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 #: 2436

De.2. **Measure Title:** Hospital-level, risk-standardized payment associated with a 30-day episode-of-care for heart failure (HF)

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

De.3. **Brief Description of Measure:** This measure estimates hospital-level, risk-standardized payment for a HF episode of care starting with inpatient admission to a short term acute-care facility and extending 30 days post-admission for Medicare fee-for-service (FFS) patients who are 65 years of age or older with a principal discharge diagnosis of HF.

IM.1.1. **Developer Rationale:** In 2019, total Medicare expenditures were $799.4 billion [1], representing 3.6% of gross domestic product (GDP). Current estimates suggest that Medicare spending will grow 7.6% per year between 2019 and 2028 [1]. This growth in spending underscores the need to create incentives for high value care. Measuring costs in a way that is transparent to consumers and fair to providers is an important component of understanding and controlling costs of care and rewarding value. Measuring condition-specific costs of care is needed to identify high value care.

HF is a common condition in the elderly with a substantial range in payments due to different practice patterns, making it an ideal condition for assessing relative value for an episode of care that begins with an acute hospitalization. HF is one of the top three leading causes of hospitalization for Americans over 65 years old [2] and is projected to cost the US up to $70 billion in direct and indirect costs by 2030 [3].

In part due to increasing Medicare spending on HF care and geographic variation in resource use and spending, Centers for Medicare and Medicaid Services (CMS) payment programs have aimed to incentivize reductions in spending as determined by Medicare payments made to providers and institutions, as well as improved clinical outcomes for an episode of HF care.

Medicare payments are difficult to interpret in isolation. Some high payment hospitals may have better clinical outcomes when compared with low payment hospitals; other high payment hospitals may not. For this reason, the value of hospital care is more clearly assessed when pairing hospital payments with hospital quality. A measure of payments for Medicare patients during an episode of care for HF aligned with current quality of care measures will facilitate profiling hospital value (payments and quality). This measure, which uses standardized payments, reflects differences in the management of care for patients with HF both during hospitalization and immediately post-discharge. By focusing on one specific condition, value assessments may provide actionable feedback to hospitals and incentivize targeted improvements in care.

This measure is intended to align with current quality measures to facilitate profiling hospital value (payments and quality). Given that HF is a condition with substantial variability in costs of care, aligning this payment measure with quality measures (e.g., RSMRs) will allow the assessment of hospital value. By evaluating their RSPs and RSMRs for HF, hospitals have an opportunity to consider actionable improvements and efficiencies on a broader scale to impact value of care. This measure provides transparency on the payments made for Medicare beneficiaries who have had HF. Hospitals receive detailed information on how they compare with other institutions regarding the amount and venues of resources expended on patients. As such, the measure provides insight to hospitals that is not otherwise possible.

References:

1. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NHE-Fact-Sheet

2. https://www.hcup-us.ahrq.gov/faststats/NationalDiagnosesServlet

3. Heidenreich, P.A., Albert, N.M., Allen, L.A., Bluemke, D.A., Butler, J., Fonarow, G.C., Ikonomidis, J.S., Khavjou, O., Konstam, M.A., Maddox, T.M., Nichol, G., Pham, M., Piña, I.L., & Trogdon, J.G. (2013). Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. Circulation: Heart Failure, 6(3):606-619.

De.1. **Measure Type:** Cost/Resource Use

S.5. **Data Source:** Claims

Enrollment Data

S.3. **Level of Analysis:** Facility

IF Endorsement Maintenance – Original Endorsement Date: Nov 07, 2014 **Most Recent Endorsement Date:** Feb 10, 2015

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?**

Preliminary Analysis: Maintenance of Endorsement

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

Criteria 1: Importance to Measure and Report

1a. High impact or high resource use:

The measure focus addresses:

− a demonstrated high-impact aspect of healthcare (e.g., affects large numbers, leading cause of morbidity/mortality, high resource use [current and/or future], severity of illness, and patient/societal consequences of poor quality).

AND

1b. Opportunity for Improvement:

Demonstration of resource use or cost problems and opportunity for improvement, i.e., data demonstrating considerable variation cost or resource across providers

---

1a. High Impact or high resource use.

- The developer cites that heart failure (HF) is one of the top three leading causes of hospitalization for Americans over 65 years old and is projected to cost the US up to $70 billion in direct and indirect costs by 2030.
- This measure estimates hospital-level, risk-standardized payment for a HF episode of care starting with inpatient admission to a short term acute-care facility and extending 30 days post-admission for Medicare fee-for-service (FFS) patients who are 65 years of age or older with a principal discharge diagnosis of HF.

1b. Opportunity for Improvement:

- The developer reports a mean risk-standardized payment (RSP) of $17,722 with a range of $13,171 – $27,996 during the reporting period of July 1, 2016-June 30, 2019.
- The median hospital RSP in the combined three-year dataset was $17,607 (interquartile range of $16,817 – $18,513).
- The developer provided a distribution of hospital-level measure scores stratified by the proportion of patients with social risk factors (dual eligibility and low Agency of Healthcare Research and Quality [AHRQ] Socioeconomic Status [SES]).
- Measure scores do not vary significantly as a function of facilities' proportion of patients with social risk factors.

Questions for the Committee:

- *Has the developer demonstrated this is high impact, high-resource use area to measure?*
- *Is there a sufficient variation in performance across hospitals that warrants a national performance measure?*

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

Committee Pre-evaluation Comments:
Criteria 1: Importance to Measure and Report (including 1a, 1b)

1a. High Impact or High Resource Use: Has the developer adequately demonstrated that the measure focus addresses a high-impact aspect of healthcare (e.g., affects large numbers, leading cause of morbidity/mortality, high resource use [current and/or future], severity of illness, and patient/societal consequences of poor quality)?

- Yes
- Yes.
- Yes
- Yes, health care spending continues to be a significant proportion of the US GDP and variation in cost is important in understanding value. The developer notes heart failure is one of the top 3 leading causes of hospitalization for Americans over 65.
- Yes
- Yes

1b. Opportunity for improvement: Was current performance data on the measure provided? Has the developer demonstrated there is a resource use or cost problem and opportunity for improvement, i.e., data demonstrating, considerable variation in cost or resource use across providers?

- Yes
- The mean risk-adjusted payment (RSP) is $17K (range $13K to $28K) from July of 2016 through June of 2019. The median RSP is $18K with IQR of $17K to $19K. The latter seems to indicate that the measure is maturing.
- Yes
- The developer reports a mean risk-standardized payment (RSP) of $17,722 with a range of $13,171 – $27,996 during the reporting period of July 1, 2016-June 30, 2019. This seems like a limited amount of variation although it may be more meaningful in the aggregate because of the volume of heart failure admissions.
- Yes
- Yes

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

2b. Validity: Alignment of Specifications with Intent (includes threats to validity [e.g., attribution, costing method, missing data]) Testing; Exclusions; Risk-Adjustment; Meaningful Differences; Multiple Data Sources; and Disparities.

Measure evaluated by Scientific Methods Panel? ☒ **Yes** ☐ **No**

Evaluators: NQF Scientific Methods Panel

R: H-5, M-3, L-0, I-0

V: H-2, M-4, L-2, I-0

Measure evaluated by Technical Expert Panel? ☐ **Yes** ☒ **No**

Evaluators: N/A

Reliability

2a1. Specifications:

The measure is well defined and precisely specified so that it can be implemented consistently within and across organizations and allow for comparability. All measures that use the ICD classification system must use ICD-10-CM.

2a2. Reliability testing:

Demonstration that 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.

2a1. Specifications

- The SMP did not raise any concerns related to the measure specifications.
- This measure uses Medicare administrative claims. Standardized Medicare payment rates were assigned to each service based on claim type, facility type, and place of service codes, which were then summed by individual patients.

2a2. Reliability Testing:

- The developer conducted measure-score level reliability testing: calculating the intra-class correlation coefficient (ICC) using a split sample (i.e., test-retest) method for hospitals with 25 admissions or more.
- This reliability of a measurement is the degree to which repeated measurements of the same entity using non-overlapping random samples agree with each other.
- Using the Spearman-Brown prediction formula, the developer found that the agreement between the two independent assessments of the risk-standardized payment for each hospital was 0.781.
- The SMP did not raised any major concerns and passed the measure on reliability (H-5, M-3, L-0, I-0).

Questions for the Committee regarding reliability:

- *Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?*
- *Do you have any concerns with the reliability testing that was not identified by the Scientific Methods Panel?*

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

Committee Pre-evaluation Comments:
Criteria 2a: Reliability

2a1. Reliability-Specifications: Describe any additional concerns you have with the reliability of the specifications that were not raised by the Scientific Methods Panel: Describe any data elements that are not clearly defined: Describe any missing codes or descriptors: Describe any elements of the logic or calculation algorithm or other specifications (e.g., risk/case-mix adjustment, survey/sampling instructions) that are not clear: Describe any concerns you have about the likelihood that this measure can be consistently implemented:

- None
- No concern.
- only CORE can implement this score but it appears they can do so consistently
- No additional concerns with the reliability of the specifications or that the measure can be consistently implemented. Measure is specified for use with claims data.
- No concerns
- No concerns

2a2. Reliability-Testing: Has the developer demonstrated that 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? Describe any additional concerns you have with the reliability testing results or approach that were not raised by the Scientific Methods Panel.

