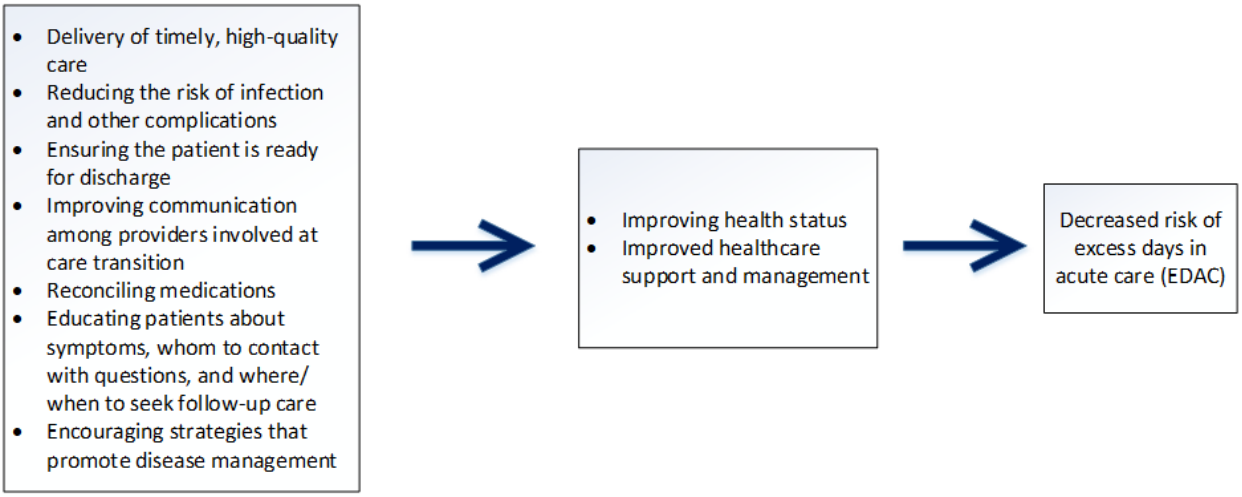
☐ Appropriate use measure:

☐ 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.

Figure 1. HF EDAC Logic Model



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

1a.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.)

****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 excess days in acute care 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 in Figure 1 indicates some of the many care processes that can influence post-discharge acute care utilization after a hospitalization for heart failure. These complex and critical aspects of care – such as communication between providers, patient education, patient safety, and coordinated transitions to the outpatient environment – all contribute to patient outcomes but are difficult to measure by individual process measures. Interventions during and after a hospitalization can be effective in reducing utilization rates in geriatric populations (Benbassat et al., 2000; Naylor et al., 1999; Coleman et al., 2006; Courtney et al., 2009; Koehler et al., 2009) and, particularly, for older patients (Carroll et al., 2007; Young et al., 2003; Bondestam et al., 1995; Ades et al, 1992; Carlhed et al., 2009). Several randomized trials have reduced 30-day readmission rates by 20-40% (Jack et al., 2009; Coleman et al., 2004; Courtney et al., 2009; Garasen et al., 2007; Koehler et al., 2009; Mistiaen et al., 2007; Naylor et al., 1994; Naylor et al., 1999; van Walraven et al., 2002; Weiss et al., 2010; Krumholz et al., 2012; Balaban et al., 2008; Patel et al., 2018). These types of interventions have also been demonstrated to be cost-saving (Naylor et al., 1999; Naylor et al., 2004; Koelling et al., 2005; Krumholz et al., 2002; Stauffer et al., 2011). Outside the randomized controlled trial setting, there is also increasing evidence that hospitals and health plans have been able to reduce readmission rates through more generalizable quality improvement initiatives (Gerhardt et al., 2012; Stauffer et al., 2011; Graham et al., 2012; Harrison et al., 2011; Hernandez et al., 2010; Radhakrishnan et al., 2018).

In the case of heart failure, specifically, a number of strategies have been shown to reduce re-hospitalization. For example, Al Damlugi and colleagues showed that better care coordination at discharge reduced the likelihood of a readmission: discharge summaries that were transmitted to any outpatient clinician were associated with lower odds of readmission, and discharge summaries that included elements related to transitions of care were also associated with lower odds of readmission (Al Damluji et al., 2015). In addition, a meta-analysis found that interventions such as patient education and patient education combined with other

interventions were the most beneficial; interventions that included one or more interventions were 1.4 to 6.8 times less likely to be readmitted (Wan et al., 2017). A review article examining effective strategies to prevent hospitalization and rehospitalization found that comprehensive discharge support, including individualized instruction and early post-discharge follow-up (by phone, home visit, or in clinic) in the immediate post-discharge period reduced mortality and/or readmissions in clinical trials (Horwitz and Krumholz, 2019). Studies have also reported reductions in emergency department (ED) visit rates for patients with other conditions after implementation of interventions that focused on the inpatient and outpatient settings (Bondestam et al., 1995).

The current process-based performance measures cannot capture all the ways that care within the hospital might influence outcomes. As a result, many stakeholders, including patient organizations, are interested in outcomes measures that allow patients and providers to assess relative outcomes performance among hospitals (Bratzler et al., 2007).

In the context of the Centers for Medicare and Medicaid Services' (CMS's) publicly reported readmission measures, the increasing use of ED visits and observation stays has raised concerns that current readmission measures do not capture the full range of unplanned acute care in the post-discharge period (Vashi et al., 2013; Rising et al., 2012; Feng et al., 2012). Observation stays can occur in many different parts of the hospital, including dedicated treatment rooms, the ED, or inpatient units. In particular, there is concern that high use of observation stays could in some cases replace readmissions, and that hospitals with high rates of observation stays in the post-discharge period may therefore have low readmission rates that do not accurately reflect the quality of care (Vashi et al., 2013; Nuckols et al., 2018).

References:

Ades PA, Huang D, Weaver SO. 1992. Cardiac rehabilitation participation predicts lower rehospitalization costs. Am Heart J 123(4 Pt 1):916-921.

Salim Al-Damluji M, Dzara K, Hodshon B, Punnanithinont N, Krumholz HM, Chaudhry SI, Horwitz LI. 2015. Association of discharge summary quality with readmission risk for patients hospitalized with heart failure exacerbation. Circ Cardiovasc Qual Outcomes. 8(1):109-111.

Balaban RB, Weissman JS, Samuel PA, Woolhandler S. Redefining and redesigning hospital discharge to enhance patient care: a randomized controlled study. *J Gen Intern Med.* 2008;23(8):1228-1233.

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

Bondestam E, Breikss A, Hartford M. 1995. Effects of early rehabilitation on consumption of medical care during the first year after acute myocardial infarction in patients > or = 65 years of age. Am J Cardiol 75(12):767-771.

Bratzler, DW, Nsa W, Houck PM. Performance measures for pneumonia: are they valuable, and are process measures adequate. Current Opinion in Infectious Diseases. 20(2):182-189, April 2007.

Carlhed R, Bojestig M, Peterson A, et al. Improved clinical outcome after acute myocardial infarction in hospitals participating in a Swedish quality improvement initiative. Circulation. Cardiovascular Quality & Outcomes. 2009;2(5):458-464.

Carroll DL, Rankin SH, Cooper BA. 2007. The effects of a collaborative peer advisor/advanced practice nurse intervention: cardiac rehabilitation participation and rehospitalization in older adults after a cardiac event. J Cardiovasc Nurs 22(4):313-319.

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

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

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

Feng Z, Wright B, Mor V. Sharp rise in Medicare enrollees being held in hospitals for observation raises concerns about causes and consequences. Health affairs (Project Hope). Jun 2012;31(6):1251-1259.

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

Gerhardt G, Yemane A, Hickman P, Oelschlaeger A, Rollins E, Brennan N. Medicare Readmission Rates Showed Meaningful Decline in 2012. *Medicare & Medicaid Research Review.* 2013;3(2):E1-E12.

Graham J, Tomcavage J, Salek D, Sciandra J, Davis DE, Stewart WF. Postdischarge monitoring using interactive voice response system reduces 30-day readmission rates in a case-managed Medicare population. *Medical Care.* 2012;50(1):50-57.

Harrison PL, Hara PA, Pope JE, Young MC, Rula EY. The impact of postdischarge telephonic follow-up on hospital readmissions. *Population Health Management.* 2011;14(1):27-32.

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

Horwitz L, and Krumholz, H. 2019. Systems-based strategies to reduce hospitalizations in patients with heart failure. UpToDate, last updated March 2019.

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.

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.

Lloyd-Jones D et al, American Heart Association Statistics Committee

Hernandez AF, Greiner MA, Fonarow GC, et al. Relationship between early physician follow-up and 30-day readmission among Medicare beneficiaries hospitalized for heart failure. *Jama.* 2010;303(17):1716-1722.

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

Koehler BE, Richter KM, Youngblood L, Cohen BA, Prengler ID, Cheng D, et al. Reduction of 30-day postdischarge hospital readmission or emergency department (ED) visit rates in high-risk elderly medical patients through delivery of a targeted care bundle. J Hosp Med 2009;4(4):211-218.

Koelling, TM. 2005. Multifaceted outpatient support can improve outcomes for people with heart failure. Commentary. Evid Based Cardiovasc Med 9 Koelling, TM. 2005. Multifaceted outpatient support can improve outcomes for people with heart failure. Commentary. Evid Based Cardiovasc Med 9 (2):138-41.

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

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.

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

Naylor, MD, Brooten D, Campbell R, et al. 1999. Comprehensive discharge planning and home follow-up of hospitalized elders: a randomized clinical trial. JAMA 281 (7):613-20.

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

Naylor MD, Brooten D, Campbell R, et al. 2004. Transitional care of older adults hospitalized with heart failure: a randomized, controlled trial. J Am Geriatr Soc 52 (5):675-84.

Nuckols TK, Fingar KR, Barrett ML, et al. Returns to Emergency Department, Observation, or Inpatient Care Within 30 Days After Hospitalization in 4 States, 2009 and 2010 Versus 2013 and 2014. J Hosp Med. 2018;13(5):296-303. doi:10.12788/jhm.2883.

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.

Rising KL, White LF, Fernandez WG, Boutwell AE. Emergency Department Visits After Hospital Discharge: A Missing Part of the Equation. Annals of Emergency Medicine. (0).

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

Torio CM, Moore BJ. National Inpatient Hospital Costs: The Most Expensive Conditions by Payer, 2013. HCUP Statistical Brief # 204. Available at: http://www.hcup-us.ahrq.gov/reports/statbriefs/sb204-Most-Expensive-Hospital-Conditions.pdf. Published May 2016. Accessed August 23, 2020.

van Walraven C, Seth R, Austin PC, Laupacis A. Effect of discharge summary availability during post-discharge visits on hospital readmission. J Gen Intern Med 2002;17(3):186-92.

Vashi AA, Fox JP, Carr BG, et al. Use of hospital-based acute care among patients recently discharged from the hospital. JAMA. Jan 23 2013;309(4):364-371.

Wallace L., Grady J., Djordjevic, D., et al. 2019 Condition-Specific Measures Updates and Specifications Report Hospital-Level 30-Day Risk-Standardized Readmission Measures: Acute Myocardial Infarction – Version 12.0, Chronic Obstructive Pulmonary Disease – Version 8.0, Heart Failure – Version 12.0, Pneumonia – Version 12.0. 2019; https://qualitynet.org/inpatient/measures/readmission/methodology. Available as of April 23, 2019.

Wan TTH, Terry A, Cobb E, McKee B, Tregerman R, Barbaro SDS. 2017. Strategies to Modify the Risk of Heart Failure Readmission: A Systematic Review and Meta-Analysis. Health Serv Res Manag Epidemiol. 18;4:1-16.

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

Young W, Rewa G, Goodman SG, et al. 2003. Evaluation of a community-based inner-city disease management program for postmyocardial infarction patients: a randomized controlled trial. CMAJ Canadian Med Assn J 169(9):905-910.

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

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

_____

1a.4 OTHER SOURCE OF EVIDENCE

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

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

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

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

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. Safely transitioning patients from hospital to home requires a complex series of tasks which would be cumbersome to capture individually as process measures: timely and effective communication between providers, prevention of and response to complications, patient education about post-discharge care and self-management, timely follow-up, and more. Suboptimal transitions contribute to a variety of adverse events post-discharge, including ED evaluation, need for observation, and readmission.

Measures of unplanned readmission already exist, but there are no current NQF-endorsed measures for ED and observation stay utilization for this condition. It is thus difficult for providers and consumers to gain a complete picture of post-discharge outcomes. Moreover, separately reporting each of these outcomes encourages "gaming," such as re-categorizing readmission stays as observation stays to avoid a readmission outcome. By capturing a range of acute care events that are important to patients, we can produce a more complete picture of post-discharge outcomes that better informs consumers about care quality and incentivizes global improvement in transitional care.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(**This is required for 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.*

We show below the distribution of HF EDAC across the three most recent three-year reporting periods (July 1, 2016-June 30, 2019, July 1, 2015-June 30, 2018, and July 1, 2014-June 30, 2017) for all hospitals. For the most recent reporting period, the range of performance is -70.1 to 259 EDAC per 100 admissions, and the median EDAC is -0.3 per 100 admissions.

We provide the results below with the measure as specified, as requested in the instructions. However, we provide additional analyses removing VA admissions in section 4b1 below to allow for comparison across performance periods for evaluation of improvement.

Periods//YEAR1619//YEAR1518//YEAR1417

Number of Hospitals//4642//4534//4577

Number of Admissions//1286352//1199343//1169795

Mean(SD)//3.3(25.5)//3.3(25.3)//3.4(25.1)

Range(Min to Max)//-70.1 to 259//-66.6 to 340.3//-65 to 147.8

Minimum//-70.1//-66.6//-65.0

10th percentile//-25.6//-25.5//-25.5

20th percentile//-17.5//-17.4//-17.2

30th percentile//-11.1//-10.9//-10.8

40th percentile//-4.7//-4.7//-4.9

50th percentile//-0.3//0.1//0.0

60th percentile//6.4//6.4//6.1

70th percentile//13.0//13.4//12.9

80th percentile//22.3//22.7//22.1

90th percentile//35.6//35.4//36.6

Maximum//259.0//340.3//147.8

We show below the distribution of HF EDAC across the three most recent three-year reporting periods (2016-2019, 2015-2018, and 2014-2017) for hospitals with at least 25 admissions. The range of performance for the most recent reporting period (2016-2019) was -59.7 to 154.4 EDAC per 100 admissions; the median was 2.3 EDAC per 100 admissions.

Periods//YEAR1619//YEAR1518//YEAR1417

Number of Hospitals//3713//3643//3690

Number of Admissions//1275344//1188842//1159275

Mean(SD)//4.3(24.9)//4.3(24.8)//4.5(25.2)

Range(Min to Max)//-59.7 to 154.4//-66.6 to 143.2//-65 to 147.8

Minimum//-59.7//-66.6//-65.0

10th percentile//-25.4//-25.7//-25.2

20th percentile//-16.5//-17.2//-16.5

30th percentile//-10.1//-9.8//-9.8

40th percentile//-3.5//-3.4//-3.6

50th percentile//2.3//2.4//2.4

60th percentile//8.5//8.4//8.4

70th percentile//14.8//15.2//14.6

80th percentile//24.0//24.1//23.9

90th percentile//36.3//36.1//37.5

Maximum//154.4//143.2//147.8

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 EDAC by Proportion of Dual Eligible Patients:

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

Dates of Data: July 2016 through June 2019

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

Description of Social Risk Variable//Dual Eligibility

Quartile//hospitals in the first quartile for the proportion patients with of dual-eligible status //Hospitals in the fourth (highest) quartile for the proportion patients with of dual-eligible status

Social Risk Proportion(%)//q1:(0-10.2%)//q4:(24.5-100%)

# of Hospitals//928//928

Maximum//148.8//154.4

90th percentile//31.0//45.5

75th percentile//13.8//28.4

Median//-0.7//6.6

25th percentile//-14.6//-11.2

10th percentile//-26.9//-24.1

Minimum//-54.0//-59.7

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

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

Dates of Data: July 2016 through June 2019

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

Description of Social Risk Variable //AHRQ SES Index

Social Risk Proportion (%)//q1:(0-8.5%)//q4:(35.1-100%)

# of Hospitals//921//921

Maximum//114.3//130.4

90th percentile//24.9//42.6

75th percentile//11.8//24.4

Median//-3.6//6.6

25th percentile//-17.1//-9.6

10th percentile//-29.5//-22.5

Minimum//-56.1//-59.7

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://www.qualitynet.org/inpatient/measures/edac/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 **Attachment:** NQF_datadictionary_HF-EDAC_Spring2021.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.

N/A

S.4. Numerator Statement *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) 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 a count of the number of days the patient spends in acute care within 30 days of discharge from an eligible index admission for HF. We define days in acute care as days spent in an ED, admitted to an observation unit, or admitted as an unplanned readmission for any cause to a short-term acute care hospital, within 30 days from the date of discharge from the index HF hospitalization.

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).*

Outcome Definition

The measure counts ED treat-and-release visits, observation stays, and readmissions to any short-term acute care hospital for any cause within 30 days of the discharge date of the index HF admission, excluding planned readmissions as defined below. Each ED treat-and-release visit is counted as one half-day (0.5 days). Observation stays are recorded in terms of hours and converted for the measure into half-days (rounded up). Each unplanned readmission day is counted as one full day (1 day). We count all eligible outcomes occurring in the 30-day period, even if they are repeat occurrences. Thus, an unplanned readmission that follows a planned readmission is still counted.

Rationale: From a patient perspective, days in acute care from any cause is an adverse event. In addition, making inferences about quality issues based solely on the documented cause of an acute care event is difficult. For example, a patient with HF who develops a hospital-acquired infection may ultimately be readmitted for sepsis. In this context, considering the readmission to any acute care setting to be unrelated to the care that the patient received for HF during the index admission would be inappropriate. Multiple events are counted in order to capture the full patient experience in the post-discharge period. Outcomes occurring within 30 days of discharge can be influenced by hospital care. The 30-day time frame is a clinically meaningful period for hospitals to collaborate with their communities to reduce days in acute care.

All eligible outcomes occurring in the 30-day period are counted, even if they are repeat occurrences. For example, if a patient returns to the ED three times on three different days, we count each ED visit as a half-day. Similarly, if a patient has two unplanned hospitalizations within 30 days, the days spent in each are counted. Therefore, the measure may include multiple ED visits, observation stays, and/or readmissions per patient. This approach is taken in order to capture the full patient experience in the post-discharge period. If a hospitalization or observation stay extends beyond the 30-day window, only those days within the 30-day window are counted.

The measure incorporates "exposure time" (the number of days each patient survives after discharge, up to 30). This exposure time is included to account for differential risk for EDAC after discharge among those patients who do not survive the full post-discharge period.

Planned Readmission Algorithm (Version 4.0)

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

The Planned Readmission Algorithm has three fundamental principles:

1. A few specific, limited types of care are always considered planned (transplant surgery, maintenance chemotherapy/immunotherapy, rehabilitation);

2. Otherwise, a planned readmission is defined as a non-acute readmission for a scheduled procedure; and

3. Admissions for acute illness or for complications of care are never planned.

The algorithm was developed in 2011 as part of the Hospital-Wide Readmission measure. In 2013, CMS applied the algorithm to its other readmission measures. In applying the algorithm to condition- and procedure-specific measures, teams of clinical experts reviewed the algorithm in the context of each measure-specific patient cohort and, where clinically indicated, adapted the content of the algorithm to better reflect the likely clinical experience of each measure's patient cohort. For the CMS 30-day HF EDAC measure, CMS used the Planned Readmission Algorithm without making any changes. The Planned Readmission Algorithm is updated annually to ensure changes in coding are captured to maintain the algorithms relevance.

For more details on the Planned Readmission Algorithm, please see the report titled "Condition-Specific Measures Updates and Specifications Report Hospital-Level 30-Day Risk-Standardized Excess Days in Acute Care Measures for HF, version 4.0" posted in data field S.1 or at

https://www.qualitynet.org/inpatient/measures/edac/methodology.

Definition of Emergency Department Visit and Observation Stay

We defined ED visits and observation stays using specified billing codes or revenue center codes identified in Medicare hospital outpatient claims and physician carrier claims. The codes that define ED visits and observation stays are in the attached Data Dictionary.

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

The target population for this measure is Medicare FFS beneficiaries aged 65 years and older hospitalized at non-Federal and VA acute care hospitals for HF.

The cohort includes admissions for patients discharged from the hospital with a principal diagnosis of HF (codes in the attached Data Dictionary) and with continuous 12 months Medicare enrollment prior to admission. CMS publicly reports this measure 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 inclusion criteria:

1. Have a principal diagnosis of HF;

2. Enrolled in Medicare FFS Part A and Part B for the 12 months prior to the date of the admission, and enrolled in Part A during the index admission, or those who are VA beneficiaries;

3. Aged 65 or over;

4. Discharged alive from a non-federal short-term acute care hospital (including Indian Health Service hospitals) and critical access hospitals; and,

5. Not transferred to another acute care facility.

Cohort codes are included in the attached data dictionary.

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

The measure excludes index hospitalizations that meet any of the following exclusion criteria:

1. Without at least 30 days of post-discharge enrollment in Medicare FFS

2. Discharged against medical advice

3. HF admissions within 30 days of discharge from a prior HF index admission

4. 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.

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 measure excludes index hospitalizations that meet any of the following exclusion criteria:

1. Without at least 30 days of post-discharge enrollment in Medicare FFS (in the case of patients who are not VA beneficiaries), determined by examining the Medicare Enrollment Database (EDB).

Rationale: The 30-day outcome cannot be assessed in this group since claims data are used to determine whether a patient visited the ED, was placed under observation, or was readmitted.

2. Discharged against medical advice, 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 prior HF index admission, 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 is not considered 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, identified via claims data

Rationale: These patients represent a clinically distinct group (ICD-10-PCS code list).

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; this measure is not stratified.

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

Statistical risk model

If other:

S.12. Type of score:

Other (specify):

If other: Excess days in acute care (EDAC) per 100 discharges

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 EDAC following hospitalization for HF using a random effects hurdle model. This model consists of the two-part logit/truncated Poisson model specifications for days in acute care and includes two random effects for hospitals – one for the logit part and one for the truncated Poisson part – with a non-zero covariance between the two random effects. This strategy accounts for within-hospital correlation of the observed outcome and accommodates the assumption that underlying differences in quality across hospitals lead to systematic differences in outcomes.

Specifically, CMS calculates EDAC, for each hospital, as the difference ("excess") between a hospital's predicted days and expected days per 100 discharges. "Predicted days" is the average number of days a hospital's patients spent in acute care after adjusting for the risk factors (included in the attached data dictionary). "Expected days" is the average number of risk-adjusted days in acute care a hospital's patients

would have been expected to spend if discharged from an average performing hospital with the same case mix. We risk adjust the day count to account for age, gender, and comorbidities. The model used is appropriate for count data, and we incorporate exposure time to account for survival times shorter than 30 days. To be consistent with the reporting of the CMS 30-day AMI, HF, and pneumonia readmission measures, CMS multiplies the measure result by 100 such that the final EDAC measures represent EDAC per 100 discharges.

To assess hospital performance for each reporting period, we re-estimate the parameter estimates using the years of data in that period.

The random effects hurdle models are described fully in the original measure methodology report

(Horwitz et al., 2015).

References:

1. Horwitz L, Wang C, Altaf F, et al. 2015. Excess Days in Acute Care after Hospitalization for Heart Failure (Version 1.0) Final Measure Methodology Report.

https://www.qualitynet.org/inpatient/measures/edac/methodology

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

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

If other, please describe in S.18.

Claims, 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:

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

For development purposes, we obtained the Medicare Part B hospital and physician outpatient claims from the Chronic Condition Data Warehouse (CCW) 100% condition-specific datasets.

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

Reference:

Fleming C, Fisher ES, Chang CH, Bubolz TA, Malenka DJ. Studying outcomes and hospital utilization in the elderly: The advantages of a merged data base for Medicare and Veterans Affairs hospitals. Medical Care. 1992; 30(5): 377-91. Data sources for the all-payer updateData sources for the Medicare FFS measure:

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

For development purposes, we obtained the Medicare Part B hospital and physician outpatient claims from the Chronic Condition Data Warehouse (CCW) 100% condition-specific datasets.

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

Reference:

Fleming C, Fisher ES, Chang CH, Bubolz TA, Malenka DJ. Studying outcomes and hospital utilization in the elderly: The advantages of a merged data base for Medicare and Veterans Affairs hospitals. Medical Care. 1992; 30(5): 377-91. Data sources for the all-payer update

S.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)*

Emergency Department and Services, 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

2020_EDAC_MU_SpecsReport_-2-.pdf,NQF_2880_HF_EDAC_Testing_Spring2021_010521_FINAL-637541844968243377.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*)**:** 2880
Measure Title: Excess days in acute care (EDAC) after hospitalization for heart failure
Date of Submission: **1/5/2021**

Type of Measure:

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

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 (including the Master Beneficiary Summary File), VHA Administrative Data	☒ other: Census Data/American Community Survey, VHA Administrative Data, Medicare Enrollment Data (including the Master Beneficiary Summary File)

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, Medicare Enrollment Database (EDB) data. 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 EM 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

For measure development, we used three years of Medicare administrative claims data (July 2010–June 2013). 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 three years of data (July 2010 – June 2013) into two equal samples: the **Development Dataset** and the **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 Section 2b3.6. Statistical Risk Model Discrimination Statistics Section 2b3.7. Statistical Risk Model Calibration Statistics	Entire Cohort: Dates of Data: July 1, 2010 – June 30, 2013 This cohort was randomly split for initial model testing. First half of split sample -Number of Admissions: 590,448 -Number of Measured Hospitals: 4,626 mean age = 81 years; % male = 44.1 Second half of split sample -Number of Admissions: 590, 447 -Number of Measured Hospitals: 4,634 mean age = 81 years; % male = 44.1
Endorsement Maintenance (EM) 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 Section 2b3.6. Statistical Risk Model Discrimination Statistics Section 2b4 Meaningful Differences	Dates of Data: July 2016-June2019 Number of admissions = 1,286,352 Patient Descriptive Characteristics: Mean age=80.5; % Male = 48.4 Number of measured hospitals (total): 4,642 Number of hospitals with at least 25 admissions: 3,713 This cohort was randomly split into two halves for split-sample reliability testing. First half of split sample

Dataset	Applicable Section in the Testing Attachment	Description of Dataset
		- Number of Admissions: 644,305. - Number of measured hospitals: 4,642. Patient Descriptive Characteristics: Mean age =80.5; % Male =48.4. Second half of split sample - Number of Admissions: 642,047. - Number of measured hospitals: 4,593. Patient Descriptive Characteristics: Mean age=80.5; % Male = 48.3.
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 EDAC 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 consistently captured in a reliable fashion for all patients in this measure. There is a large body of literature linking various SRFs to worse health status, greater use of the emergency department, 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 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 7 variables): percentage of people in the labor force who are unemployed, percentage of people living below poverty level, median household income, median value of owner-occupied dwellings, percentage of people ≥25 years of age with less than a 12th grade education, percentage of people ≥25 years of age completing ≥4 years of college, and percentage of households that average ≥1 people per room).

The 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 sociodemographic status (SDS) indicators. We ultimately did not include the ADI at the time, partly due to the fact that the coefficients used to derive ADI had not been updated for many years. Recently, the coefficients for ADI have been updated and therefore we compared the ADI with the AHRQ SES Index and found them to be highly correlated. In this submission, we present analyses using the census block level, the most granular level possible using ACS data. A census block group is a geographical unit used by the US Census Bureau which is between the census tract and the census block. It is the smallest geographical unit for which the bureau publishes sample data. The target size for block groups is 1,500 and they typically have a population of 600 to 3,000 people. We used 2013-2017 ACS data and mapped patients' 9-digit ZIP codes via vendor software to the census block group level. Given the variation in cost of living across the country, the median income and median property value components of the AHRQ SES Index were adjusted by regional price parity values published by the Bureau of Economic Analysis (BEA). This provides a better marker of low SES neighborhoods in high expense geographic areas. We then calculated an AHRQ SES Index score for census block groups that can be linked to 9-digit ZIP codes. We used the percentage of patients with an AHRQ SES index score equal to or below 46.0 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.

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/pdf-report/second-impact-report-to-congress. Accessed January 4, 2021.

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.

H. F. Chen, A. Nevola, T. M. Bird, 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.

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 calculated split-sample reliability for the EDAC measures. We are unable to calculate signal-to-noise reliability because it is only computable for single one-way random effect models where the between-hospital variance is scalar. Our EDAC measures employ a hurdle model with correlated random effects where the hospital random effects follow a multivariate normal distrubtion with means of zero. Specifically, the hurdle model has two components: a random-effect logit model to estimate the odds of having at least one EDAC day within the entire population and a random-effect zero-truncated Poisson model to estimated EDAC days within patients with at least one EDAC day. Thus, there are two between- hospital variances, and there is no clear method for creating a single signal to noise estimate using two signal estimates.

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 (which we refer to here as split-sample reliability) 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 aged 65 years and older, 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 published standards. Specifically, we used a combined 2016-2019 sample, randomly split it into two approximately equal subsets of patients, and calculated EDAC for each hospital for each sample. The agreement of the two EDAC scores 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 split-sample 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.