- None
- None
- none
- The developer has demonstrated that the measure data elements are repeatable, producing the same results a high proportion of the time and precise enough to distinguish differences in performance.
- No concerns
- No concerns

Validity

2b1. Specifications align with measure intent:
The measure specifications are consistent with the measure intent and captures the most inclusive target population.
2b2. Validity Testing:
Demonstration that the measure data elements are correct and/or the measure score correctly reflects the cost of care or resources provided.
2b3. Exclusions:
Exclusions are supported by the clinical evidence**,** AND/OR There is a rationale or analysis demonstrating that the measure results are sufficiently distorted due to the magnitude and/or frequency of then on-clinical exclusions; AND Measure specifications for scoring include computing exclusions so that the effect on the measure is transparent (i.e., impact clearly delineated, such as number of cases excluded, exclusion rates by type of exclusion); AND If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be

specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

2b4. Risk Adjustment:

For resource use measures and other measures when indicated: an evidence-based risk-adjustment strategy is specified and is based on patient factors (including clinical and sociodemographic risk factors) that influence the measured outcome and are present at start of care, and has demonstrated adequate discrimination and calibration, OR rationale/data support no risk-adjustment/-stratification.

2b5. Meaningful Differences:

Data analysis demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/ clinically meaningful differences in performance.

2b6. Multiple Data Sources:

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

2c. Disparities: If disparities in care have been identified, measure specifications, scoring, and analysis allow for identification of disparities through stratification of results (e.g., by race, ethnicity, socioeconomic status, gender), OR rationale rationale/data justifies why stratification is not necessary or not feasible.

2b1. Specifications Align with Measure Intent:

- Attribution:
    - This measure attributes payments incurred during the 30-day episode to the original admitting hospital. The developer states that payments are assigned "to the admitting hospital because decisions made at the admitting hospital affect payments for care in the inpatient setting as well as the post-discharge and recovery periods for an HF."
- Costing approach:
    - This measure uses Medicare Standardized pricing.

2b2. Validity Testing:

- The SMP passed the measure on validity (H-2, M-4, L-2, I-0)
- Face Validity
    - The developer convened a 16-member Technical Expert Panel to assess the following: "This is a measure of payments for Medicare patients for a 30-day HF episode of care. The measure removes policy adjustments that are independent of care decisions and risk-adjusts based on case mix. The measure is intended to provide CMS a tool to compare payments across hospitals nationally to identify hospitals that have notably higher or lower payments associated with HF care. To what extent does the committee agree that this measure accomplishes this purpose?"
    - 8 of the 16 TEP members responded; 1 somewhat agreed, 3 moderately agreed, and 5 strongly agreed.
- Empirical Validity of Performance Measure Score
    - Measure compared to Medicare Spending per Beneficiary (MSPB) measure is a risk-adjusted, price-standardized measure that assesses Medicare Part A and Part B payments for services provided to Medicare beneficiaries for episodes that spanning from three days prior to an inpatient hospital admission through 30 days after discharge.
    - The developers found a correlation coefficient of 0.543 meaning that hospitals with higher spending across all Medicare FFS beneficiaries correlated with hospitals with higher spending on patients hospitalized with HF.

2b3. Clinical Inclusions and Exclusions/Evidence to Support Clinical Logic

- This measure applies 9 exclusions:
  - 1. Discharged alive on the day of admission or the following day who were not transferred to another
  - 2. Unreliable data
  - 3. Incomplete administrative data in the 30 days following the start of the index admission if discharged alive
  - 4. Enrolled in the Medicare hospice program any time in the 12 months prior to the index
  - 5. Discharged against medical advice (AMA)
  - 6. Transferred to a federal hospital
  - 7. not matched to admission in the HF mortality measure
  - 8. Missing index DRG weight where provider received no payment
  - 9. LVAD or transplant

2b4. Risk adjustment

- The hospital-level episode-of-care RSP for each measure is estimated using a hierarchical generalized linear model (HGLM).
- The developer used a team of clinicians to reviewed all 189 Condition Categories (CCs) and excluded those that were not relevant to the Medicare population or that were not clinically relevant to the HF payment outcome (e.g., attention deficit disorder, female infertility).
- Clinically relevant CCs were selected as candidate variables; some of these CCs were combined into clinically coherent groups. Other adjustment variables included age.
- To inform variable selection, the developer performed a modified approach to stepwise generalized linear model regression.
- The developer reviewed these results and decided to retain all risk-adjustment variables above a 90% cutoff (i.e., to retain variables that were significant at the p<0.05 level in at least 90% of the bootstrap samples).
- For this endorsement maintenance submission, the quasi-R2 slightly decreased to 0.031, suggesting that about three percent of the variation in payment could be explained by patient-level risk factors.
- The developer tested the impact of dual eligible status and the AHRQ SES index as social risk factors. The developer found that the two social risk factors did have slightly lower payment after adjustment for other risk factors in the multivariate model but the addition of these social risk factors had limited impact on model performance, little change in measure scores and measure scores estimated with hospitals with and without dual eligibility were highly correlated.

2b5: Meaningful Differences

- The developer notes that of the 4,502 hospitals in the study cohort, 409 had a payment "Less than the National Average Payment," 2,515 had a payment "No Different than the National Average Payment," and 542 had a payment "Greater than the National Average Payment."
- 1,036 were classified as "Number of Cases Too Small" (fewer than 25) to reliably estimate the hospital's RSP.

2b6. Multiple Data Sources

- N/A – this measure used Medicare administrative claims

2c. Disparities

- The developer provided a distribution of hospital-level measure scores stratified by the proportion of patients with social risk factors (dual eligibility and low Agency of Healthcare Research and Quality [AHRQ] Socioeconomic Status [SES]).

- Measure scores do not vary significantly as a function of facilities' proportion of patients with social risk factors.

Questions for the Committee regarding validity:

- *Do you have any concerns regarding the validity of the measure (e.g., correlations, exclusions, risk-adjustment approach, etc.)?*
- *Does the SC have any concerns related to the risk adjustment model (e.g., the r-squared values, lack of social risk factor adjustment)*

Staff preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 2b: Validity

2b1. Validity -Testing: Describe any concerns you have with the testing approach, results and/or the Scientific Methods Panel's evaluation of validity. Describe any concerns you have with the consistency of the measure specifications with the measure intent. Describe any concerns regarding the inclusiveness of the target population. Describe any concerns you have with the validity testing results: Does the testing adequately demonstrate that the measure data elements are correct and/or the measure score correctly reflects the cost of care or resources provided?

- Yes
- None
- yes
- No concerns with the testing approach or the SMP’s evaluation of validity.
- No concerns
- No concerns

2b2. Additional threats to validity: Describe any concerns of threats to validity related to attribution, the costing approach, or truncation (approach to outliers): Attribution: Does the accountable entity have reasonable control over the costs/resources measured? Is this approach aspirational (intending to drive change) or was it developed based on current state? Costing Approach: Do the cost categories selected align with the measure intent, target population and care settings? Is the approach for assigning dollars to resources agreeable? Truncation (approach to outliers): What is the threshold for outliers (i.e., extremely high cost or low cost cases) and are they handled appropriately?

- No concerns
- No concerns over any of the above-described validity issues.
- no new issues
- No concerns regarding the attribution or costing approach.
- No concerns
- No concerns

2b3. Additional Threats to Validity: Exclusions: Describe any concerns with the consistency exclusions with the measure intent and target population: Describe any concerns with inappropriate exclusion of any patients or patient groups:

- None
- None
- Credible argument that medicare advantage patients should be included to increase the validity of results
- No concerns regarding exclusions.
- No concerns
- No concerns

2b4/2c. Additional Threats to Validity: Risk Adjustment Is there a conceptual relationship between potential social risk factor variables and the measure focus? How well do social risk factors that were available and analyzed align with the conceptual description provided? Has the developer adequately described their rationale for adjusting or stratifying for social risk factors? Are all of the risk-adjustment variables present at the start of care (if not, do you agree with the rationale provided)? Describe any concerns with the appropriateness of risk adjustment (case-mix adjustment) development and testing: Do analyses indicate acceptable results?

- Yes. No concerns.

- Is there a conceptual relationship between potential social risk factor variables and the measure focus? Yes. How well do social risk factors that were available and analyzed align with the conceptual description provided? Dual eligibility and AHRQ SES index were tested and determined to have very minimal influence in the model performance. Has the developer adequately described their rationale for adjusting or stratifying for social risk factors? Yes. Are all of the risk-adjustment variables present at the start of care (if not, do you agree with the rationale provided)? Yes. Describe any concerns with the appropriateness of risk adjustment (case-mix adjustment) development and testing: None. Do analyses indicate acceptable results? Yes.
- I think we need a different approach for looking at social risk factors because the testing data do not match the reality that race and income do impact outcomes. Testing the addition of just one or two SES factors obscures the impact due to partial effects for other variables in the model (e.g. comorbidities). Stratifying results by SES would likely reveal more disparities, and testing SES in a different way.
- The developer tested dual eligible status and the AHRQ SES index as social risk factors. Ultimately, they were not included because they had limited impact on performance. I have no concerns regarding the appropriateness of risk adjustment.
- No concerns
- Yes

2b5. Threats to Validity: Meaningful Differences Describe any concerns with the analyses demonstrating meaningful differences among accountable units:

- None
- The developer states that the measure may not be reliable in approximately one-third of the hospitals because those hospitals have less than 25 cases.
- none
- The developer notes that of the 4,502 hospitals in the study cohort, 409 had a payment “Less than the National Average Payment,” 2,515 had a payment “No Different than the National Average Payment,” and 542 had a payment “Greater than the National Average Payment.” The developer reports a mean risk-standardized payment (RSP) of $17,722 with a range of $13,171 – $27,996 during the reporting period of July 1, 2016-June 30, 2019. Across the episode cost of care measures, I am wondering how we can assess whether above (or below) average costs are “appropriate” or not. Is there acceptable and normal amount of variability?
- No concerns
- No concerns

2b6. Threats to Validity: Missing Data/Carve Outs Describe any concerns you have with missing data that constitute a threat to the validity of this measure: Carve Outs: Has the developer adequately addressed how carve outs in the data source are handled (or should be handled for other users)? For example, if pharmacy data is carved out (missing) from the data set, can a measure that focuses on cost of care the target clinical population still be valid?

- No concerns
- No concerns.
- should look at the impact of excluding pharmacy data on ratings
- No concerns.
- No concerns
- No concerns

Criterion 3. Feasibility

3. Feasibility

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

- This measure uses administrative claims data.
- The developer indicates that all data elements for this measure are in defined fields in electronic claims.
- The developer also indicates that there a no fees associated with the use of this measure.

Questions for the Committee:

- *Are there any concerns regarding feasibility?*

Staff 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)? Describe your concerns about how the data collection strategy can be put into operational use: Describe any barriers to implementation such as data source/availability, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary tools (e.g., risk adjuster or grouper instrument):

- I do not see barriers to implementation. The measure uses administrative claims data and there is no fee associated with the use of this measure -
- No concerns as this measure is based on Medicare claims data, which is routinely collected.
- only CORE can implement this method
- The measure seems feasible to implement since it uses claims data.
- Feasibility rating high
- No concerns

Criterion 4: Usability and Use

Use

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

4a.1. Accountability and Transparency.

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

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

---

4a1. 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
  - https://www.medicare.gov/care-compare
- Payment Program
  - Hospital Inpatient Quality Reporting (IQR) Program
  - https://qualitynet.org/inpatient/iqr

4a2.Feedback on the measure by those being measured or others

- The developer reports that each hospital receives their respective 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 notes that since the measure is risk-standardized using data from all hospitals, hospitals cannot independently calculate their score.
- The accountable entities and other stakeholders can submit questions or comments the QualityNet online portal. The developer states that experts on measure specifications, calculation, or implementation, prepare responses to those inquiries and reply directly to the query.
- The developer has received inquiries from accountable entities regarding requests for the SAS code used to calculate measure results; questions about preview reports; questions about how to interpret the outcome and how performance categories are calculated; questions about the reporting period; and questions about measure specifications, including risk adjustment, outcome definition, and inclusion/exclusion criteria.
- The developer reports that there were no questions or issues raised by stakeholders requiring additional analysis or changes to the measure since the last endorsement maintenance cycle.

Additional Feedback:

- N/A

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?*

Staff preliminary rating for Use: ☒ Pass ☐ No Pass

Usability

4b. Usability.

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.

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).

4b1. Improvement results

- The developer reports a median hospital 30-day RSP of $17,607 for the HF payment measure for the 3-year period between July 1, 2016 and June 30, 2019.
- Median RSP decreased by 2.6% from July 2017-June 2018 (median RSP: $17,781) to July 2018-June 2019 (median RSP: $17,310).

4b2. Unintended consequences

- The developer did not identify any unintended consequences during measure development and test.

4b2.Potential harms

- The developer did not identify any potential harms.

Questions for the Committee:

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *What benefits, potential harms or unintended consequences should be considered?*

Staff 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? Is the measure being used in any other accountability applications? Are the performance results disclosed and available outside of the organizations or practices whose performance is measured? Is a credible plan for implementation provided?

- No concerns
- How is the measure being publicly reported? Yes, it is publicly reported in Care Compare. Is the measure being used in any other accountability applications? Yes, it is used for in Inpatient Quality Reporting program that impacts reimbursements to individual providers. Are the performance results disclosed and available outside of the organizations or practices whose performance is measured? Yes. Is a credible plan for implementation provided? Not applicable as the measure is already publicly reported.
- yes
- The developer reports the measure results annually on a public website.
- Yes
- No concerns

4a2. Use - Feedback on the measure: Describe any concerns with the feedback received or how it was adjudicated by the measure developer: 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?

- Developer demonstrates that they addressed the feedback provided.
- No concerns on the feedback from the measured entities (accountable care organizations) received by the measure developer and how the feedback is acted upon to improve the measure.
- after all these years of using the measure, I was disappointed not to see more data showing the correlation of this measure to quality, yet the assumption that less is always better.
- No concerns with feedback. The developer provides reports and resources on the measure. The public website also has a Q&A function.
- Yes
- No concerns

4b1. Usability – Improvement: Has the measure developer demonstrated that the use of this measure is helping to drive improvements in cost or efficiency? Has the developer adequately described how 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?

- No concerns
- The median 30-day RSP is stated to be $18K for HF payment measure between July of 2016 and June of 2019. Between this period, the median RSP has gone down by 2.6%.
- did not provide data that shows care is high-quality, just that cost declined
- The developer reported that the median RSP decreased by 2.6% from July 2017-June 2018 (median RSP: $17,781) to July 2018-June 2019 (median RSP: $17,310). Was the developer able to discern what was driving this change?
- Yes

- Is there additional analysis of the leading contributors to variability in episode cost? Table 2 on page 51 demonstrates the total cost differences between quartiles, as well as quartile differences for PAC, inpatient facilities and physician payments, with PAC contributing ~60% to the overall differences between quartiles. Is it possible that rolling up the measure to the entire episode costs masks the opportunities for improvement?

4b2. Usability – Benefits vs. harms: Describe any unintended consequences and note how you think the benefits of the measure outweigh them:

- None noted
- None
- none
- To what degree do we think the measure is driving change and how do hospitals understand how to benchmark themselves. Is it more informative than motivational? I wonder if the burden of producing the measure outweigh the benefits.
- No potential harms identified
- N/A

Criterion 5: Related and Competing Measures

- The developers identified the following related measures:
    - 0229 : Hospital 30-Day, All-Cause, Risk-Standardized Mortality Rate (RSMR) Following Heart Failure (HF) Hospitalization
    - 0330 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization
    - 2158 : Medicare Spending Per Beneficiary (MSPB) Hospital
    - 2431 : Hospital-level, risk-standardized payment associated with a 30-day episode-of-care for Acute Myocardial Infarction (AMI)
    - 2579 : Hospital-level, risk-standardized payment associated with a 30-day episode of care for pneumonia (PN)
    - 3474 : Hospital-level, risk-standardized payment associated with a 90-day episode of care for elective primary total hip and/or total knee arthroplasty (THA/TKA)

Harmonization

- The developer indicates that these measures are harmonized to the extent possible.

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

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

- None noted
- "Yes, the measure developer enlists the following measures as related, and they all have been harmonized to the extent possible.: • 0229 : Hospital 30-Day, All-Cause, Risk-Standardized Mortality Rate (RSMR) Following Heart Failure (HF) Hospitalization • 0330 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization • 2158 : Medicare Spending Per Beneficiary (MSPB) Hospital • 2431 : Hospital-level, risk-standardized payment associated with a 30-day episode-of-care for Acute Myocardial Infarction (AMI) • 2579 : Hospital-level, risk-standardized payment associated with a 30-day episode of care for pneumonia (PN) • 3474 : Hospital-level, risk-standardized payment associated with a 90-day episode of care for elective primary total hip and/or total knee arthroplasty (THA/TKA) "
- none
- There are related measures – heart failure mortality and readmissions, MSPB, episode-of-care payment measures for other conditions (AMI, pneumonia, total hip/total knee arthroplasty) – and they are harmonized to the extent possible.
- None
- Given the attention to 30-day readmission rates, how much do readmissions contribute to overall episode cost variability?

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 06/17/2021

- **No NQF members have submitted support/non-support choice as of this date.**

Combined Scientific Methods Panel Preliminary Analysis of Scientific Acceptability

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 2436

Measure Title: Hospital-level, risk-standardized payment associated with a 30-day episode-of-care for heart failure (HF)

RELIABILITY: SPECIFICATIONS

1. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?** ☒ **Yes** ☐ **No**

   Submission document: "MIF_xxxx" document, items S.1-S.22

   NOTE*: NQF staff will conduct a separate, more technical, check of eCQM specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.*

2. **Briefly summarize any concerns about the measure specifications.**

   Panel Member 1: None

Panel Member 2: None

Panel Member 4: No concerns

Panel Member 5: None

Panel Member 6: NA

Panel Member 8: None

RELIABILITY: TESTING

Type of measure:

☐ **Outcome (including PRO-PM)** ☐ **Intermediate Clinical Outcome** ☐ **Process**

☐ **Structure** ☐ **Composite** ☒ **Cost/Resource Use** ☐ **Efficiency**

Data Source:

☐ **Abstracted from Paper Records** ☒ **Claims** ☐ **Registry**
☐ **Abstracted from Electronic Health Record (EHR)** ☐ **eMeasure (HQMF) implemented in EHRs**
☐ **Instrument-Based Data** ☒ **Enrollment Data** ☒ **Other (please specify)**

Panel Member 1: Census Data/American Community Survey, Medicare Fee Schedules, CMS Wage Index Data.
Panel Member 2: ACS, CMS administrative data sets
Panel Member 3: Medicare Enrollment Database (including the Master Beneficiary Summary File), Medicare Fee Schedules, CMS Wage Index Data.
Panel Member 5: Medicare Enrollment Database (including the Master Beneficiary Summary File), Medicare Fee Schedules, CMS Wage Index Data.
Panel Member 6: Census Data/American Community Survey, Medicare Enrollment Database (including the Master Beneficiary Summary File), Medicare Fee Schedules, CMS Wage Index Data.)
Panel Member 8: Census Data/American Community Survey, Medicare Enrollment Database (including the Master Beneficiary Summary File), Medicare Fee Schedules, CMS Wage Index Data.)
Panel Member 9: Census Data/American Community Survey, Medicare Enrollment Database (including the Master Beneficiary Summary File), Medicare Fee Schedules, CMS Wage Index Data.)