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 and seek to avoid variables which do not meet this standard.

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

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.

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.

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*)

Split-Sample Reliability

We show split-sample reliability results (ICC with Spearman Brown adjustment) below in Table 2, for different volume cutoffs.

Table 2. Split-sample reliability for the HF EDAC measure score

Hospitals Included for ICC calculation	Split-Sample ICC with Spearman Brown Adjustment	Number of Hospitals	Number of Admissions	Percent of Hospitals	Percent of Admissions
>=2 admissions	0.465	4,593	1,286,303	98.9	100
>=25 admissions	0.527	3,713	1,275,344	80.0	99.1
>=50 admissions	0.573	3,139	1,254,618	67.6	97.5
>=100 admissions	0.632	2,491	1,207,900	53.7	93.9
>=200 admissions	0.662	1,914	1,124,154	41.2	87.4
>=300 admissions	0.698	1,515	1,024,819	32.6	79.7

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

Measure score reliability, calculated using the split-sample approach, ranges from 0.465 for hospitals with at least two admissions to 0.698 for hospitals with at least 300 admissions.

The interpretation of reliability results depends on context: the reliability method used (e.g. split sample vs signal-to-noise), the type of measure (e.g. process vs. outcome), the complexity of the measure (e.g. complex clinical measure with risk-adjusted or not), as well as provider volume, variation in the measure score, and the implementation approach, all play into the interpretation of reliability results. While there are different

published scales for assessment of reliability (Adams, 2010; Landis & Koch, 1977), the scales are insufficiently nuanced, and there is currently no widely accepted standard. However, split-sample reliability is considered to be the lower bound of true reliability (Adams, 2009).

One approach to determining the adequacy of a reliability result is to compare the results of reliability testing in similar contexts using the same method. We identified several studies that we think support the assessment of this measure as having adequate reliability when assessing split-sample (or test-retest) reliability in the context of hospital measurement. For example, Hall et al calculated test-retest reliability for determining comorbidities from chart abstraction (Hall et al., 2006). In this study, multiple abstracters abstracted the same charts and the results were used to calculate four different common comorbidity scores. For three of the indices, test-retest reliabilities ranged from 0.59-0.68, with the fourth (the Charlson comorbidity score) achieving 0.80. We would argue that chart abstraction, with test-retest reliabilities in the 'moderate' to 'substantial' range, should be inherently more reliable than measuring hospital quality. In addition, Cruz et al report reliabilities for collecting risk factor information from patients presenting to an emergency department with potential acute coronary syndrome (ACS) (Cruz et al., 2009). Each patient was queried twice, once by a clinician and once by a trained research assistant, and the reliabilities for a range of risk factors were calculated; these ranged from 0.28 (associated symptoms) to 0.69 (cardiac risk factors), with all other factors in the 0.30-0.56 range. Finally, Hand et al report test-retest reliabilities for bedside clinical assessment of suspected stroke (Hand et al., 2006). Pairs of observers independently assessed suspected stroke patients; findings were recorded on a standard form to promote consistency. The reliabilities were calculated for the full range of diagnostic factors: for vascular factors, reliabilities ranged from 0.47-0.69 with only four of eight above 0.6; for history, they ranged from 0.37-0.65 with only five of 12 above 0.6; other categories were similar (though reliability=1 for whether the patients were conscious).

CMS currently publicly reports results for this measure for hospitals with at least 25 admissions, and hospitals are assessed in performance categories based on 95% confidence intervals of "better," "worse," or "no different" than the national average. In the public interest of transparency by the government, there is a trade-off when selecting a volume cutoff, which limits the number of hospitals with publicly-reportable data. Please note that this measure is used in a pay-for-reporting and not in a pay-for-performance program.

In summary, we interpret the reliability of this measure as adequate in the context of split-sample reliability for a clinical risk-adjusted outcome measure.

References:

Adams J. The reliability of provider profiling, a tutorial. RAND Corporation, 2009. Available at: https://www.rand.org/content/dam/rand/pubs/technical_reports/2009/RAND_TR653.pdf. Accessed on October 13, 2020.

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

Cruz CO, Meshberg EB, Shofer FS, McCusker CM, Chang AM, Hollander JE. Interrater reliability and accuracy of clinicians and trained research assistants performing prospective data collection in emergency department patients with potential acute coronary syndrome. Ann Emerg Med. 2009 Jul;54(1):1-7.

Hall SF, Groome PA, Streiner DL, Rochon PA. Interrater reliability of measurements of comorbid illness should be reported. J Clin Epidemiol. 2006 Sep;59(9):926-33.

Hand PJ, Haisma JA, Kwan J, Lindley RI, Lamont B, Dennis MS, Wardlaw JM. Interobserver agreement for the bedside clinical assessment of suspected stroke. Stroke. 2006 Mar;37(3):776-80.

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

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)*

Validity of Claims-Based Measures

Our team has demonstrated for a number of prior measures the validity of claims-based measures for profiling hospitals by comparing either the measure results or individual data elements against medical records. CMS validated six NQF-endorsed measures currently in public reporting (AMI, heart failure, and pneumonia mortality and readmission) with models that used chart-abstracted data for risk adjustment. Specifically, claims model validation was conducted by building comparable models using abstracted medical chart data for risk adjustment for heart failure patients (National Heart Failure data) (Krumholz et al. 2006; Keenan et al. 2008), AMI patients (Cooperative Cardiovascular Project data) (Krumholz, Wang, et al. 2006), and pneumonia patients (National Pneumonia Project dataset) (Bratzler et al. 2011). When both models were applied to the same patient population, the hospital risk-standardized rates estimated using the claims-based risk-adjustment models had a high level of agreement with the results based on the medical record model, supporting the use of the claims-based models for public reporting. This measure uses the same risk-adjustment variables that were previously validated in the chart review studies.

Validity Indicated by Established Measure Development Guidelines

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

Validity as Assessed by External Groups

Throughout measure development, we obtained expert and stakeholder input via three mechanisms in the initial, early phase of development: a discussion with an advisory Methodology Workgroup, discussions with a national TEP, and a 30-day public comment period in order to increase transparency and to gain broader input on the measure.

The Methodology Workgroup meeting addressed key issues related to measure methodology, including weighing the pros and cons of and measure specifications, modeling, and use (e.g., defining the measure cohort and outcome) to ensure the measure is meaningful, useful, and well-designed. The group provided a forum for focused expert review and discussion of technical issues during measure development.

List of Methodology Workgroup Members:

1) Arlene Ash, PhD; University of Massachusetts Medical School (Professor and Division Chief)

2) Jeremiah Brown, MS, PhD; The Dartmouth Institute for Health Policy and Clinical Practice (Assistant Professor of Health Policy and Clinical Practice)

3) Grant Ritter, PhD, MS, MA; Schneider Institute for Health Policy & Heller Graduate School (Senior Scientist)

4) Patrick Romano, MD, MPH; University of California Davis School of Medicine (Professor of Medicine and Pediatrics)

In alignment with the CMS MMS, we convened a TEP to provide input and feedback during measure development from a group of recognized experts in relevant fields. To convene the TEP, we released a public call for nominations and selected individuals to represent a range of perspectives, including physicians, consumers, purchasers, as well as individuals with experience in quality improvement, performance measurement, and health care disparities. We held two structured TEP conference calls consisting of a presentation of key issues, our proposed approach, and relevant data, followed by open discussion among TEP members. We solicited additional input and comments from the TEP via e-mail between meetings.

Following completion of the preliminary model, we solicited public comment on the measure through the CMS site link http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/CallforPublicComment.html. The public comments were then posted publicly for 30 days. The resulting input was taken into consideration during the final stages of measure development, and led to supplementary analyses reported in the application (1b.4).

Face Validity as Determined by Technical Expert Panel

One means of confirming the validity of this measure was face validity assessed by our TEP, which included 16 members, including patient representatives, expert clinicians, researchers, providers, and purchasers.

List of TEP members:

1) Kevin E. Driesen, PhD, MPH, MA; Center for Rural Health Mel and Enid Zuckerman College of Public Health, University of Arizona (Assistant Professor & Director of the Arizona Rural Hospital Flexibility Program)

2) David Engler, PhD; America's Essential Hospitals (Senior Vice President for Leadership and Innovation)

3) Timothy Farrell, MD; University of Utah School of Medicine (Assistant Professor of Medicine, Geriatrics; Adjunct Professor of Family Medicine)

4) Karen Farris, PhD; University of Michigan College of Pharmacy (Charles R. Walgreen III Professor of Pharmacy Administration; Director of the Social and Administrative Pharmacy Graduate Program)

5) Maura C. Feldman, MSW; Blue Cross Blue Shield of Massachusetts, Inc. (Director for Hospital Performance Measurement and Improvement)

6) Jay A. Gold, MD, JD, MPH; Meta Star, Inc. (Vice President & Chief Medical Officer)

7) Sally Hinkle, DNP, MPA, RN; Temple University Hospital (Director of Performance Improvement & Clinical Value)

8) Amy J.H. Kind, MD, PhD; University of Wisconsin School of Medicine and Public Health (Assistant Professor of Geriatrics)

9) Marjorie King, MD, FACC, MAACVPR; Helen Hayes Hospital (Director of Cardiac Services)

10) Eugene Kroch, PhD; University of Pennsylvania (Adjunct Faculty at the Health Care Systems Department); Premier, Inc. (Vice President & Chief Scientist) University of Pennsylvania; Philadelphia, PA

11) Keith D. Lind, JD, MS, BSN; American Association of Retired Persons (AARP) Public Policy Institute (Senior Policy Advisor)

12) Grace McConnell, PhD; Patient representative

13) Michael A. Ross, MD, FACEP; Emory University School of Medicine (Medical Director of Observation Medicine and Chest Pain Center; Professor of Emergency Medicine)

14) Mark Louis Sanz, MDl; International Heart Institute of Montana (Interventional Cardiologist)

15) Paul Takahashi, MD; Mayo Clinic College of Medicine (Associate Professor of Medicine)

16) Patient representative

We systematically assessed the face validity of the measure score as an indicator of quality by soliciting the TEP members' agreement with the following statement: "The risk-standardized acute care days obtained from the measures as specified can be used to distinguish between better and worse quality hospitals."

We measured agreement on a six-point scale: 1=Strongly disagree, 2=Moderately disagree, 3=Somewhat disagree, 4=Somewhat agree, 5=Moderately agree, 6=Strongly agree.

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, justify the use of face validity only. To meet this requirement for the HF EDAC measure, we identified and assessed the measure's correlation with other measures that target the same domain of quality for the same or similar populations. The goal was to identify if better performance on this measure was related to better performance on other relevant structural or outcome measures. After literature review and consultations with measure experts in the field, there were very few measures identified that assess the same domains of quality. While ideally we would compare outcome measures with process measures that would be predicted to be associated with the outcome based on the logic model presented in the evidence form, there are no existing validated process measures with publicly available data that can be used for this purpose. In addition, CMS has been moving away from process measures and has been removing process measures from use; process measures can also often be "topped out" and therefore not useful for comparison purposes. Developers may also have limited ability to access proprietary data.

Given these challenges, 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 CMS' *Care Compare*, formerly *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, timeliness of care). The readmission group is comprised of the readmission measures that are publicly reported on *Care Compare*. The readmission group score is derived from a latent-variable model that identifies an underlying quality trait for that group. The readmission group score is on a higher-is-better scale. For the validity testing presented in this testing form, we used readmission group scores from Medicare FFS hospitals from January 2020. 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 summary score**: CMS's Overall Hospital Star Rating assesses hospitals' overall performance (shown on *Care 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, timeliness of care). Each group is comprised of individual measures that are reported on *Care 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 summary score using fixed weights; overall hospital summary scores are then clustered, using k-means clustering, into five groups and are assigned one to five stars (the hospital's Star Rating), with more stars indicating a better rating. The hospital summary score is also on a higher-is-better scale. For the validity testing presented in this testing form, we used hospitals' overall summary scores from Medicare FFS hospitals from January 2020. The full methodology for the Overall Hospital Star Rating can be found at https://www.qualitynet.org/inpatient/public-reporting/overall-ratings/resources
3. **Heart Failure (HF) Readmission Measure**: The HF readmission measure estimates a hospital-level 30-day, all-cause, risk-standardized readmission rate (RSRR) for patients discharged from the hospital with a principal diagnosis of heart failure (HF). The HF readmission measure complements the HF EDAC measure because it provides information on a narrower range of unplanned acute care utilization following initial 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 readmission measure can be found at: https://qualitynet.cms.gov/files/5eaaf35ee8ffc8001f999241?filename=2020_Rdmn_CSR.pdf.

We examined the relationship of performance on the HF EDAC measure scores with each of these external measures of hospital quality. For the external measures, the comparison was against performance within quartiles of readmission group scores, overall hospital summary scores, or HF readmission scores.

We hypothesized the strength and the direction of the relationship for each measure. For the HF EDAC measure, a lower measure score means better performance, therefore for comparator measures where better performance is hypothesized to be related to be better performance on HF (such as the Star Rating overall summary score and readmission group score), the direction of the association should be negative. We predicted the HF EDAC scores would be more strongly, positively correlated associated with the readmission group score than the Star Rating overall summary score. We also hypothesized that HF EDAC and HF readmission measure scores would show a strong, positive correlation.

In addition to providing empirical evidence, we have found multiple sources that support that EDAC, particularly the readmission component, can represent a signal of hospital quality. Readmissions have been shown to be associated with low hospital quality. As outlined in the evidence attachment, included with this submission, 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.

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.

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.

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.

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

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.

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.

Koehler BE, Richter KM, Youngblood L, et al. Reduction of 30-day post-discharge 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

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):1683-1701.

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.

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.

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.