Level of Analysis:

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

Measure is:

☐ **New** ☒ **Previously endorsed (**NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.)

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

3. **Reliability testing level** ☒ **Measure score** ☐ **Data element** ☐ **Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure**
   ☒ **Yes** ☐ **No**
5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of **patient-level data** conducted?

☐ **Yes** ☒ **No**

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

   Submission document: Testing attachment, section 2a2.2

   Panel Member 1: Overall measure score reliability was assessed by calculating the intra-class correlation coefficient (ICC) using a split sample (i.e., test-retest) method.

   Panel Member 2: Split sample is a standard approach to assessing reliability in these measures.

   Panel Member 4: The method used by the developer is reasonable -- they estimate reliability by calculating an ICC using a split sample approach to conduct test-retest measure score reliability.

   Panel Member 5: ICC using a split sample, as a reasonable approach

   Panel Member 6: ICC split sample

   Panel Member 8: Split sample approach makes sense, particularly since the volume of HF admissions is so high.

   Panel Member 9: To measure the reliability of the measure developer calculated the intra-class correlation coefficient (ICC) using a split sample (i.e., test-retest) method. Hospital performance is measured using a random subset of patients from a measurement period, and then measured again using a second random subset exclusive of the first from the same measurement period, and then comparing the two measures. The extent to which the measures of these two subsets agree provide evidence that the measure is assessing an attribute of the hospital, not of the patients.

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

   Submission document: Testing attachment, section 2a2.3

   Panel Member 1: ICC = 0.781

   Panel Member 2: Split sample correlation of 0.781 indicates adequate reliability

   Panel Member 4: Across 4,502 hospitals, the agreement between RSP for each hospital was 0.781, indicating sufficient test-retest reliability

   Panel Member 5: Using the Spearman-Brown prediction formula, the agreement between the two independent assessments of the risk-standardized payment (RSP) for each hospital was 0.781.

   Panel Member 6: ICC(2,1) = 0.781. This is above to the committee's 0.70 cutoff.

   Panel Member 8: The authors find a split sample reliability score of 0.781 - this is better than the AMI measure. Looks very good.

   Panel Member 9: As a metric of agreement, they calculated the ICC for hospitals with 25 admissions or more. Using the Spearman-Brown prediction formula, the agreement between the two independent assessments of the risk-standardized payment (RSP) for each hospital was 0.781. This shows good reliability.

8. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? NOTE: If multiple methods used, at least one must be appropriate.

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

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

9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?

   Submission document: Testing attachment, section 2a2.2

☒ **Yes**

☒ **No**

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

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

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

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

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

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

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

Panel Member 1: The estimated reliability of the measure based on split sample methodology is 0.78 and is considered high.

Panel Member 2: Split sample ICC 0.781

Panel Member 4: Although the ICC was strong, test-retest is not as strong as other methods to estimate facility level precision or lack of measurement noise.

Panel Member 5: Reasonable approach with the score

Panel Member 6: Split half results were good.

Panel Member 8: This measure is high-to-moderate in terms of reliability give strong results from the split sample analysis.

Panel Member 9: The ICC based on a sample of more than 400K admissions over 3 years with more than 3,700 hospitals showed relatively strong repeatability 0.78 correlation.

VALIDITY: TESTING

12. **Validity testing level:** ☒ **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:**

☒ **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: Face Validity as Determined by Technical Expert Panel (TEP) comprising of 16 members, including patient representatives, expert clinicians, researchers, providers, and purchasers. Measure Score Validity was sought to be obtained through established measure development guidelines including NQF guidance for outcomes measures (National Quality Forum, 2010), CMS Measure Management System (MMS) guidance, and the guidance articulated in the American Heart Association scientific statement, “Standards for Statistical Models Used for Public Reporting of Health Outcomes” In addition, measure score validity was also assessed by external Groups that included expert and stakeholder input via three mechanisms: regular consultations with an expert health economist, a national TEP, and a 30-day public comment period in order to increase transparency and to gain broader input into the measure.

    Panel Member 2: Face validity. Comparison of measure and hospital cost per Medicare Beneficiary by quartile

    Panel Member 4: The developer used 2 approaches to assess validity. TEP members assessed face validity and the developer looked at the correlation between the HF payment measure and the MSPB (Medicare parts A and B) for Medicare benes.

    Panel Member 5: Demonstrated internal and external validity using the score

    Panel Member 6: Tested correlation with MSPB - MSPB is an all - claim measures. It is not clear that there is value in comparing the HF-specific measure to MSPB.

    Panel Member 8: TEP review and criterion related validity make sense. Since there is no established gold standard, the authors use MSPB, the same measure they used for AMI. HF admissions are expensive, but the conceptual link seems weak.

    Panel Member 9: (NOTE: the developer did not check Face validity but the evaluation is included) Development of the measure was guided by an expert consultant and a TEP. One approach to validity of the measure was face validity assessed by a TEP of 16 members. To measure empirical validity, they identified and assessed the measure’s correlation with other measures that target the same domain for similar populations, including the hospital Medicare Spending per Beneficiary (MSPB). They used the unweighted Pearson’s correlation coefficient. Because the MSPB measure assesses payments for all Medicare FFS patients for all conditions during the measurement period, and the HF payment measure is focused on a single diagnosis, they predicted that HF payment measure scores would be weakly-to-moderately, positively correlated with MSPB measure scores. As a final test of measure score validity, they present a measure of internal validity of the outcome by examining the distribution of payment types across the quartiles of risk-standardized payments. They hypothesized that detailed level observed payments would be greater in hospitals with higher risk-standardized payments.

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

    Submission document: Testing attachment, section 2b2.3

    Panel Member 1: Face validity: Among the 8 TEP members who provided a response, 1 responded “Somewhat Agree,” 3 responded “Moderately Agree,” and 4 reported “Strongly Agree” that this measure accomplished the purposes of measuring payments for Medicare patients for a 30-day HF episode of care Empiric validity: Validity is supported by the correlation in the expected strength and direction with a related and valid payment measure. In addition, the observed payment breakdowns appropriately align with the distribution of the provider-level risk-standardized payments.

Panel Member 2: Validity of measure appears adequate

Panel Member 4: I did not consider the face validity results as they are less useful and appropriate for re-endorsement. There was a decent correlation (.54) between the HF RSP and MSPB score. Across quartiles of RSPs, mean total payments per patient went up monotonically in the expected direction.

Panel Member 5: The validity of the HF Payment measure is supported by three types of evidence: face validity results derived from a systematic survey of a Technical Expert Panel (TEP), empiric validity demonstrated by correlations, and internal consistency. The validity of the HF Payment measure is supported by face validity as indicated by the Technical Expert Panel (TEP) vote. There was unanimous TEP support for the face validity of the measure: 8 of 8 TEP members strongly, mostly, or somewhat agreed with the validity statement. The validity of the measure is further supported by the empiric evidence that shows a correlation in the expected strength and direction with a related and valid payment measure. Finally, the observed payment breakdowns appropriately align with the distribution of the provider-level risk-standardized payments.

Panel Member 6: Correlation is higher here than other cost metrics submitter, but unclear if this is just by chance.

Panel Member 8: Better correlation between MSPB and HF episode (0.543) than we saw on the AMI measure; general support from TEP.

Panel Member 9: Only 8 (half) of 16 TEP members responded to the face validity question; of those, 1 responded somewhat agree, 3 moderately agreed and 4 strongly agreed that the measure accomplished its purpose to enable CMS to identify hospitals with notably higher and lower payments for HF episodes. The correlation of the HF payment measure with the MSPB measure was higher than a similar measure (AMI) at 0.543 indicating higher spending on Medicare FFS beneficiaries is associated with higher spending on HF patients. Finally, the observed payment breakdowns aligned with the distribution of the provider-level risk-standardized payments.

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: Same concerns about the exclusion criteria #1 and #3 as were for the measure #2431 (AMI measure)

Panel Member 2: None

Panel Member 4: No concerns

Panel Member 5: None.

Panel Member 6: NA

Panel Member 8: No concerns

Panel Member 9: Exclusions are few and seem appropriate.

19. **Risk Adjustment**

Submission Document: Testing attachment, section 2b3

19a. **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☒ **Stratification**

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: "We used the ICD-9-to-CC assignment map, which is maintained by CMS and posted at www.qualitynet.org." (see 2b3.3a.) I am curious about the fact that most hospitals have now moved onto ICD-10s. If that is the case, then how will providers generate CCs? Converting ICD10 to ICD9 to CC? If so, won't it impose a layer of implementation difficulty? Otherwise, the rationale for the risk-adjustment model, including the variable selection, is well-described.

Panel Member 2: Measure uses a list of comorbidities and can include multiple comorbidities rather than HCC highest approach. Quasi-R-square of measure is extremely low, around 0.03. Need to have substantive experts on HF assess whether factors known to affect costs of post-acute care are adequately represented in the risk adjustment model.

Panel Member 4: While the overall R-squared is low, it appears to be similar to other cost measures. Some concerns about not using Dual eligible status and AHRQ SES index as SRF.

Panel Member 5: Reasonable approach.

Panel Member 6: The risk adjustment approach is similar to previous submission - they updated model for 2019 data.

Panel Member 8: Very comprehensive analysis on social risk factors, parsimonious model, adequate performance

Panel Member 9: Developers started with 189 condition categories and used a team of clinicians to exclude those not relevant to the Medicare population or the HF payment outcome. Other adjustment variables included age, history of PCI and history of CABG. They performed a modified approach to stepwise GLM regression. First, created 1,000 bootstrap samples and ran a GLM including all candidate variables for each sample and summarized results to show which variables was significantly associated with HF payment in each sample. They retained all risk adjustment variables that were significant at P<0.05 level in at least 90% of bootstrap samples. The developers noted that potential causal mechanisms by which social risk factors influence costs following HF are varied and complex. Although studies have assessed the relationship between patient social risk factors (e.g., gender, SES and race) and payment associated with an HF, few studies directly address the complex causal pathways. A literature review identified four potential mechanisms at the patient- and hospital-level: (1) Health at admission and other patient characteristics: patients with social risk factors such as low SES may have more comorbid conditions at the time of admission related to historical or lifelong social disadvantage, but developers contend this was controlled for with the adjustment for comorbidities to account for health at admission; (2) selection of patients into different quality hospitals: patients with social risk factors may be more likely to live near to and be admitted to lower quality hospitals, but developers pointed to a recent study that found between hospital differences in readmission rates were small at hospitals treating a minimum volume of patients within different race/income groups; (3) care within the hospital: social risk factors can contribute to costs if patients do not receive equivalent or patient-centered care within a facility. For example, a study using linked hospital and census data found that low income or minority patients may experience differential, lower quality, or discriminatory care within a given facility; and (4) post-discharge care: social risk factors can contribute to costs if patients receive or have access to more or less high-value post-discharge care (e.g., HF patients with social risk factors were less likely to have a follow up visit and more likely to have a readmission or ED visit within 30 days of discharge and HF patients with lower education were less likely to participate in cardiac rehabilitation). Impact of SES factors were measured AFTER adjusting for demographic and clinical factors. Mean observed payments are $930 higher for dual-eligible patients compared with non-dual enrolled patients ($18,440 vs. $17,511) and mean observed payments were $96 lower for patients with low AHRQ SES compared with patients without low AHRQ SES Index. This shows that for hospitals serving a large percent of duals, NOT ADJUSTING for dual status could significantly increase their costs, potentially penalizing facilities that care for vulnerable patients. They also examined the payment ratio when adding each of the social risk factors to the multivariate model. The payment ratio for the low AHRQ SES variable was 0.98 when added to the model with the clinical risk factors; the payment ratio for the dual eligibility variable was 1.01. When both variables are added to the model, the payment ratio for the low AHRQ SES variable was 0.98 and payment ratio for the dual eligibility variable is unchanged. Developers also examined the impact of adding each social risk factor separately on measure scores and found that when adding the low AHRQ SES variable to the model, the median change in measure scores (risk-standardized payments or RSPs) was very small, and in a negative direction: -$7.55 (interquartile range [IQR] (-$97.80 – $77.50). When the dual eligibility variable was added to the model the median change in hospitals' RSPs was also small: $1.30 (interquartile range [IQR] (-$2.50 – $5.00). They also examined the correlation between measure scores (risk-standardized payments) calculated with the baseline model and with either social risk factor included in the model and found the measures scores were highly correlated: the correlation coefficient between RSPs for each hospital with and without the low AHRQ SES variable is 0.989; the correlation coefficient between RSPs for each hospital with and without the dual eligibility variable is >0.999. They conclude that these results demonstrate that overall, risk adjustment for either social risk variable has a small impact on measure scores. In summary, 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, and the use of the measure (in a payment program or for public reporting). The HF payment measure is not in a payment program; the measure is used only in public reporting. In addition, the Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation's (ASPE's) recommends that resource use measures that are used for public reporting should not be risk adjusted for social risk (ASPE 2020). DECISION: not adjust for the social risk factors. FOR CONSIDERATION: A recent study reported continued disparities in the 3 hospital readmission measures and suggests that there ARE differences in performance across racial/ethnic groups, contrary to what the developers report. RACIAL AND ETHNIC DISPARITIES IN 30-DAY READMISSION RATES IN MEDICARE ADVANTAGE AND MEDICARE FEE-FOR-SERVICE, Innovation in Aging, 2018, Vol.,2, No S1, M. Rivera-Hernandez, O. Panagiotou, M. Rahman, A. Kumar, V. Mor, A. Trivedi. It is unclear whether improvements in readmission rates have accompanied improvements for all White, African-American and Hispanic beneficiaries, and whether differences among these groups have narrowed over time. Using national Medicare Provider and Analysis Review Files linked with the Medicare Advantage Healthcare Effectiveness Data and Information Set, we examined 30-day all-cause readmission for three conditions: acute myocardial infraction (AMI), heart failure (HF), and pneumonia from 2011 and 2014 by race and ethnic group. We calculated raw differences and also used generalized linear models adjusting for demographic characteristics and up to 39 co-morbidities. In 2011, unadjusted rates were 18.9% (African-American) vs. 17.1% (Hispanic) vs. 15.2% (white) for AMI; 24.6% vs. 24.5% vs. 22.0% for HF; and 21.5% vs. 20.0%, vs. 17.9% for pneumonia. Among African-Americans, adjusted readmission rates declined by 3.2 percentage points in AMI, 4.3 percentage points in HF, and 3.4 percentage points in pneumonia. The corresponding declines among whites were 2.9, 3.2, and 2.4; and among Hispanics were 3.0, 4.1, and 3.4. The magnitude of the disparity between white patients and African-Americans narrowed significantly for HF and pneumonia (0.6 [CI: 1.2 to 0.2] and 0.8 [1.3 to 0.3]). Among Hispanics and whites, the magnitude of the disparity significantly changed for pneumonia (0.9 [CI: 1.5 to 0.3]) but was unchanged for the two other conditions. We observed similar trends in stratified analysis by Medicare Advantage vs. fee-for-service enrollment. Interventions that are focused on minority groups and minority-serving hospitals are still needed to decrease disparities in readmission rates.   Approach to assessing model performance: computer 4 summary statistics including R-squared, over-fitting indices, distribution of Standardized Pearson Residuals and Predictive Ratios. For this endorsement maintenance submission, the quasi-R2 slightly increased up to 0.078, suggesting that about eight percent of the variation in payment could be explained by patient-level risk factors.

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

    Submission document: Testing attachment, section 2b4.

    Panel Member 1: None.

    Panel Member 2: Substantial variation in risk adjusted costs/episode.

    Panel Member 4: no concerns

    Panel Member 5: None. The variation in rates and the proportion of outliers suggests that there are meaningful differences across hospitals in risk-standardized payments associated with a 30-day episode of care for patients with HF.

    Panel Member 6: NA

    Panel Member 8: None.

    Panel Member 9: The distribution of measure scores across hospitals shows that there are meaningful differences, though the distribution is fairly tight. The range of risk-standardized payments across the 4,502 hospitals with a measure score was $13,171-$27,996. Hospitals in the 10th percentile have risk-standardized payments that are about 8.5% lower than the median; hospitals in the 90th percentile have payments that are about 10.6% higher than the median.

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 1: None.

    Panel Member 2: NA

    Panel Member 4: not applicable

    Panel Member 5: NA

    Panel Member 6: NA

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

    Submission document: Testing attachment, section 2b6.

Panel Member 1: None.

Panel Member 2: About 4% missing as not reported in administrative data. Would like to know more.

Panel Member 4: No concerns

Panel Member 5: None

Panel Member 6: NA

Panel Member 8: None.

For cost/resource use measures ONLY:

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

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

Panel Member 6: It is unclear is the cost should be attributed to the hospital vs physician. The hospital may not have control over the total cost. - Note toe NQF staff - please correct my response to 2431 to include this statement

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

Panel Member 1: None.

Panel Member 2: No concerns about attribution. Standard concerns about standardized pricing masking actual resource differences across hospitals.

Panel Member 5: None

Panel Member 6: NA

Panel Member 8: Again, I'm a little vague on the exact price standardization method. Worth noting, even within a single market, individuals hospitals have a number of adjustment factors that can affect the price of an admission. Certain AMCs, for example, may look more expensive than community hospitals in the same market b/c of these differences. This could create some systematic bias.

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: The measure passed all the relevant tests pertaining to validity criterion with high confidence.

Panel Member 2: Correlation with hospital standardized overall spending per beneficiary. Low risk adjustment quasi-R-square.

Panel Member 3: To demonstrate a moderate level, the developer must show an empirical association between the implicit quality construct and the material outcome

Panel Member 4: Moderately strong correlation between MSPB and HF payment measure.

Panel Member 5: No data element assessment.

Panel Member 6: Moderate correlation with MSPB, but unclear if that is a relevant comparison measure. Concerns regarding the attribution logic also.

Panel Member 8: Solid, but not overwhelming evidence.

Panel Member 9: Results above show good validity of the measure. The distribution of HF risk standardized payment measure scores was fairly tight, indicating the differences across deciles are quite small after risk adjustment. Also concerned about large difference in mean cost with and without the social risk factors (dual status and low SES), which indicates hospitals serving large percentages of patients with one or both of these factors could be significantly penalized given the large variation at the aggregate level for the measure.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

    ☐ **High**

    ☐ **Moderate**

    ☐ **Low**

    ☐ **Insufficient**

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

ADDITIONAL RECOMMENDATIONS

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

Panel Member 6: The three cost measures (AMI, HF and PN) all include the attribution of all costs to the hospital. Would like the group to discuss that logic to determine acceptability.