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

Systematic Assessment of Face Validity

The results of the TEP rating of agreement with the validity statement were as follows:

Mean rating = 5, moderately agree

Table 3. Face Validity Poll Results for HF EDAC

Rating	# of Responses (N=12)	Percent (%)	Cumulative Percent (%)
6 (Strongly agree)	4	33.3%	33.3%
5 (Moderately agree)	6	50.0%	83.3%
4 (Somewhat agree)	1	8.3%	91.7%
3 (Somewhat disagree)	0	0.0%	91.7%

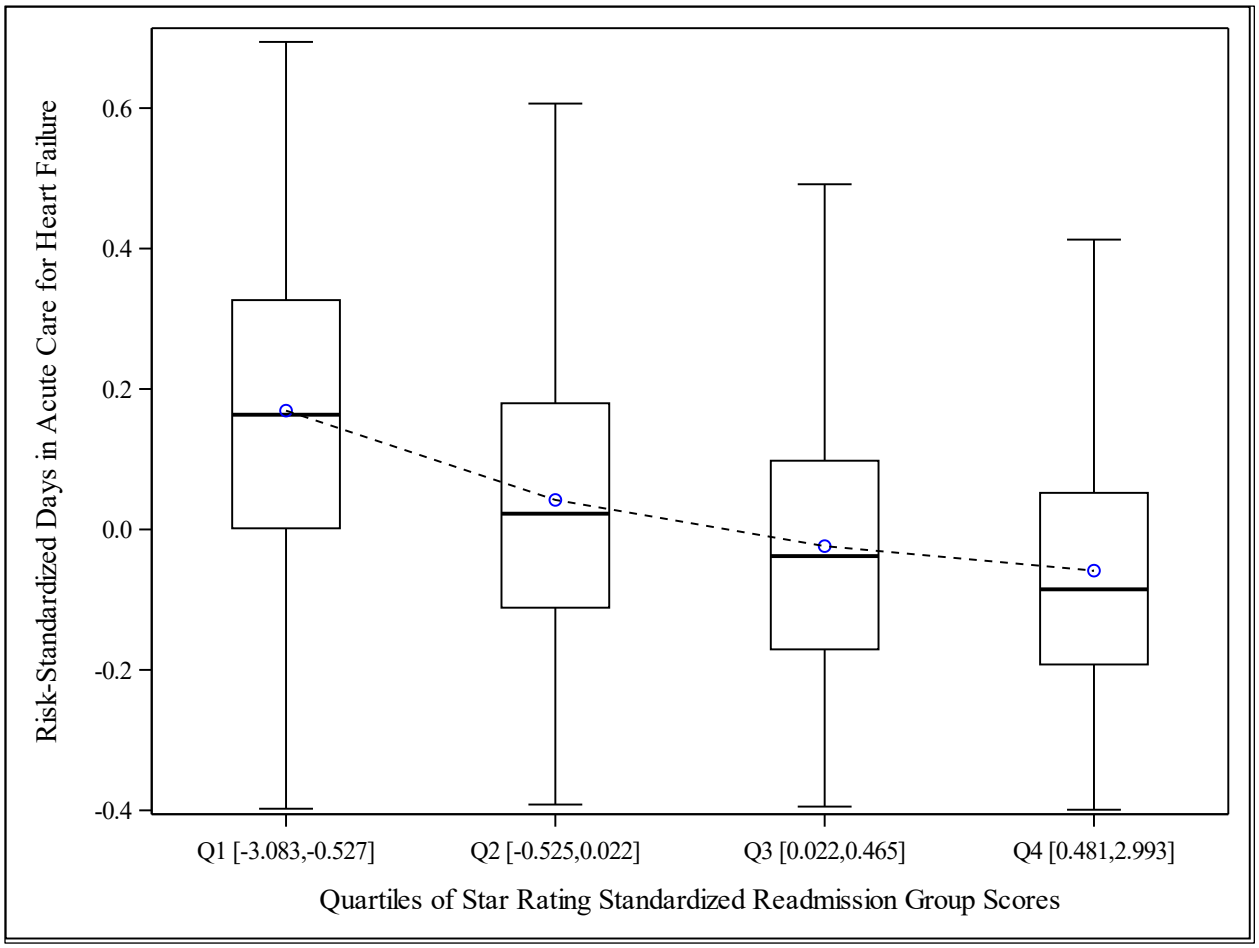
2 (Moderately disagree)	1	8.3%	100.0%
1 (Strongly disagree)	0	0.0	100.0%

Empiric Validity

Correlation between HF EDAC Scores and Star Rating Readmission Group Scores

Figure 1 shows the box-whisker plots of the HF EDAC scores within each quartile of Star Rating Readmission group scores. The blue circles represent the mean HF EDAC score within each quartile of the Star Rating readmission group score. The correlation between HF EDAC scores and Star-Rating readmissions group score is -0.418 (p<.0001), which suggests that hospitals with lower HF EDAC scores (better performance) are more likely to have higher Star Rating readmission group scores (better performance).

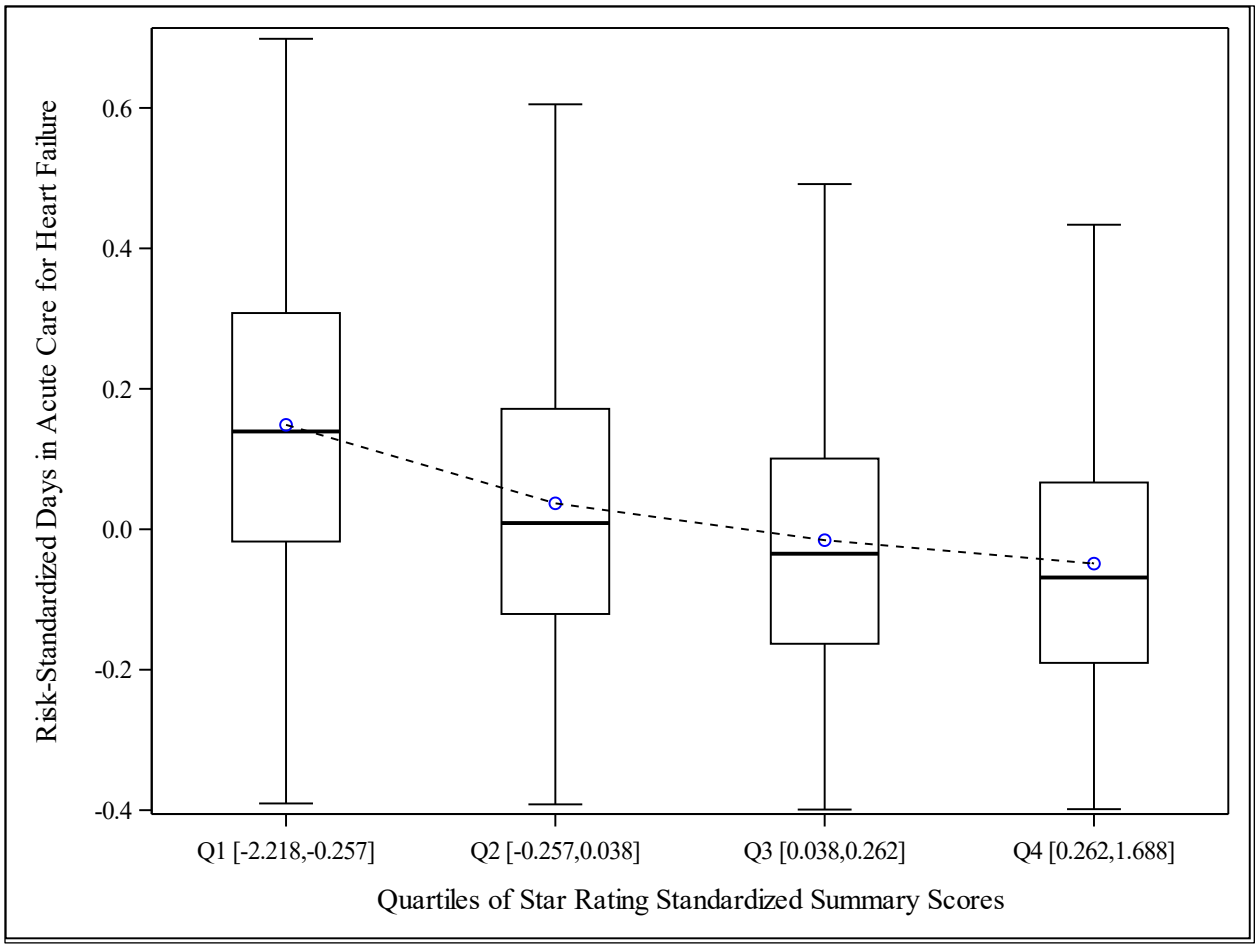
Figure 1. Box-whisker plots of the HF EDAC scores within each quartile of Star Rating readmission scores (n=4,257)



Correlation between HF EDAC Scores and Overall Hospital Star Rating Scores

Figure 2 shows the box-whisker plots of the HF EDAC scores within each quartile of Star Rating overall summary scores. The blue circles represent the mean HF EDAC measure score within each quartile of Star Rating summary score. The correlation between HF EDAC scores and Star-Rating summary score is -0.371 (p<.0001) which suggests that hospitals with lower HF EDAC scores (better performance) are more likely to have higher Star Rating summary scores (better performance).

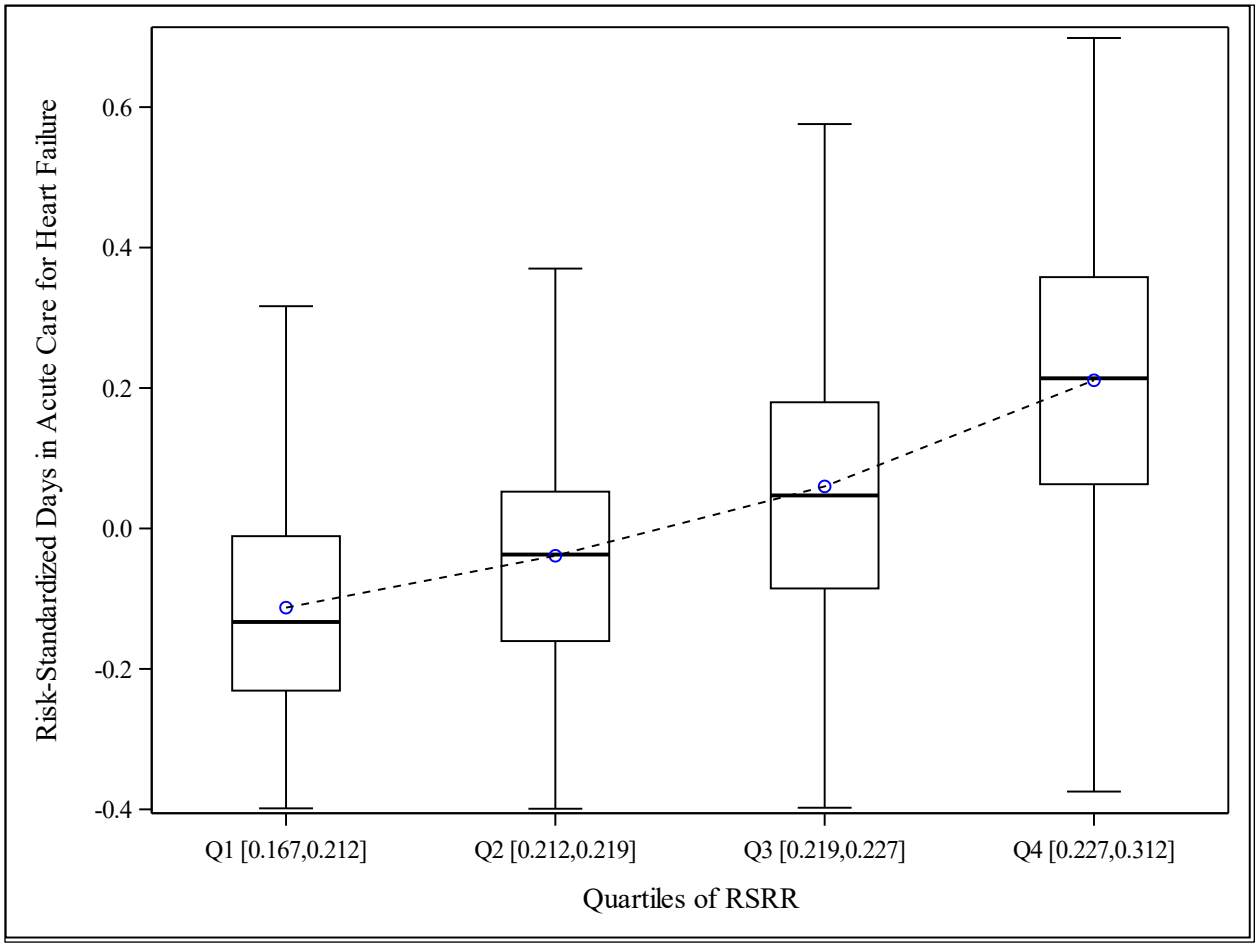
Figure 2. Box-whisker plots of the HF EDAC measure scores within each quartile of Star Rating overall summary scores (n=4,380)



Correlation between HF EDAC Scores and HF Readmission Measure Scores

Figure 3 shows the box-whisker plots of the HF EDAC measure scores and the HF Readmission measure scores (risk-standardized readmission rates or RSRRs.) The blue circles represent the mean HF EDAC measure score within each quartile of the HF Readmission measure score. The correlation between HF EDAC scores and HF RSRRs is 0.574 ($p<.0001$), which suggests that hospitals with lower HF EDAC scores (better performance) are more likely to have lower HF RSRRs (better performance).

Figure 3. Box-whisker plots of the HD EDAC scores and HF Readmission measure scores (RSRRs) (n=4,642)



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?*)

The validity of the HF EDAC measure is supported by three types of evidence: (1) strong face validity as gauged by feedback from Technical Expert Panel (TEP), (2) external empiric comparisons with other quality measures, and (3) validity of the outcome.

Face validity

The validity of the measure is supported by strong face validity results, as measured by systematic feedback from the TEP. As shown in Table 3 above, 11 of 12 (91.7%) TEP members strongly, moderately, or somewhat agreed with the statement: "The risk-standardized acute care days obtained from the measures as specified can be used to distinguish between better and worse quality hospitals."

Empirical Validity Testing

The validity of the measure is further supported by the empiric validation results which demonstrate a correlation (in the expected strength and direction) between the HF EDAC measure and other quality measures, such as the Star Rating readmission group score, the Star Rating overall summary score, and the HF readmission measure score. As expected, we found a stronger association between the HF EDAC measure score and the HF readmission measure score, compared with the association with the Star Ratings readmission group score or the Star Rating summary score. Please note that while demonstrating validity through correlations with other valid measures is an accepted approach, developers may be limited as to the the publicly available data that are avialable for these types of studies.

Validity of the Outcome

The validity of the outcome is supported by the relationship between care processes and the outcome of hospital readmission, emergency room visit, or observation stay. As discussed in the evidence attachment (included with this submission), interventions during and after a hospitalization can be effective in reducing utilization rates in geriatric populations. For HF specifically, implementing interventions that focus on the inpatient and outpatient settings have resulted in reductions in emergency department (ED) visit rates, and high-quality discharge summaries and timely transmission of the discharge summary are associated with reduced risk of readmission.

2b2. EXCLUSIONS ANALYSIS

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

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 the impact of exclusions on the cohort, we examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion. 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*)

Table 4 shows the distribution of exclusions among hospitals with 25 or more admissions before applying exclusion criteria **using the EM Testing Dataset.**

Table 4. Distribution of exclusions among hospitals with 25 or more admissions (EM Testing Dataset)

Exclusion	N	%	Distribution across hospitals (N=2,161 before applying exclusion criteria: Min, 25th, 50th, 75th percentile, Max)
1. Without at least 30 days post-discharge enrollment in FFS Medicare for index admission	8,592	0.60	(0.00,0.00,0.29,0.83,9.56)
2. Discharged against medical advice (AMA)	9,569	0.67	(0.00,0.00,0.49,0.98,9.38)

3. HF admissions within 30 days of a prior HF admission	112,210	7.90	(0.00,5.60,7.27,8.86,19.2)
4. With a procedure code of LVAD implantation or heart transplantation either at the index admission or 12 months prior	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 without at least 30 days post-discharge enrollment in FFS Medicare for index admissions in non-VA hospitals) accounts for 0.60% of all index admissions excluded from the initial index cohort. This exclusion is needed since the 30-day EDAC outcome cannot be assessed in this group since claims data are used to determine whether a patient was readmitted.

Exclusion 2 (patients who are discharged AMA) accounts for 0.67% of all index admissions excluded from the initial index cohort. This exclusion is needed for acceptability of the measure to hospitals, who do not have the opportunity to deliver full care and prepare the patient for discharge. Given that a very small percentage of patients are being excluded, it is unlikely this exclusion affects the measure score.

For **Exclusion 3** (patients with admissions within 30 days of a prior index admission), if a patient has an admission within 30 days of discharge from the index admission, that admission is not included in the cohort so that admission can be both an index admission and readmission. This is consistent with the HF readmission measure. This exclusion accounts for 7.90% of all index admissions excluded from the initial index cohort.

Exclusion 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) accounts for 0.31% of all index admissions excluded from the initial index cohort. These patients are excluded because they represent a clinically distinct group.

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 2b5.

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

We used the final risk-adjustment variables in the current CMS 30-day HF readmission measure in order to harmonize with the existing EDAC measure. We verified the adequacy of this risk-adjustment strategy with the EDAC outcome by comparing the discrimination of the models with a full set of all comorbidities to the more parsimonious existing risk models. We found no meaningful improvement in model discrimination with the full set. The final model measure adjusts for variables (age, gender, comorbid diseases, and indicators of patient frailty) that are clinically relevant and have strong relationships with the outcome. For each patient, risk-adjustment variables are obtained from inpatient, outpatient, and physician Medicare administrative claims data extending 12 months prior to, and including, the index admission.

A summary of the methods used for selecting final risk-adjustment variables for the HF readmission measure (and by proxy the HF EDAC measure) is detailed below. The HF readmission measure development employed 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. The final ICD-10 code set is in the attached 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 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 presented here) 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)

- Cancers (CC 8-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 31 variables.

Social Risk Factors

We weigh social risk factor (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 conceptual pathways by which SRFs may influence risk of the outcome, as well as the statistical methods for SRF empiric testing. The conceptual framework 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 components that make up the EDAC outcome (ED visits, observation stays, and readmissions), few studies directly address causal pathways or examine the role of the hospital in these pathways (see, for example, Chang et al 2007; Gopaldas et al., 2009; Kim et al., 2007; LaPar et al., 2010; 2012; Lindenauer et al., 2013; Trivedi et al., 2014; Buntin et al., 2017; Hamadi et al., 2019). 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 EDAC.

The social risk factors that have been examined in the literature can be categorized into three domains: (1) patient-level variables, (2) neighborhood/community-level variables, and (3) hospital-level variables.

Patient-level variables describe characteristics of individual patients, and include the patient's income or education level (Eapen et al., 2015). Neighborhood/community-level variables use information from sources such as the American Community Survey as either a proxy for individual patient-level data or to measure environmental factors. Studies using these variables use one dimensional measures such as median household income or composite measures such as the AHRQ-validated SES index score (Blum et al., 2014). 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 EDAC 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 EDAC 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 EDAC risk is that patients may not receive equivalent care within a facility. For example, patients with SRFs 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 SRFs, such as income or wealth, may affect the likelihood of EDAC 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 conceptual 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 EDAC 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 EDAC (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 EDAC for patients of low SES were largely due to lower quality/higher EDAC 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 EDAC 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, for both of the two selected SRFs (low-SES and dual-eligibility), we decomposed the effect of a given SRF on the risk of EDAC as follows: Let denote a binary indicator of the SRF's status of patient i at hospital j, and X_j denote the percent of patients with the SRF at hospital j. Then we added X_{ij} into the original model adjusting for comorbidities only and broke down $X_{ij} = (X_{ij} - X_j) + X_j$, in which we let the first component,

(X_{ij} - X_j), represent the patient level social risk variable and the second component, X_j,, represent the hospital level social risk variable. By adding the SRF into the original risk-adjustment model and decomposing it into patient and hospital level variables, we can simultaneously estimate the SRF's within-hospital or patient level effect ($X_{patient}$) and between-hospital level effect ($X_{hospital}$) on the risk of EDAC; then we can assess, after controlling for the effects of comorbidities, whether the two levels of effects are independent and whether one level of effect contributes more than the other. The decomposition analysis allows us to calculate the effects of: 1) hospitals with higher or lower proportions of low-SES patients or patients dually-eligible for Medicare and Medicaid on the risk of EDAC for an average patient; and 2) patients' low-SES or dual-eligibility on their risk of EDAC when they are 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 EDAC based on the fitted model.

Specifically, to estimate the average hospital-level effect of a SRF, we calculated the predicted probabilities of EDAC for the following scenarios: (1) assuming all patients did not have the SRF (X_{ij} =0 for all i and j) and were seen at hospitals with a percent of patients with the SRF at the 5th percentile (P5) of the observed percent of patients with the SRF of all hospitals; (2) assuming all patients did not have the SRF and were seen at hospitals with a percent of patients with the SRF at the 95th percentile (P95); (3) assuming all patients did have the SRF (X_{ij} =1 for all i and j) and were seen at hospitals with a percent of patients with the SRF at the 5th percentile (P5); (4) assuming all patients did have the SRF and were seen at hospitals with a percent of patients with the SRF at the 95th percentile (P95). The estimated average hospital-level effect is calculated as ((2)-(1) + (4)-(3))/2 (denoted as P95-P5). Then, to estimate the average patient-level effect of a SRF, we calculated the predicted probabilities of EDAC for scenarios assuming all patients did or did not have the SRF (X_{ij} =0 or 1 for all i and j) and were seen at hospitals with the percent of patients with the SRF at nine selected percentiles (0th, 5th, 10th, 25th, 50th, 75th, 90th, 95th, and 100th). Then, we calculated the difference in predicted probabilities between patients with and without the risk factor who were seen at hospitals with the same percent of patients with the SRF at each of the nine percentiles (DELTAp, p=1, …, 9). We calculated the average of those differences in predicted probabilities as (DELTA1+…DELTA9)/9 (denoted as Delta) as the patient-level effect.

In summary, the difference in predicted probabilities of EDAC for an average patient between seen at hospitals with a percent of patients with the SRF at the 95th and 5th percentiles (P95-P5) of hospital percent of patients with the SRF estimates the hospital-level effect of the SRF on the risk of EDAC. We used the 5th and 95th percentiles rather than the maximum and minimum to avoid outlier values. The difference in predicted probabilities between patients with or without the SRF seen at an average hospital (Delta) estimates the patient-level effect of the SRF on the risk of EDAC. If P95-P5 is greater than Delta, it suggests that the hospital-level effect of the SRF is greater than the patient-level effect. That is, the hospital-level effect of the SRF contributes more than the patient-level effect on patients' risk of EDAC.

We also performed the same analysis for several clinical risk variables selected from the comorbidities included in the original risk-adjustment model to contrast the relative contributions of patient- and hospital-level effects of clinical risk variables to the relative contributions of the within and between-hospital level effects of SRFs on patients' risk of EDAC.

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/pdf-report/second-impact-report-to-congress. Accessed January 4, 2021.

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.

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

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. 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. 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.

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.

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

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

Lindenauer PK, Lagu T, Rothberg MB, et al. Income inequality and 30 day outcomes after acute myocardial infarction, heart failure, and pneumonia: retrospective cohort study. BMJ. 2013 Feb 14; 346:f521. doi: 10.1136/bmj.f521.

Mather JF, Fortunato GJ, Ash JL, et al. Prediction of pneumonia 30-day readmissions: a single-center attempt to increase model performance. Respir Care. 2014 Feb; 59(2):199-208. doi: 10.4187/respcare.02563. Epub 2013 Aug 13.

McHugh MD, Carthon JM, Kang XL. Medicare readmissions policies and racial and ethnic health disparities: a cautionary tale. Policy Polit Nurs Pract. 2010 Nov; 11(4):309-16. doi: 10.1177/1527154411398490.

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

Pandey A, Keshvani N, Khera R, et al. Temporal Trends in Racial Differences in 30-Day Readmission and Mortality Rates After Acute Myocardial Infarction Among Medicare Beneficiaries [published online ahead of print, 2020 Jan 8]. *JAMA Cardiol*. 2020;5(2):136-145.

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.

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

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

Spatz ES, Beckman AL, Wang Y, Desai NR, Krumholz HM. Geographic Variation in Trends and Disparities in Acute Myocardial Infarction Hospitalization and Mortality by Income Levels, 1999-2013. *JAMA Cardiol*. 2016;1(3):255-265. doi:10.1001/jamacardio.2016.0382.

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. 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 EM Testing Dataset with parameter estimates and 95 percent credible intervals (CI).

Table 5. HF EDAC Model: Median Parameter Estimates of Risk Variables from the Logit and Poisson Models (July 2016-June 2019)

Variable	Part 1: Logit Model		Part 2: Poisson Model	
	Median	CI	Median	CI
Age minus 65 (years above 65, continuous)	-0.001	(-0.001, -0.000)	-0.006	(-0.007, -0.006)
Male	0.026	(0.019, 0.035)	0.008	(0.005, 0.012)
History of coronary artery bypass graft (CABG) surgery	0.027	(0.019, 0.036)	0.005	(0.001, 0.009)
Metastatic cancer and acute leukemia (CC 8)	0.192	(0.167, 0.214)	0.051	(0.041, 0.060)
Cancer (CC 9-14)	0.027	(0.018, 0.037)	0.002	(-0.002, 0.005)
Diabetes mellitus (DM) or DM complications (CC 17-19, 122-123)	0.054	(0.046, 0.062)	0.020	(0.016, 0.023)
Protein-calorie malnutrition (CC 21)	0.113	(0.102, 0.125)	0.078	(0.074, 0.083)
Other significant endocrine and metabolic disorders; disorders of fluid/electrolyte/acid-base balance (CC 23-24)	0.117	(0.108, 0.125)	0.027	(0.023, 0.031)
Liver or biliary disease (CC 27-32)	0.093	(0.082, 0.104)	0.045	(0.041, 0.048)
Peptic ulcer, hemorrhage, other specified gastrointestinal disorders (CC 36)	0.055	(0.044, 0.067)	0.011	(0.007, 0.015)
Other gastrointestinal disorders (CC 38)	0.092	(0.083, 0.104)	-0.012	(-0.016, -0.009)
Severe hematological disorders (CC 46)	0.181	(0.159, 0.204)	0.055	(0.046, 0.063)
Iron deficiency or other/unspecified anemias and blood disease (CC 49)	0.098	(0.088, 0.106)	0.079	(0.075, 0.083)
Dementia or other specified brain disorders (CC 51-53)	0.062	(0.053, 0.071)	-0.015	(-0.019, -0.012)
Drug/alcohol abuse/dependence/psychosis (CC 54-56)	0.121	(0.111, 0.132)	-0.025	(-0.029, -0.021)
Major psychiatric disorders (CC 57-59)	0.054	(0.040, 0.067)	-0.000	(-0.005, 0.005)
Depression (CC 61)	0.023	(0.012, 0.033)	-0.021	(-0.024, -0.017)
Other psychiatric disorders (CC 63)	0.108	(0.099, 0.118)	-0.000	(-0.004, 0.003)
Hemiplegia, paraplegia, paralysis, functional disability (CC 70-74, 103-104, 189-190)	0.059	(0.046, 0.074)	0.024	(0.019, 0.030)

Variable	Part 1: Logit Model		Part 2: Poisson Model	
	Median	CI	Median	CI
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 diagnosis codes 799.01 and 799.02, for discharges prior to October 1, 2015)	0.071	(0.061, 0.080)	0.064	(0.061, 0.068)
Congestive heart failure (CC 85)	0.086	(0.076, 0.097)	0.019	(0.014, 0.023)
Acute coronary syndrome (CC 86-87)	0.123	(0.114, 0.133)	0.002	(-0.001, 0.005)
Coronary atherosclerosis or angina (CC 88-89)	0.062	(0.054, 0.072)	-0.017	(-0.021, -0.013)
Valvular and rheumatic heart disease (CC 91)	0.062	(0.055, 0.069)	0.039	(0.036, 0.042)
Specified arrhythmias and other heart rhythm disorders (CC 96-97)	0.065	(0.056, 0.075)	0.016	(0.012, 0.019)
Other and unspecified heart disease (CC 98)	0.054	(0.044, 0.062)	-0.001	(-0.004, 0.002)
Stroke (CC 99-100)	0.045	(0.031, 0.061)	-0.020	(-0.025, -0.015)
Vascular or circulatory disease (CC 106-109)	0.063	(0.053, 0.070)	0.015	(0.012, 0.018)
Chronic obstructive pulmonary disease (COPD) (CC 111)	0.109	(0.100, 0.116)	0.046	(0.043, 0.050)
Fibrosis of lung or other chronic lung disorders (CC 112)	0.067	(0.053, 0.081)	0.016	(0.011, 0.020)
Asthma (CC 113)	0.053	(0.043, 0.064)	-0.006	(-0.010, -0.001)
Pneumonia (CC 114-116)	0.062	(0.054, 0.070)	0.047	(0.044, 0.050)
Dialysis status (CC 134)	0.257	(0.240, 0.276)	-0.100	(-0.106, -0.094)
Renal failure (CC 135-140)	0.156	(0.147, 0.165)	0.120	(0.117, 0.124)
Nephritis (CC 141)	0.031	(0.020, 0.044)	0.011	(0.006, 0.015)
Other urinary tract disorders (CC 145)	0.072	(0.064, 0.080)	0.014	(0.010, 0.017)
Decubitus ulcer or chronic skin ulcer (CC 157-161)	0.073	(0.063, 0.084)	0.088	(0.084, 0.092)

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 SRF testing results based on the current EM Testing Dataset

Please note that for these analyses we encountered an issue with missing data in the ACS data. As described above in section 1.8, we created the ZIP-code-specific low-SES datafile based on the latest ACS data and obtained patients' low-SES based on their ZIP codes of residence in the CMS claims data. Patients' low-SES could be missing for two reasons: (1) patients' ZIP codes were missing from the claims data; or, (2) patients' ZIP codes were not present in the latest ACS data. Given that there was no ACS data available for the areas in U.S. territories, we found that the missing rates of patients' low-SES at hospitals in U.S. territories were extremely high (about 90% or above). Moreover, all patients with low-SES seen at hospitals in U.S. territories were residents of U.S. states and could not be representative of the population of those hospitals. Therefore,

we do not report the results for hospitals in U.S. territories for some the analyses with hospital-level results (i.e., variation in prevalence of low-SES, the change in EDAC after adding patients' low-SES for risk adjustment).