Panel Member 9: I believe the SMP and Standing Committee should consider the following recent analysis which is relevant to the 3 hospital readmission measures being evaluated during this cycle: The authors point out that The Hospital Readmissions Reduction Program publicly reports and financially penalizes hospitals according to 30-day risk-standardized readmission rates (RSRRs) exclusively among traditional Medicare (TM) beneficiaries but not persons with Medicare Advantage (MA) coverage and note that this approach may not accurately reflect hospitals' readmission rates for older adults. RESULTS: There were 748 033 TM patients (mean [SD] age, 76.8 [83] years; 360 692 [48.2%] women) and 295 928 MA patients (mean [SD] age, 77.5 [7.9] years; 137 422 [46.4%] women) hospitalized and discharged alive for AMI; 1 327 551 TM patients (mean [SD] age, 81 [8.3] years; 735 855 [55.4%] women) and 457 341 MA patients (mean [SD] age, 79.8 [8.1] years; 243 503 [53.2%] women) for CHF; and 2 017 020 TM patients (mean [SD] age, 80.7 [8.5] years; 1 097 151 [54.4%] women) and 610 790 MA patients (mean [SD] age, 79.6 [8.2] years; 321 350 [52.6%] women) for pneumonia. The 30-day RSRRs for TM and MA patients were correlated (correlation coefficients, 0.31 for AMI, 0.40 for CHF, and 0.41 for pneumonia) and the TM-based RSRR systematically underestimated the RSRR for all Medicare patients for each condition. Of the 2820 hospitals with 25 or more admissions for at least 1 of the outcomes of AMI, CHF, and pneumonia, 635 (23%) had a change in their penalty status for at least 1 of these conditions after including MA data. Changes in hospital performance and penalty status with the inclusion of MA patients were greater for hospitals in the highest quartile of MA admissions. Conclusions and relevance: In this cohort study, the inclusion of data from MA patients changed the penalty status of a substantial fraction of US hospitals for at least 1 of 3 reported conditions. This suggests that policy makers should consider including all hospital patients, regardless of insurance status, when assessing hospital quality measures. Association of Inclusion of Medicare Advantage Patients in Hospitals' Risk-Standardized Readmission Rates, Performance, and Penalty Status, January 2021, JAMA Network Open 4(2):e2037320; 10.1001/jamanetworkopen.2020.37320, Orestis A Panagiotou, Kirsten R Voorhies, Laura M Keohane, …., Amal N Trivedi

Developer Submission

Brief Measure Information

NQF #: 2436

De.2. **Measure Title:** Hospital-level, risk-standardized payment associated with a 30-day episode-of-care for heart failure (HF)

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

De.3. **Brief Description of Measure:** This measure estimates hospital-level, risk-standardized payment for a HF episode of care starting with inpatient admission to a short term acute-care facility and extending 30 days post-admission for Medicare fee-for-service (FFS) patients who are 65 years of age or older with a principal discharge diagnosis of HF.

IM.1.1. **Developer Rationale:** In 2019, total Medicare expenditures were $799.4 billion [1], representing 3.6% of gross domestic product (GDP). Current estimates suggest that Medicare spending will grow 7.6% per year between 2019 and 2028 [1]. This growth in spending underscores the need to create incentives for high value care. Measuring costs in a way that is transparent to consumers and fair to providers is an important component of understanding and controlling costs of care and rewarding value. Measuring condition-specific costs of care is needed to identify high value care.

HF is a common condition in the elderly with a substantial range in payments due to different practice patterns, making it an ideal condition for assessing relative value for an episode of care that begins with an acute hospitalization. HF is one of the top three leading causes of hospitalization for Americans over 65 years old [2] and is projected to cost the US up to $70 billion in direct and indirect costs by 2030 [3].

In part due to increasing Medicare spending on HF care and geographic variation in resource use and spending, Centers for Medicare and Medicaid Services (CMS) payment programs have aimed to incentivize reductions in spending as determined by Medicare payments made to providers and institutions, as well as improved clinical outcomes for an episode of HF care.

Medicare payments are difficult to interpret in isolation. Some high payment hospitals may have better clinical outcomes when compared with low payment hospitals; other high payment hospitals may not. For this reason, the value of hospital care is more clearly assessed when pairing hospital payments with hospital quality. A measure of payments for Medicare patients during an episode of care for HF aligned with current quality of care measures will facilitate profiling hospital value (payments and quality). This measure, which uses standardized payments, reflects differences in the management of care for patients with HF both during hospitalization and immediately post-discharge. By focusing on one specific condition, value assessments may provide actionable feedback to hospitals and incentivize targeted improvements in care.

This measure is intended to align with current quality measures to facilitate profiling hospital value (payments and quality). Given that HF is a condition with substantial variability in costs of care, aligning this payment measure with quality measures (e.g., RSMRs) will allow the assessment of hospital value. By evaluating their RSPs and RSMRs for HF, hospitals have an opportunity to consider actionable improvements and efficiencies on a broader scale to impact value of care. This measure provides transparency on the payments made for Medicare beneficiaries who have had HF. Hospitals receive detailed information on how they compare with other institutions regarding the amount and venues of resources expended on patients. As such, the measure provides insight to hospitals that is not otherwise possible.

References:

1. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NHE-Fact-Sheet

2. https://www.hcup-us.ahrq.gov/faststats/NationalDiagnosesServlet

3. Heidenreich, P.A., Albert, N.M., Allen, L.A., Bluemke, D.A., Butler, J., Fonarow, G.C., Ikonomidis, J.S., Khavjou, O., Konstam, M.A., Maddox, T.M., Nichol, G., Pham, M., Piña, I.L., & Trogdon, J.G. (2013). Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. Circulation: Heart Failure, 6(3):606-619.

De.1. **Measure Type:** Cost/Resource Use

S.5. **Data Source:** Claims

Enrollment Data

S.3. **Level of Analysis:** Facility

IF Endorsement Maintenance – Original Endorsement Date: Nov 07, 2014 **Most Recent Endorsement Date:** Feb 10, 2015

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?**

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.***

IM.1. **Opportunity for Improvement**

IM.1.1. **Briefly explain the rationale for this measure (e.g., the benefits or improvements in performance envisioned by use of this measure)**

In 2019, total Medicare expenditures were $799.4 billion [1], representing 3.6% of gross domestic product (GDP). Current estimates suggest that Medicare spending will grow 7.6% per year between 2019 and 2028 [1]. This growth in spending underscores the need to create incentives for high value care. Measuring costs in a way that is transparent to consumers and fair to providers is an important component of understanding and controlling costs of care and rewarding value. Measuring condition-specific costs of care is needed to identify high value care.

HF is a common condition in the elderly with a substantial range in payments due to different practice patterns, making it an ideal condition for assessing relative value for an episode of care that begins with an acute hospitalization. HF is one of the top three leading causes of hospitalization for Americans over 65 years old [2] and is projected to cost the US up to $70 billion in direct and indirect costs by 2030 [3].

In part due to increasing Medicare spending on HF care and geographic variation in resource use and spending, Centers for Medicare and Medicaid Services (CMS) payment programs have aimed to incentivize reductions in spending as determined by Medicare payments made to providers and institutions, as well as improved clinical outcomes for an episode of HF care.

Medicare payments are difficult to interpret in isolation. Some high payment hospitals may have better clinical outcomes when compared with low payment hospitals; other high payment hospitals may not. For this reason, the value of hospital care is more clearly assessed when pairing hospital payments with hospital quality. A measure of payments for Medicare patients during an episode of care for HF aligned with current quality of care measures will facilitate profiling hospital value (payments and quality). This measure, which uses standardized payments, reflects differences in the management of care for patients with HF both during hospitalization and immediately post-discharge. By focusing on one specific condition, value assessments may provide actionable feedback to hospitals and incentivize targeted improvements in care.

This measure is intended to align with current quality measures to facilitate profiling hospital value (payments and quality). Given that HF is a condition with substantial variability in costs of care, aligning this payment measure with quality measures (e.g., RSMRs) will allow the assessment of hospital value. By evaluating their RSPs and RSMRs for HF, hospitals have an opportunity to consider actionable improvements and efficiencies on a broader scale to impact value of care. This measure provides transparency on the payments made for Medicare beneficiaries who have had HF. Hospitals receive detailed information on how they compare with other institutions regarding the amount and venues of resources expended on patients. As such, the measure provides insight to hospitals that is not otherwise possible.

References:

1. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NHE-Fact-Sheet

2. https://www.hcup-us.ahrq.gov/faststats/NationalDiagnosesServlet

3. Heidenreich, P.A., Albert, N.M., Allen, L.A., Bluemke, D.A., Butler, J., Fonarow, G.C., Ikonomidis, J.S., Khavjou, O., Konstam, M.A., Maddox, T.M., Nichol, G., Pham, M., Piña, I.L., & Trogdon, J.G. (2013). Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. Circulation: Heart Failure, 6(3):606-619.

IM.1.2. **Provide performance scores on the measure as specified** (current and over time) **at the specified level of analysis.** (This is required for endorsement maintenance. Include mean, stddev, 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 subcriterion on improvement (U.3.1.) under Usability and Use.**

We examine the distribution of hospital payment scores to demonstrate the variation, current and over time, in payment among measured hospitals.

Current results: Within the measure reporting period (July 1, 2016-June 30, 2019), 987,227 admissions were included in this analysis, representing care at 4,502 hospitals. The mean risk-standardized payment (RSP) was $17,722, and the range was $13,171-$27,996. The median hospital RSP in the combined three-year dataset was $17,607 (IQR $16,817- $18,513).

The distribution of hospital HF RSPs for the three-year reporting period (2016-2019) are shown below, followed by the distribution of RSPs for individual years (2016/2017, 2017/2018,2018/2019).