Prevalence of SRFs across hospitals

Table 6: Variation in prevalence of each social risk factor across measured entities

Social Risk Factor (SRF)	Prevalence of the SRF (IQR)
Dual Eligible	16.0% (9.60%-25.1%)
Low AHRQ SES	19.0% (6.60%-36.0%)

The prevalence of SRFs in the HF cohort varies across measured entities. The median percentage of dual-eligible patients was 16.0% (IQR 9.60%-25.1%) and the median percentage of patients with low AHRQ SES [an AHRQ SES index score adjusted for cost of living at the census block group level equal to or below 46.0 (lowest quartile)] was 19.0% (IQR 6.60%-36.0%).

Table 7: Comparison of observed days in acute care per 100 discharges in patients with and without social risk

Social Risk Factors	Mean Observed Days in Acute Care
Dual Eligibility (vs. Non-Dual)	175 (vs.144)
Low AHRQ SES (vs. SES score above 46.0)	164 (vs.145)

Patient-level mean observed days in acute care after HF admission are higher for dual-eligible patients compared with non-dual enrolled patients (175 vs. 144 mean acute days per 100 discharges). Mean observed days in acute care for patients with low AHRQ SES (index score equal to or below 46.0) are somewhat higher compared with patients without low AHRQ SES (164 vs. 145 mean acute days per 100 discharges).

Incremental effects of SRF variables in a multivariable model

Table 8: Parameter Estimates of social risk factor variables

Social risk factor	Adding Either SRF Individually		Adding Both SRFs Simultaneously	
	Logit model	Poisson model	Logit model	Poisson model
Low AHRQ SES	-0.006 (p=0.0015)	0.067 (p<0.0001)	-0.006 (p=0.0014)	0.061 (p<0.0001)
Dual Eligibility	0.001 (p=0.6920)	0.065 (p<0.0001)	-0.001 (p=0.5662)	0.053 (p<0.0001)

*p<.05

We also find that the c-statistics for the logit part of the hurdle model are almost unchanged with the addition of either or both of the SRFs into the model (Table 9).

Table 9. C-statistic for models with and without SRFs

HF EDAC Models	C-Statistic
Base Model (without social risk factors)	0.59
Base Model plus Low AHRQ SES	0.59
Base Model plus dual eligibility	0.59
Base Model plus Low AHRQ SES and dual eligibility	0.59

Impact on measure scores

We then examined the impact of adding each SRF separately on measure scores. We found that the addition of either SRF to the model has little to no effect on hospital performance, as measured by the median change in the score and by the correlation coefficient between measure scores, with and without the social risk factor (Table 10).

Table 10: Change in measure score and correlation coefficients comparing the base model with and without each social risk factor

Metric	Change in measure scores (per 100 discharges)	Change in measure scores (per 100 discharges)	Measure Score Correlation
Social Risk Factor	Median	IQR	Pearson Correlation Coefficient
Low AHRQ SES	0.50	-0.80 – 2.40	0.977
Dual Eligibility	-0.10	-0.30 – 0.10	0.999

The median change in hospitals' EDACs when adding the low AHRQ SES Index score indicator to the model is 0.50 EDAC per 100 discharges (interquartile range [IQR] -0.80 – 2.40 EDAC per 100 discharges) (Table 10). The correlation coefficient between EDACs for each hospital with and without the low AHRQ SES Index score indicator is 0.977.

The median change in hospitals' EDACs when adding a dual eligibility indicator to the existing model is -0.10 EDAC per 100 discharges (interquartile range [IQR] -0.30 – 0.10 EDAC per 100 discharges). The correlation coefficient between EDACs for each hospital with and without the dual eligibility indicator added is 0.999.

Contextual Effect Analysis

As described in 2b3.3a, we performed a decomposition analysis for each SRF variable to assess whether there was a corresponding contextual effect. 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.

Most of the patient-level and hospital-level effects of the dual eligible and low AHRQ SES variables were significant in the logistic and Poisson part of the HF EDAC hurdle model (Table 11). This indicates that both the patient- and hospital-level dual eligible effects of the SRFs are associated with an increased risk of acute care and expected duration of that care at the patient and hospital levels.

Because both the patient- and hospital-level effects contribute to an increased risk, if the dual eligibility and low-SES variables were added into the model to adjust for patient-level differences, then some of the differences in both risk of acute care and expected duration of care between hospitals would also be adjusted for, potentially obscuring a signal of hospital quality.

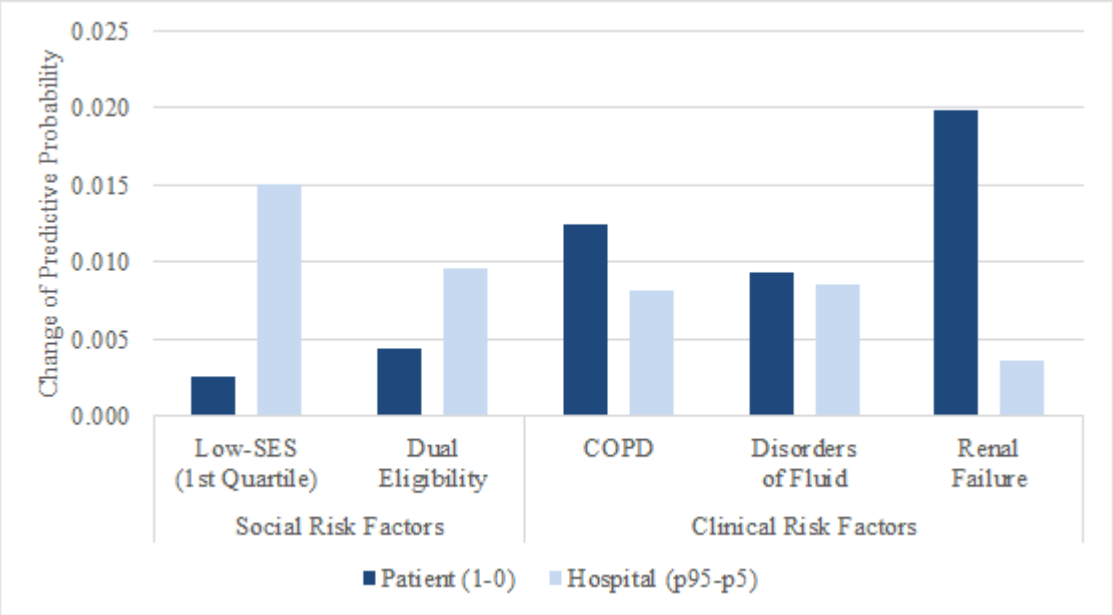
Table 11. Parameter Estimates for Hospital-Level and Patient-Level in 2020 from Decomposition Analysis

Parameter	Estimate (standard error), p-value	
	Logistic model	**Poisson model**
Low AHRQ SES – Patient Level	**-0.008 (0.002), p=0.0002**	**0.047 (0.005), p=<.0001**
Low AHRQ SES – Hospital Level	**0.068 (0.019), p=0.0003**	**0.335 (0.018), p=<.0001**
Dual-Eligible – Patient Level	**-0.001 (0.002) p=0.790**	**0.060 (0.006), p<.0001**
Dual-Eligible – Hospital Level	**0.185 (0.025), p<.0001**	**0.110 (0.025), p<.0001**
COPD – Patient-level	**0.046 (0.002), p<.0001**	**0.103 (0.004), p<.0001**
COPD – Hospital Level	-0.055 (0.032), p=.088	0.659 (0.032), p<.0001
Disorders of Fluid – Patient Level	0.027 (0.002), p<.0001	0.118 (0.005), p<.0001
Disorders of Fluid – Hospital Level	0.576 (0.041), p<.0001	0.003 (0.047), p=0.957
Renal Failure – Patient Level	0.120 (0.002),	0.159 (0.005),

Parameter	Estimate (standard error), p-value	
	Logistic model	Poisson model
Low AHRQ SES – Patient Level	**-0.008 (0.002), p=0.0002**	**0.047 (0.005), p=<.0001**
Low AHRQ SES – Hospital Level	**0.068 (0.019), p=0.0003**	**0.335 (0.018), p=<.0001**
Dual-Eligible – Patient Level	**-0.001 (0.002) p=0.790**	**0.060 (0.006), p<.0001**
Dual-Eligible – Hospital Level	**0.185 (0.025), p<.0001**	**0.110 (0.025), p<.0001**
COPD – Patient-level	**0.046 (0.002), p<.0001**	**0.103 (0.004), p<.0001**
	p<.0001	p<.0001
Renal Failure – Hospital Level	0.527 (0.036), p<.0001	-0.190 (0.041), p<.0001

However, as mentioned above in section 2b.3.3a, the patient-level and hospital-level coefficients shown in Table 11 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, to quantitatively compare the relative size of the patient and hospital effects, we calculated a range of predicted probabilities of EDAC based on the fitted model (Figure 4).

Figure 4. Decomposition analysis showing the patient-level and hospital-level effects for each social risk factor (HF EDAC)*



*These values are not comparable to Table 11 because the DE variable is binary and the AHRQ SES variable is continuous, therefore, to compare the two, we calculated a range of predicted probabilities of EDAC based on the fitted model; see Section 2b3.3a for details.)

As shown in Figure 4, as expected, the clinical risk factors shown for comparison have a larger patient-level effect compared to their hospital-level effects. In contrast, both the low AHRQ SES variable and the dual eligible variable have a larger hospital-level effect compared to the patient-level effect.

Social Risk Factor Adjustment Summary

The analyses above show that patients with either of two social risk factors (low AHRQ SES Index or dual eligibility) are at increased risk of EDAC, even after adjusting for other risk factors in a multivariable model. However, median changes in measure scores between the adjusted and unadjusted measures are small, and measure scores estimated for hospitals with and without either social risk factor are highly correlated.

Nevertheless, the residual risk suggests the need to consider whether to add the two variables as risk adjusters to the measure's risk model to ensure fairness to hospitals that care for such patients. As presented in the conceptual model (section 2b3.3a), the relationship may reflect that patients with social risk factors are receiving differential care within hospitals, that hospitals are missing opportunities to mitigate social risk factors they can address, that patients with these social risk factors disproportionately get care at lower-quality hospitals, or that patient factors that are difficult for hospitals to address are driving differences in the outcome. The extent to which each of these or other factors are contributing to the measured relationship is unclear, however empirically we found that for the HF EDAC measure, both the dual eligible and low AHRQ SES variables had a greater hospital-level component compared with the patient-level component.

CMS' decision regarding whether or not to adjust for social risk factors is based both on the empiric results (impact on model and measure scores), the conceptual model (for example, hospitals are better able to mitigate the influence of social risk factors on the measured outcome than clinicians) and the use of the measure (in a payment program or for public reporting). The HF EDAC measure is not in a payment program; the measure is used only in public reporting.

In making the decision about whether or not to risk adjust for these factors, CMS also considers the potential unintended consequence of adjusting, and the fairness to patients and hospitals that care for patients with social risk factors of the unadjusted measure score. If the relationship is driven by poorer quality, adjusting will mask the disparity in care. In contrast, an unadjusted measure will illuminate quality differences and create an

incentive to mitigate them. Not adjusting, however may disadvantage providers who care for low SES patients, and unintentionally create an incentive for hospitals to care for fewer patients with social risk factors, potentially reducing access care. CMS considers this risk limited, given the correlations between the measure scores calculated with and without social risk factors in the model.

In consideration of the benefits of a measure that can illuminate the potential disparities for beneficiaries with the two social risk factors and that there is little evidence of unintended consequences, CMS decided not to adjust this measure for either dual eligibility or the AHRQ SES Index. The decision to not adjust is also consistent with Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation's (ASPE's) recommendation that quality measures that are used for public reporting should not be risk adjusted (ASPE 2020).

Ongoing research aims to identify valid patient-level social risk factors and highlight disparities related to social risk. As additional variables become available, they will be considered for testing and inclusion within the measure. There are also alternative ways to account for social risk as part of measure program implementation. For the readmission measures (but not this measure) CMS confidentially reports disparities to hospitals so that they have more detailed, actionable information about their patient population's social risk.

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/pdf-report/second-impact-report-to-congress. Accessed January 4, 2021.

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 2b3.9

Dataset

The model selection process was performed using one half (the development sample) of the random three-year split sample.

Approach to Determining Model Specifications

Because the outcome, number of days in acute care, is novel not only for quality measurement but also in the literature as a measure of utilization, we considered a range of model specifications. We performed a number of analyses to determine the best model specification for the number of days in acute care. This is a pseudo-count variable (similar to a count variable, but taking half-integer values for half-days of acute care), and we therefore considered models that were generalized count models. All model development was performed using the development sample.

Inspection of the distribution of the outcome determined that the number of event days was highly skewed, with a large number of zeroes. Thus, we considered models appropriate for skewed data, including approaches that modeled the zero-day outcomes and non-zero day outcomes separately. We only considered approaches that allowed us to incorporate exposure time to account for differential risk.

First, using only patients with non-zero days, we estimated a generalized linear model (GLM) using a Poisson specification, and applied a Park test (Manning and Mullahy, 2001); the Park test indicated that Poisson was the best fit for our outcome. The Poisson model is commonly used for modeling count data and can be generalized to dependent variables that take non-integer values, such as ours.

We then considered three different model specifications for the full set of outcomes (zero and non-zero days): Poisson, zero-inflated Poisson (ZIP), and two-part logit/Poisson ("hurdle" model). For each model, we included an offset for the number of days the patient survived discharge, up to 30 (i.e., the exposure time). For the hurdle model, we included exposure time as an offset for each part because the Poisson part included only observations with non-zero days; it was technically a 'truncated' Poisson model.

For each of the three specifications listed above, we estimated (non-hierarchical) generalized linear models with days in acute care as the outcome. We compared the three different model specifications for the outcome using the following criteria: Akaike information criterion (AIC), Baysian information criterion (BIC), and log-likelihood.

Table 12

Criterion	Poisson	Zero-inflated Poisson	Two-part logit/Poisson
Akaike information criterion (AIC)	6,290,000	3,940,000	3,930,000
Bayesian information criterion (BIC)	6,290,000	3,940,000	3,930,000
Log-likelihood	-3,095,000	-1,970,000	-1,965,000

We selected the best model based on these statistics and judgment regarding the technical challenges of extending each to a random effects model for the measure. The AIC is a measure of the relative quality of statistical models for a given set of data. The best performing model was the two-part logit/ Poisson model, which had the smallest AIC. This model also made the most sense conceptually, with the likelihood of returning for acute care being modelled separately from the number of days of acute care received.

Assessing Model Discrimination and Calibration

Discrimination: We computed two different statistics – one for the logit part of the model and one for the Poisson part – using the development sample. For the logit model of zero versus non-zero days, which includes all patients in the cohort, we calculated the c-statistic. For the Poisson model of non-zero days, which includes only patients with some acute care, we calculated the deviance R^2. The deviance R2 is computed from the difference in the log-likelihoods between the final model and an empty model (no covariates) attributed to each observation, averaged over all observations (Cameron, Windmeijer, 1996).

Calibration Statistics

In a generalization of the calibration statistics for logistic models, we calculated the linear prediction Z = XB and W = XC using the coefficients B and C from the development sample and data X from the validation sample. We then estimated a model using the same functional form but only two independent variables, Z for the truncated Poisson part and W for the logit part. The intercepts and coefficients of Z and W in these second models are reported as ($\gamma 0$, γ 1) the calibration statistics for each part of the model. The closer they are to (0, 1), the better the model calibration (Harrell, 2013).

Calibration Plot

To further assess model calibration we constructed calibration plots with mean predicted and mean observed days in acute care plotted against decile of predicted utilization rate (predicted days/exposure days).

Approach to Annual Model Validation

CORE's measures undergo an annual measure reevaluation process, which ensures that the risk-standardized models are continually assessed and remain valid, given possible changes in clinical practice and coding standards over time. Modifications made to measure cohorts, risk models, and outcomes are informed by review of the most recent literature related to measure conditions or outcomes, feedback from various stakeholders, and empirical analyses, including assessment of coding trends that reveal shifts in clinical practice or billing patterns. Input is solicited from a workgroup composed of up to 20 clinical and measure experts, inclusive of internal and external consultants and subcontractors.

We provide a link to the 2020 measure re-evaluation report for this measure. The report describes what CORE did for 2020 public reporting, including:

- Updated the ICD-10 code-based specifications used in the measure. Specifically:
    - Incorporated the code changes that occurred in the FY 2019 version of the ICD-10-CM/PCS (effective with October 1, 2018+ discharges) into the cohort definitions and risk model;
    - Applied version 2019.1 (beta version) of the Agency for Healthcare Research and Quality (AHRQ) Clinical Classification Software (CCS) for ICD-10-CM/PCS to the planned readmission algorithm; and
    - Applied a modified version of the FY 2019 V22 CMS-Hierarchical Condition Category (HCC) crosswalk that is maintained by RTI International to the risk models.
- Monitored code frequencies to identify any warranted specification changes due to possible changes in coding practices and patterns;
- Reviewed potentially clinically relevant codes that "neighbor" existing codes used in the measures to identify any warranted specification changes;
- Reviewed select pre-existing ICD-10 code-based specifications with our workgroup to confirm the appropriateness of specifications unaffected by the updates;
- Evaluated and validated model performance for the three years combined (July 2016-June 2019); and
- Evaluated the stability of the risk-adjustment models over the three-year measurement period by examining the model variable frequencies, model coefficients, and the performance of the risk-adjustment model in each year (July 2016-June 2017, July 2017-June 2018, and July 2018-June 2019).

References

Cameron AC and Windmeijer FAG. R-Squared Measures for Count Data Regression Models with Applications to Health-Care Utilization. Journal of Business & Economic Statistics, Vol. 14, No. 2 (Apr., 1996), pp. 209-220.

Harrell FE. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. Springer New York; 2013.

Manning WG, Mullahy J. Estimating log models: to transform or not to transform? Journal of health economics. 2001;20(4):461-494.

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

Dataset

The model discrimination statistics were calculated using the development sample:

Discrimination Statistics (Development Dataset):

C-statistic for logit part of model: 0.587

Deviance R2 for truncated Poisson part of model: 0.026 (2.6%)

Updated Discrimination Statistics (EM Testing Dataset):

C-statistic (Logistic Model): 0.59

Deviance R^2 (Poisson Model): 0.027 (2.7%)

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

The model discrimination statistics were calculated using both the development and validation samples; see section 1.7.

Calibration Statistics (y0, y1):

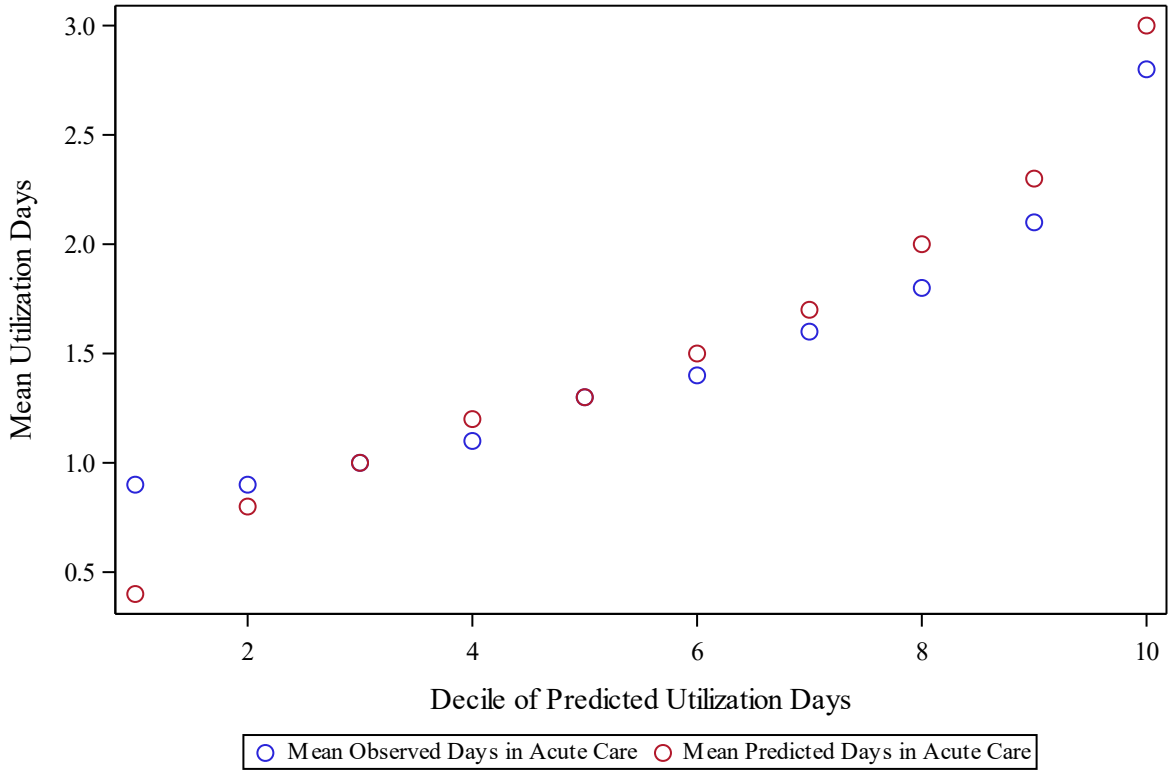
Logit part of model: (-0.10, 0.98)

Poisson part of model: (-0.04, 0.97)

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 (EM Testing Dataset).

Figure 5. Risk Decile Plot for the HF EDAC measure



2b3.9. Results of Risk Stratification Analysis:

N/A. This measure is not stratified.

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*)

The c-statistic of 0.59 indicates fair model discrimination, suggesting a moderate ability to distinguish patients with high risk from low risk of having at least one excess day in acute care. The deviance R^2 of 0.027 indicates that patients' clinical risk factors can explain 2.7% of the variation in the numbers of excess days in acute care. We report the deviance R^2 to present how much the Poisson model controls for differences in patients' comorbid characteristics. There is no rule of thumb for interpretation of the deviance R^2 for Poisson models with count outcome variables. However, Cameron et al.'s simulation study (1996) reported generally low deviance R^2 values (a maximum of 0.18) as predicting the numbers of inpatient days during a 12-month period. Given that the count outcome variable of the Poisson part of the EDAC hurdle model is the number of days in acute care within 30 days after discharge, a deviance R^2 much lower than 0.18 is expected. For the performance and calibration of the overall hurdle model, we provide risk-decline plots, described below and shown in Figure 5.

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). The strength of the evidence provided for endorsement maintenance (risk-decile plots and c-statistic) supports that the models remain valid for use with current data.

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 hospital-level 30-day all-cause EDAC measure score is estimated using a hurdle model with correlated random effects. This model consists of a logit model and a zero-truncated Poisson model for excess days in acute care and includes two correlated random effects for hospitals – one for the logit part and one for the zero-truncated Poisson part – with a non-zero covariance between the two random effects. This strategy accounts for within-hospital correlation of the observed outcome and accommodates the assumption that underlying differences in quality across hospitals lead to systematic differences in outcomes.

Specifically, the measure calculates EDAC, for each hospital, as the difference ("excess") between a hospital's "predicted days" and "expected days" per 100 discharges. "Predicted days" is the average number of days a hospital's patients spent in acute care after adjusting for the clinical risk factors. "Expected days" is the average number of risk-adjusted days in acute care a hospital's patients would have been expected to spend if discharged from an average performing hospital with the same case mix. To be consistent with the reporting of the CMS 30-day AMI, heart failure, and pneumonia readmission measures, measure scores are multiplied by 100 so that the final EDAC measures represent EDAC per 100 discharges.

We characterize the degree of variation by:

1) Reporting the distribution of the measure score, and

2) Reporting performance categories:

   To categorize hospital performance, the measure estimates each hospital's excess "days" in acute care and the corresponding 95% credible interval (CI). Excess "days" refers to the difference between the hospital's predicted days and expected days, per 100

   discharges. CMS assigns hospitals to a performance category by comparing each

   hospital's CI surrounding the hospital's excess "days" to zero. The reference to zero

   reflects the expectation that the hospital's "days" will be no different than an average performing hospital with a similar case mix. Comparative performance for hospitals with

   25 or more eligible cases is classified as follows:

   - "Fewer days than average" if the entire 95% CI surrounding the hospital's days is below zero. [Patients who are discharged from a hospital in this category spend fewer days in acute care than patients discharged from an average-performing hospital with a similar case mix.]
   - "Average" if the 95% CI surrounding the hospital's days includes zero. [Patients who are discharged from a hospital in this category spend about the same number of days in acute care after discharge as patients discharged from an average-performing hospital with a similar case mix.]
   - "More days than average" if the entire 95% CI surrounding the hospital's days is above zero. [Patients who are discharged from a hospital in this category spend more days in acute care than patients discharged from an average-performing hospital with a similar case mix.]

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

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)

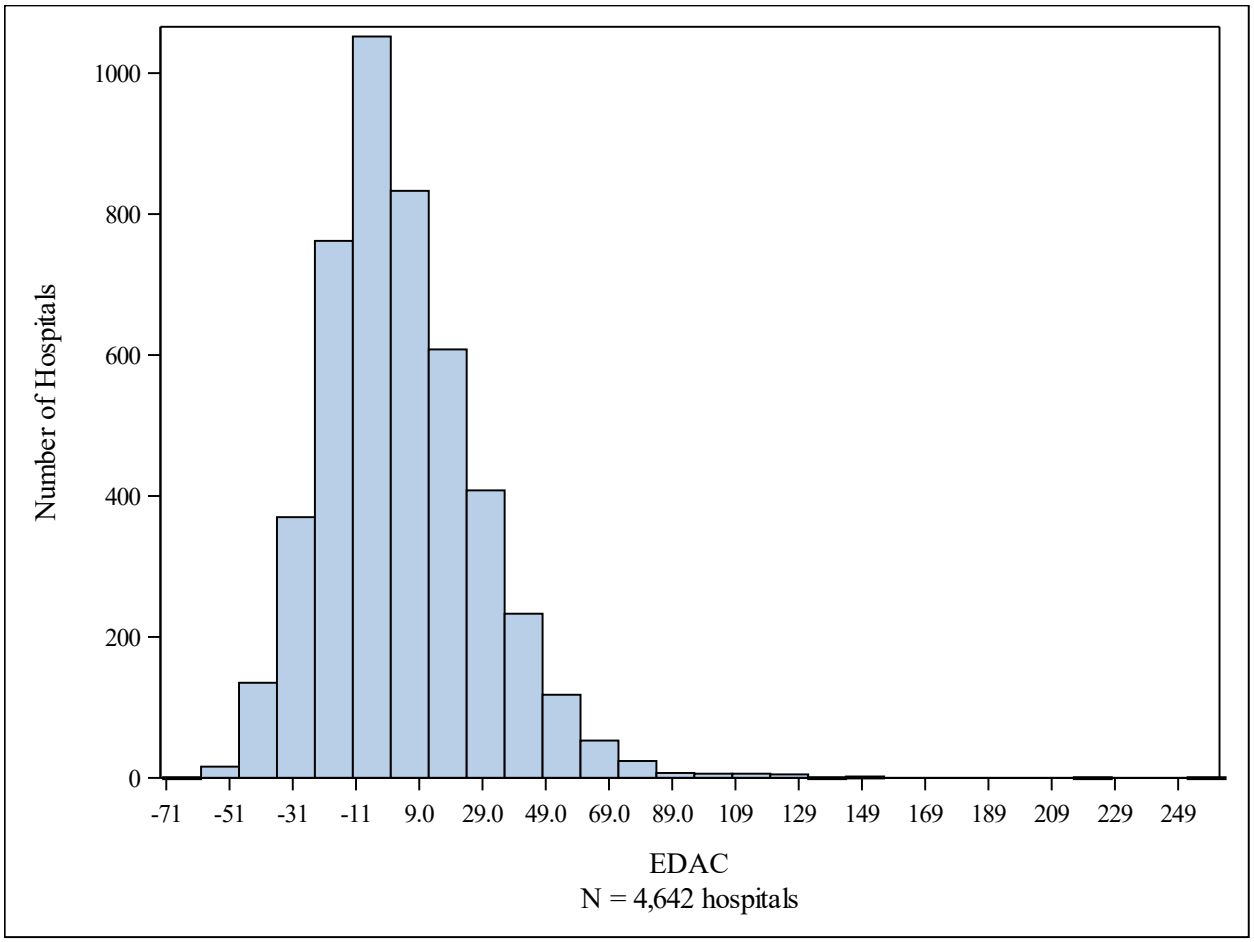
Analyses of Medicare FFS data show substantial variation in EDAC among hospitals.

The distribution of the measure score for all hospitals is shown below in Table 13 and Figure 6. Compared with average-performing hospitals, hospitals in the 10th percentile (better performance) have about 26 fewer excess days in acute care per 100 discharges, and hospitals in the 90th percentile (worse performance) have 35.6 more excess acute days in acute care per 100 discharges.