Distribution for the 3-year reporting period (July 1, 2016-June 30, 2019):

Number of Hospitals: 4,502

Number of Admissions: 987,227

Mean(SD): 17,722(1,368)

Range(Min-Max): 13,171-27,996

Minimum: 13,171

10th percentile: 16,106

20th percentile: 16,632

30th percentile: 16,987

40th percentile: 17,312

50th percentile: 17,607

60th percentile: 17,914

70th percentile: 18,308

80th percentile: 18,804

90th percentile: 19,482

Maximum: 27,996

Distribution of hospital-level RSPs for each individual year:

Periods//YEAR1617//YEAR1718//YEAR1819

Number of Hospitals//4,380//4,351//4,344

Number of Admissions//322,586//333,072//331,569

Mean(SD)//17,774(985)//17,891(962)//17,427(968)

Range(Min-Max)//14,567-23,164//14,748-22,846//13,929-22,393

Minimum//14,567//14,747//13,929

10th percentile//16,654//16,786//16,313

20th percentile//17,024//17,140//16,670

30th percentile//17,277//17,408//16,947

40th percentile//17,493//17,607//17,141

50th percentile//17,670//17,781//17,310

60th percentile//17,879//18,000//17,520

70th percentile//18,131//18,245//17,795

80th percentile//18,501//18,576//18,141

90th percentile//19,029//19,169//18,675

Maximum//23,164//22,846//22,393

IM.1.3. **If no or limited performance data on the measure as specified is reported in IM.1.2., 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

IM.1.4. **Provide disparities data from the measure as specified** (current and over time) **by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** (This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) **This information also will be used to address the subcriterion on improvement (U.3.1.) under Usability and Use.**

The distribution of hospital-level measure scores stratified by the proportion of patients with social risk factors (dual eligibility and low AHRQ SES) are shown below. For either risk factor, measure scores do not vary meaningfully as a function of facilities' proportion of patients with social risk factors.

Distribution of HF RSPs by Proportion of Dual Eligible Patients (for Hospitals with 25 or More Cases):

Quartile//Q1//Q4

Social Risk Proportion(%)//(0-10.7)//(24.1-100)

# of Hospitals//866//866

Maximum//22,938//27,997

90th percentile//19,533//19,696

75th percentile//18,616//18,525

Median//17,687//17,463

25th percentile//16,807//16,564

10th percentile//16,060//15,839

Minimum//13,171//13,607

Distribution of HF RSPs by Proportion of Patients with AHRQ SES Index Scores (for Hospitals with 25 or More Cases):

Quartile//Q1//Q4

Social Risk Proportion(%)//(0-8.24)//(34.78-100)

# of Hospitals//866//865

Maximum//27,997//23,409

90th percentile//19,561//19,275

75th percentile//18,575//18,327

Median//17,649//17,354

25th percentile//16,734//16,448

10th percentile//15,983//15,701

Minimum//14,474//13,607

IM.1.5. If no or limited data on disparities from the measure as specified is reported in IM.1.4., then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.

N/A

IM.2. Measure Intent

IM.2.1. Describe intent of the measure and its components/ Rationale (including any citations) for analyzing variation in resource use in this way.

A hospital-level, episode-of-care payment measure for HF is informative for several reasons. First, it provides transparency into the differences in costs of care to Medicare for the same condition across hospitals. Second, it allows hospitals to assess the payments for patients admitted to their institution relative to other hospitals and thus may incentivize hospitals to examine their own practices and coordinate with post-discharge providers to seek new efficiencies. Finally, when paired with existing outcome measures for HF patients, it identifies institutions that, after removing the effect of geography, policy adjustments, and case mix, demonstrate good patient outcomes at low cost. Such hospitals may provide important examples of positive deviance from which other hospitals can learn.

The HF Payment measure is aligned with the HF mortality measure (NQF #0229). Other related measures of quality include the HF readmission measure (NQF #0330).

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.***

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 : Congestive Heart Failure

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

Care Coordination

De.7. Care Setting *(Select all the settings for which the measure is specified and tested):*

Inpatient/Hospital

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.)*

<WebPageURLExists nodeType="1">https://www.qualitynet.org/inpatient/measures/payment/methodology

S.2. Type of resource use measure *(Select the most relevant)*

Per episode

S.3. Level of Analysis *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED):*

Facility

S.4. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*

S.5. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*

If other, please describe in S.5.1.

Claims

Enrollment Data

S.5.1. Data Source or Collection Instrument *(Identify the specific data source or data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.)*

Data Sources

Medicare Inpatient and Outpatient Administrative Claims: This data source contains claims data for FFS inpatient and outpatient services including: Medicare inpatient hospital care, outpatient hospital services, skilled nursing facility care, some home health agency services, as well as inpatient and outpatient physician claims.

The 2020 reporting period for these analyses include Medicare administrative claims and enrollment information for patients with hospitalizations between July 1, 2016 and June 30, 2019. Medicare administrative claims for the 12 months prior to and during the index admission are used for risk adjustment. The period for public reporting of the AMI payment measure aligns with the 30-day AMI mortality and readmission measures for harmonization purposes.

Price standardization methodology:

The datasets also contain price-standardized payments for Medicare patients across all Medicare settings, services, and supplies (that is, inpatient, outpatient, SNF, home health agency, hospice, physician/clinical laboratory/ambulance services, and durable medical equipment, prosthetics/orthotics, and supplies). The CMS Standardization Methodology for Allowed Amount for 2009 through 2019 was applied to the claims to calculate the measures. Price-standardized payments for Medicare patients across all Medicare settings, services, and supplies (that is, inpatient, outpatient, SNF, home health agency, hospice, physician/clinical laboratory/ambulance services, and durable medical equipment, prosthetics/orthotics, and supplies) were calculated using standardized methodology specific to services reimbursed through Medicare parts A and B (for specific values see https://www.resdac.org/articles/cms-price-payment-standardization-overview).

Medicare Enrollment Database (EDB)

This database contains Medicare beneficiary demographic, benefit/coverage, and vital status information. This dataset was used to obtain information on enrollment, date of birth, and post-discharge mortality status. These data have previously been shown to accurately reflect patient vital status (Fleming et al. 1992).

Medicare Fee Schedules

Fee schedules are lists of pre-determined reimbursement amounts for certain services and supplies (e.g. physician services, independent clinical labs, ambulance services, durable medical equipment) and are used by Medicare in the calculation of payment to providers. We used the applicable fee schedules when calculating payments for claims that occurred in each care setting.

Federal Register Final Rules for Medicare Prospective Payment Systems and Payment Policies

Certain data necessary to calculate payments (e.g. annual base payments and conversion factors, DRG weights, wage indexes, and average length of stay) were taken from applicable Federal Register Final Rules.

CMS-published Wage Index Data

Wage index data not published in Federal Register Final Rules (such as the wage index data for Renal Dialysis Facilities) were obtained through the CMS website.

American Community Survey (2013-2017)

We used the American Community Survey (2013-2017) to derive an updated Agency for Healthcare Research and Quality (AHRQ) Socioeconomic Status (SES) index score at the patient nine-digit zip code level for use in studying the association between our measure and social risk factors (SRFs).

Reference

Fleming, C., Fisher, E., Chang, C., Bubolz, T., & Malenka, D. (1992). Studying Outcomes and Hospital Utilization in the Elderly: The Advantages of a Merged Data Base for Medicare and Veterans Affairs Hospitals. Medical Care, 30(5), 377-391.

S.5.2. Data Source or Collection Instrument Reference *(available at measure-specific Web page URL identified in S.1 OR in the file attached here) (Save file as: S_5_2_DataSourceReference)*

S.6. Data Dictionary or Code Table *(Please provide a web page URL or attachment if exceeds 2 pages. NQF strongly prefers URLs. Attach documents only if they are not available on a web page.)*

Data Dictionary:

URL:

Please supply the username and password:

Attachment: S6_Data_Dictionary-637454170807172519.xlsx

Code Table:

URL:

Please supply the username and password:

Attachment:

Construction Logic

S.7.1. Brief Description of Construction Logic

If applicable, summarize the general approach or methodology to the measure construction. This is most relevant to measures that are part of or rely on the execution of a measure system or applies to multiple measures.

This measure estimates hospital-level, risk-standardized payments for a 30-day episode of care for HF. To this end, we constructed a cohort of HF patients by examining the principal discharge diagnosis in administrative claims data. Specifically, we included Medicare fee-for-service patients 65 or older with a principal discharge diagnosis of an HF (defined by ICD-10 codes in attached data dictionary). We then applied several exclusion criteria as detailed in S.9.1.

Once our cohort was finalized, we examined all payments for these patients (including co-pays, co-insurance, and deductibles) that occurred within 30 days of the index admission. We included payments for all care settings, except Part D Medicare claims. We standardized payments across providers by removing or averaging geographic differences and removing policy adjustments from the total payment for that service. These payments were then assigned to the initial admitting hospital. As part of our model, we risk adjusted these payments for patient comorbidities listed in outpatient and inpatient claims in the 12 months prior to the index admission as well as the secondary diagnoses included in the index admission. We then used hierarchical generalized linear regression models to calculate a risk-standardized payment for each hospital.

S.7.2. Construction Logic *(Detail logic steps used to cluster, group or assign claims beyond those associated with the measure's clinical logic.)*

To construct the measure, we use Medicare administrative claims data. These data contain claims for all care settings, supplies, and services as outlined in Section S.7.8. (except Part D). Claim payment data are organized by the setting, supply, or service in which they were rendered. Standard Medicare payment rates were assigned to each service based on claim type, facility type, and place of service codes. These payments are then summed by individual patients. To create a hospital-level measure, we aggregate the payments for all eligible patients at each hospital.

S.7.2a. CONSTRUCTION LOGIC ATTACHMENT or URL: If needed, attach supplemental documentation (Save file as: S_7_2_Construction_Logic). All fields of the submission form that are supplemented within the attachment must include a summary of important information included in the attachment and its intended purpose, including any references to page numbers, tables, text, etc.