Table 13: Distribution of the HF EDAC measure score (excess days in acute care per 100 discharges)

Number of Hospitals	Number of Admissions	Mean (SD)	Min-Max	10th Percentile	25th Percentile	Median	75th Percentile	90th Percentile
4,642	1,286,352	3.2 (25.6)	-70.0 to 259.0	-26	-14	-0.3	17.3	35.6

Figure 6. Distribution (Histogram) Of Hospital-Level HF EDAC



Performance Categories

Out of 4,642 hospitals in the measure cohort, 447 had EDAC "fewer days than average," 2,467 were "average," and 799 had EDAC "more days than average." 929 were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing.

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 variation in hospital-level EDAC 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 nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

The HF EDAC measure used claims-based data for development and testing. There was no missing data in the claims-based 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 a combination of electronic sources

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

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

Attachment:

3c. Data Collection Strategy

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

3c.1. Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and

frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

IF 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)
Not in use	Public Reporting Care Compare https://www.medicare.gov/care-compare/ Payment Program CMS Hospital Inpatient Quality Reporting Program (IQR) https://qualitynet.cms.gov/inpatient/iqr

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: Care Compare, Centers for Medicare and Medicaid Services (CMS)
Purpose: Under Care 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 Care Compare website at: https://www.medicare.gov/care-compare/.
Payment Program

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

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

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

Geographic area and number and percentage of accountable entities and patients included:

The IQR program includes all participating non-federal acute care hospitals in the United States. The number and percentage of accountable hospitals included in the program, as well as the number of patients included in the measure, varies by reporting year.

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 July 1, 2016 – June 30, 2019, 4,642 hospital were included in measurement. All non-federal short-term acute care hospitals (including Indian Health Service hospitals), critical access hospitals, and VA hospitals were included in the measure calculation. However, only those hospitals with at least 25 HF admissions were included in public reporting.

Each hospital generally 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 public reporting websites in the summer 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 [total days] and type of post-discharge event). These reports facilitate quality improvement activities such as review of the number of days for each event and patterns of care; 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. CMS also provides measure frequency asked questions (FAQs), webinars, and provide a mechanism 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 programming code used to process the claims data and calculate measure results is written in Statistical Analysis System (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. FAQs: posted in April of each calendar year on QualityNet; includes general and measure-specific questions and responses, as well as infographics that explain complex components of the measure's methodology.

8. 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: provide a brief overview of measures and measure updates; 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 online portal. 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 EDAC measure, we have received the following inquiries from hospitals since the last endorsement maintenance cycle, regarding:

1. Clarification on the measure specifications including the methodology for calculating EDAC

2. The use of both physician and facility claims and use of the claim with the longer duration when both claims are present.

3. Requests for the SAS pack.

4. How to validate the measure score results they received in their preview report.

5. An admission that the hospital noted should have been excluded from the measure based on the current measure specifications.

6. How to interpret a negative measure score.

7. Why patients transferred to another facility that bills an elective procedure as an outpatient is still included in the denominator.

8. How to interpret the credible intervals that are generated with each point estimate, and the relationship to the performance categories.

9. . The overlap between the EDAC and Readmission measures

10. The difference between EDAC measures and HRRP

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

Summary of Question and Comments from Other Stakeholders:

For the HF EDAC measure, we have received the following feedback from other stakeholders since the last endorsement maintenance cycle:

1. Requests for the SAS code used to calculate measure results.

Summary of Relevant Publications from the Literature Review:

Since the last endorsement cycle, we have reviewed several articles related to EDAC following HF admissions. Relevant articles shared key themes related to: the impact of the Hospital Readmission Reduction Program (HRRP) on observation stays, which while relevant to the study, are nullified by CMS's shift to using a 'two midnight' definition of observation stays; the need for measuring EDAC, due to hospitals shifting return care away from inpatient admission to ED or observation stays; the relationship between readmission rates and rates of ED visits and observation stays; trends in readmission rates, ED visits, and observation stays. We provide more detail about these studies in the paragraph below.

We reviewed three studies that reinforce the importance of studying trends in ED visits and observation stays along with readmissions in order to get a complete picture of health care utilization. One study found an increase in overall returns to hospitals driven by an increase in observation stays and ED visits using three years of Medicare data [1]. This increase was greater than reduction in readmissions as a result of HRRP. Although no association between the reduced readmissions and increased observation stays or ED visits was determined, these findings imply a shift in return care from readmissions to ED or observation stays. In contrast, another study using HCUP data and hospital-level data from four states found a decreasing trend in overall hospital return rates driven by declining readmission rates with only a slight increase in observation stays and ED visits [2]. Yet another study found that while more than half of their patients returned to the hospital, only 25% were readmitted indicating a gap in measurement [3]. Our literature review also yielded two studies that provided

evidence that the EDAC measures and readmissions measures are correlated and that adding observation stays to the outcome will allow for more in-depth analysis of the association between hospital characteristics and the outcome [4,5].

REFERENCES

1. Wadhera RK, Joynt Maddox KE, Kazi DS, Shen C, Yeh RW. Hospital revisits within 30 days after discharge for medical conditions targeted by the Hospital Readmissions Reduction Program in the United States: national retrospective analysis. BMJ. 2019;366:l4563.

2. Nuckols TK, Fingar KR, Barrett ML, et al. Returns to Emergency Department, Observation, or Inpatient Care Within 30 Days After Hospitalization in 4 States, 2009 and 2010 Versus 2013 and 2014. J Hosp Med. 2018;13(5):296-303.

3. Shammas NW, Kelly R, Lemke J, et al. Assessment of Time to Hospital Encounter after an Initial Hospitalization for Heart Failure: Results from a Tertiary Medical Center. Cardiol Res Pract. 2018;2018:6087367.

4. Venkatesh AK, Wang C, Ross JS, et al. Hospital Use of Observation Stays: Cross-sectional Study of the Impact on Readmission Rates. Med Care. 2016;54(12):1070-1077.

5. Horwitz LI, Wang Y, Altaf FK, et al. Hospital Characteristics Associated With Postdischarge Hospital Readmission, Observation, and Emergency Department Utilization. Med Care. 2018;56(4):281-289.

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 Inpatient Prospective Payment System (IPPS) or other rule.

For example, based on stakeholder feedback, we revised the methodology used to count the number of observation stay days in the EDAC outcome. The use of both physician and facility claims (and use of the claim with the longer duration when both claims are present) was changed to use of physician claims only in cases when a facility claim is not available. This change, however, had minimal impact on measure results.

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.

To compare performance on the HF EDAC measures across performance periods, we show the distribution of measure scores for hospitals with at least 25 admission for Medicare FFS admissions only. We removed VA admissions as they only became part of the cohort during the most recent reporting period (2016-2019) and therefore we do not have trend information for VA admissions.

Our results show that over the past three reporting periods (from right to left, 2014-2017, 2015-2018, and 2016-2019) there has been improvement in measure scores across most of the distribution, from the 30th percentile through the 80th percentile.

Periods//YEAR1619//YEAR1518//YEAR1417

Number of Hospitals//3586//3643//3690

Number of Admissions//1219779//1188842//1159275

Mean(SD)//4.2(24.8)//4.3(24.8)//4.5(25.2)

Range(Min to Max)//-59.7 to 154.4//-66.6 to 143.2//-65 to 147.8

Minimum//-59.7//-66.6//-65.0

10th percentile//-25.4//-25.7//-25.2

20th percentile//-16.6//-17.2//-16.5

30th percentile//-10.1//-9.8//-9.8

40th percentile//-3.6//-3.4//-3.6

50th percentile//2.3//2.4//2.4

60th percentile//8.3//8.4//8.4

70th percentile//14.8//15.2//14.6

80th percentile//23.9//24.1//23.9

90th percentile//36.1//36.1//37.5

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.

We did not identify any unintended consequences during measure development or model testing. However, we are committed to monitoring this measure's use and assessing potential unintended consequences over time, such as the inappropriate shifting of care, increased patient morbidity and mortality, and other negative unintended consequences for patients.

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

N/A

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

0230 : Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute myocardial infarction (AMI) hospitalization

0330 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization

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

0506 : Hospital 30-day, All-Cause, Risk-Standardized Readmission Rate (RSRR) Following Pneumonia Hospitalization

1551 : Hospital-level 30-day risk-standardized readmission rate (RSRR) following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA)

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

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

2515 : Hospital 30-day, all-cause, unplanned, risk-standardized readmission rate (RSRR) following coronary artery bypass graft (CABG) surgery

2881 : Excess days in acute care (EDAC) after hospitalization for acute myocardial infarction (AMI)

2882 : Excess days in acute care (EDAC) after hospitalization for pneumonia

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

N/A

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

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

Are the measure specifications harmonized to the extent possible?

No

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

Measure harmonization: We developed the measure in the Medicare Fee-for-Service (FFS) population and completely harmonized the cohort definition and risk-adjustment strategy with those of the existing CMS 30-day HF readmission measure. Key differences: EDAC measures are based on the count of excess days spent in acute care whereas the readmission measures focus on the dichotomous presence of any readmission within the 30 days past discharge. In addition to readmission, the EDAC measure also counts observation stays and ED visits as acute care time. This difference in the outcome measure imposes differences on the statistical modeling and reporting format. The interpretations of the measures are also based on relative differences in excess days in acute care based on variations in case mix. There are no differences in data collection burden.

5b. Competing Measures

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

OR

Multiple measures are justified.

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

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

N/A

Appendix

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

Attachment **Attachment:** Heart_Failure_Excess_Days_in_Acute_Care_NQF_Appendix_01-29-16_v1.0.pdf

Contact Information

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

Co.2 Point of Contact: James, Poyer, James.Poyer@cms.hhs.gov, 410-786-2261-

Co.3 Measure Developer if different from Measure Steward: Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (CORE)

Co.4 Point of Contact: Jacqueline, Grady, jacqueline.grady@yale.edu, 203-764-5700-

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.

Yale New Haven Health Services Corporation/Center for Outcomes Research (YNHHSC/CORE) Measure Development Team Members

1. Faseeha K. Altaf, MPH- Lead Project Coordinator. Provided experience relevant to performance measurement.

2. Susannah Bernheim, MD, MHS- Project Director. Provided experience relevant to clinical content and performance measurement.

3. Nihar Desai, MD, MPH- Clinical Consultant. Provided experience relevant to clinical content and performance measurement.

4. Jacqueline Grady, MS- Supporting Analyst. Provided experience relevant to performance measurement.

5. Jeph Herrin, PhD- Statistician. Provided experience relevant to performance measurement.

6. Leora Horwitz, MD, MHS- Project Lead. Provided experience relevant to clinical content and performance measurement.

7. Zhenqiu Lin, PhD- Director of Analytics. Provided experience relevant to performance measurement.

8. Shuling Liu, PhD- Statistical Consultant. Provided experience relevant to performance measurement.

9. Chi Ngo, MPH- Research Associate. Provided experience relevant to performance measurement.

10. Arjun Venkatesh, MD, MBA- Clinical Consultant. Provided experience relevant to clinical content and performance measurement.

11. Changqin Wang, MD, MS - Lead Analyst. Provided experience relevant to performance measurement.

12. Yongfei Wang- Supporting Analyst. Provided experience relevant to performance measurement.

13. Sharon-Lise Normand, Ph.D.* - Statistical Consultant. Provided statistical expertise for the project.

*Harvard Medical School

Technical Expert Panel (TEP) Members

1. Anonymous Patient- Patient Representative. Provided patient perspective.

2. Kevin E. Driesen, PhD, MPH, MA- Assistant Professor, Mel and Enid Zuckerman College of Public Health; Director, Arizona Rural Hospital Flexibility Program. Provided experience relevant to performance measurement.

3. David Engler, PhD- Senior Vice President for Leadership and Innovation, America's Essential Hospitals. Provided experience relevant to clinical content, performance measurement, and coding and informatics.

4. Timothy Farrell, MD- Assistant Professor of Medicine, Adjunct Professor of Family Medicine, Physician Investigator; University of Utah School of Medicine. Provided experience relevant to clinical content and performance measurement.

5. Karen Farris, PhD- Charles R. Walgreen III Professor of Pharmacy Administration, Director of the Social and Administrative Pharmacy Graduate Program; University of Michigan College of Pharmacy. Provided experience relevant to performance measurement.

6. Maura C. Feldman, MSW- Director for Hospital Performance Measurement and Improvement, Blue Cross Blue Shield of Massachusetts. Provided consumer perspective.

7. Jay A. Gold, MD, JD, MPH- Senior Vice President and Chief Medical Officer, MetaStar. Provided experience relevant to clinical content and performance measurement.

8. Sally Hinkle, DNP, MPA, RN- Director of Performance Improvement and Clinical Value, Temple University Hospital. Provided experience relevant to performance measurement.

9. Amy Jo Haavisto Kind, MD, PhD - Assistant Professor of Geriatrics, University of Wisconsin School of Medicine and Public Health; Attending Physician, William S. Middleton VA. Provided experience relevant to clinical content and performance measurement.

10. Marjorie King, MD, FACC, MAACVPR- Director of Cardiac Services, Helen Hayes Hospital. Provided experience relevant to clinical content and performance measurement.

11. Eugene Kroch, PhD- Vice President and Chief Scientist, Premier. Provided experience relevant to performance measurement.

12. Keith D. Lind, JD, MS, BSN- Senior Policy Advisor, American Association of Retired Persons (AARP) Public Policy Institute. Provided consumer perspective.

13. Grace McConnell, PhD- Patient Representative. Provided patient perspective.

14. Michael A. Ross, MD, FACEP- Medical Director, Professor of Emergency Medicine; Emory University School of Medicine. Provided experience relevant to clinical content and performance measurement.

15. Mark Louis Sanz, MD- Interventional Cardiologist, International Heart Institute of Montana. Provided experience relevant to clinical content and performance measurement.

16. Paul Takahashi, MD- Associate Professor of Medicine, Mayo Clinic College of Medicine. Provided experience relevant to performance measurement.

Methodology Work Group Members

1. Arlene Ash, PhD- Professor and Division Chief, University of Massachusetts Medical School. Provided experience relevant to performance measurement.

2. Jeremiah Brown, PhD, MS- Assistant Professor of Health Policy and Clinical Practice, The Dartmouth Institute for Health Policy and Clinical Practice. Provided experience relevant to performance measurement.

4. Grant Ritter, PhD, MS, MA- Senior Scientist, Schneider Institute for Health Policy & Heller Graduate School. Provided experience relevant to performance measurement.

5. Patrick Romano, MD, MPH- Professor of Medicine and Pediatrics, University of California Davis School of Medicine. Provided experience relevant to performance measurement.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2015

Ad.3 Month and Year of most recent revision: 11, 2019

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? 2021

Ad.6 Copyright statement: N/A

Ad.7 Disclaimers: N/A

Ad.8 Additional Information/Comments: N/A