URL:
https://qualitynet.cms.gov/files/5d0d3b4e764be766b0105350?filename=HF_Pymnt_MeasMeth_Rprt_092313.pdf

Please supply the username and password:

Attachment:

S.7.3. Concurrency of clinical events, measure redundancy or overlap, disease interactions *(Detail the method used for identifying concurrent clinical events, how to manage them, and provide the rationale for this methodology.)*

This measure examines payments for a 30-day episode of care beginning with an admission for HF and extending to 30-days post-admission. We determine if a patient has HF by examining the principal discharge diagnosis code in the administrative data. If a patient has a principal discharge diagnosis of any other condition, even if this includes a secondary diagnosis of HF, this admission is not considered as an index admission. Therefore, the concurrency of clinical events is not an issue when determining what triggers the episode of care. Once, an episode is triggered, however, we include payments for all care settings, except Part D Medicare claims. The model risk adjusts for comorbidities listed in outpatient and inpatient claims in the 12 months prior to the index admission as well as the secondary diagnoses included in the index admission that are not considered complications of care.

S.7.4. Complementary services *(Detail how complementary services have been linked to the measure and provide rationale for this methodology.)*

The measure includes payments for all care settings, except Part D, that occur during the 30-day window. If a claim for a complimentary service was filed in the study window, then it would be included in the measure.

S.7.5. Clinical hierarchies *(Detail the hierarchy of codes or condition groups used and provide rationale for this methodology.)*

The measure uses a risk-adjustment model based on Condition Categories (CCs) as opposed to Medicare Advantage Hierarchical Condition Categories (HCCs). We used CCs because they provide detailed descriptions about comorbidities that may influence care decisions that affect payment for HF without assigning hierarchy. This allows conditions that would be ranked lower in the hierarchy to be considered for risk adjustment if they are medically and statistically relevant. For example, it would allow for the inclusion of both HCC 34 (peptic ulcer, hemorrhage, other specified gastrointestinal disorders) and HCC 33 (inflammatory bowel disease) rather than only HCC 33, which is considered the more “severe” condition.

S.7.6. Missing Data *(Detail steps associated with missing data and provide rationale for this methodology (e.g., any statistical techniques to impute missing data)*

:

We do not impute missing data for any of the variables included in the measure. However, if a hospitalization is missing a DRG or DRG weight, we exclude it as an index admission.

S.7.7. Resource Use Service Categories (Units) (Select all categories that apply)

Inpatient services: Inpatient facility services

Inpatient services: Evaluation and management

Inpatient services: Procedures and surgeries

Inpatient services: Imaging and diagnostic

Inpatient services: Lab services

Inpatient services: Admissions/discharges

Inpatient services: Labor (hours, FTE, etc.)

Other inpatient services

Ambulatory services: Outpatient facility services

Ambulatory services: Emergency Department

Ambulatory services: Pharmacy

Ambulatory services: Evaluation and management

Ambulatory services: Procedures and surgeries

Ambulatory services: Imaging and diagnostic

Ambulatory services: Lab services

Ambulatory services: Labor (hours, FTE, etc.)

Other ambulatory services

Durable Medical Equipment (DME)

Other services not listed

See S.7.8 for full list of care settings included

See S.7.8 for full list of care settings included

See S.7.8 for full list of care settings included

S.7.8. Identification of Resource Use Service Categories (Units)

(For each of the resource use service categories selected above, provide the rationale for their selection and detail the method or algorithms to identify resource units, including codes, logic and definitions.)

To estimate payments for a 30-day episode of care for HF we included payments for all care settings, services, and supplies, except drugs covered under Part D Medicare claims. We did not include Part D since a large proportion of Medicare beneficiaries are not enrolled in Part D and there is variation in enrollment status across and within states. Including payments for Part D services would thus bias payments upwards for hospitals with high Part D enrollment. By following patients through an episode of care for HF, CMS and hospitals can gain key insights into the drivers of payments and how practice patterns vary across providers.

We include payments for the following care settings below in the measure:

Inpatient hospital facility and physician

Outpatient hospital facility and physician

Skilled nursing facility and physician

Hospice facility and physician

Home health facility and physician

Inpatient psychiatric facility and physician

Inpatient rehab facility and physician

Long-term care hospital facility

Clinical labs facility and physician

Comprehensive outpatient rehab facility and physician

Outpatient rehab facility and physician

Renal dialysis facility and physician

Community mental health centers facility and physician

DME/POS/PEN

Observation stay facility

Part B drugs

Ambulance and ambulance physician

Emergency department facility and physician

Physician office

Federally qualified health centers facility and physician

Rural health clinics facility and physician

Ambulatory surgical centers facility and physician

We also include physician payments for the following care settings:

Indian health service free-stand facility

Indian health service provider facility

Tribal free-standing facility

Tribal facility

Military treatment facility

Independent clinic

State or local health clinic

Mass immunization center

Walk-in retail health clinic

Urgent care facility

Unassigned

Pharmacy

School

Homeless Shelter

Prison

Group Home

Mobile Unit

Temporary Lodging

Birthing Center

Intermediary Care/Mentally Retarded

Residential Substance Abuse

Psychiatric Residential Facility

Non-Residential Substance Abuse

Other Physician

Other carrier claims with HCPCS codes P9603 or P9604

In order to determine how to assign claims, we examine the place of service code for physician claims and a combination of claim type and facility type codes to determine the facility in which care was provided. Depending on the facility and physician codes we standardize payments differently. Information on how we standardize claims can be found in section S.9.6.

S.7.8a. If needed, provide supplemental resource use service category specifications in either URL (preferred) or as an attachment (Save file as S.7.8a_RU_Service_Categories):

URL: https://qualitynet.cms.gov/files/5d0d3b4e764be766b0105350?filename=HF_Pymnt_MeasMeth_Rprt_092313.pdf

Please supply the username and password:

Attachment:

Clinical Logic

S.8.1. **Brief Description of Clinical Logic** (Briefly describe your clinical logic approach including clinical topic area, whether or not your account for comorbid and interactions, clinical hierarchies, clinical severity levels and concurrency of clinical events.)

HF is a common condition in the elderly with substantial variability in payments due to different practice patterns. Quality measures for HF such as 30-day HF risk-standardized mortality rate (RSMR) are already publicly reported. In the context of its publicly reported quality measures, HF is an ideal condition in which to assess payments for Medicare patients and relative hospital value. Therefore, we created a measure of payments for a 30-day episode of care for HF that is aligned with CMS's 30-day AMI mortality and readmission measures, making it possible for CMS to assess the value of care provided for these episodes.

The measure uses Condition Categories (CCs) to adjust for patient case-mix across hospitals. Details of our risk-adjustment strategy can be found in our technical report at https://www.qualitynet.org/inpatient/measures/payment/methodology.

This measure is for patients who are admitted with HF. We determine this by examining the principal discharge diagnosis code in the administrative data. If a patient has a principal discharge diagnosis of any other condition, even if this includes a secondary diagnosis of HF, this admission is not considered as an index admission. Therefore, the concurrency of clinical events is not applicable for this measure. However, the model does risk adjust for comorbidities listed in outpatient and inpatient claims in the 12 months prior to the index admission as well as the secondary diagnoses included in the index admission that are not considered complications of care.

S.8.2. Clinical Logic *(Detail any clustering and the assignment of codes, including the grouping methodology, the assignment algorithm, and relevant codes for these methodologies.)*

We focused on a 30-day episode of care triggered by admission for an HF as identified using ICD-10 diagnosis codes described in the data dictionary. The measure includes admissions for Medicare FFS beneficiaries aged 65 years and older. A full list of codes used to identify these conditions is provided in the data dictionary.

We assigned all payments for the episode of care to the hospital that originally admitted the patient.

S.8.3. Evidence to Support Clinical Logic Described in S.8.2 *Describe the rationale, citing evidence to support the grouping of clinical conditions in the measurement population(s) and the intent of the measure (as described in IM3)*

The intent of the measure is to estimate payments for a 30-day episode of care for HF in order to gain insight into drivers of payment within and across hospitals. To profile hospital payments fairly, the measure fulfills the following criteria:

1. We standardize payments for geography to isolate payment differences related to the clinical care of patients with HF.

2. We adjust for hospital case-mix.

3. We align the HF payment measure specifications with the nationally reported 30-day HF risk-standardized mortality measure (RSMR) to identify practice patterns that may be expensive without conferring a quality benefit across an episode of care for HF.

4. We focused on a specific disease condition to provide the most meaningful feedback to hospitals and incentivize targeted improvements in care.

S.8.3a. CLINICAL LOGIC ATTACHMENT or URL: If needed, attach supplemental documentation (Save file as: S_8_3a_Clinical_Logic). All fields of the submission form that are supplemented within the attachment must include a summary of important information included in the attachment and its intended purpose, including any references to page numbers, tables, text, etc.

URL:

Please supply the username and password:

Attachment:

S.8.4. Measure Trigger and End mechanisms *(Detail the measure's trigger and end mechanisms and provide rationale for this methodology)*

When considering hospital payments, we focused on a 30-day episode of care triggered by admission for HF for several key reasons. First, hospitalizations represent brief periods of illness that require ongoing management post-discharge. Second, decisions made at the admitting hospital affect payments for care in the immediate post-discharge period. Third, attributing payments for a continuous episode of care to admitting hospitals may reveal practice variations in the immediate and extended care of the illness that can result in increased payments. Fourth, a 30-day preset window provides a standard observation period by which to compare all hospitals. Lastly, we designed the HF payment measure to be aligned with HF quality measures, i.e. CMS´s publicly reported 30-day HF mortality and readmission measures.

S.8.5. Clinical severity levels *(Detail the method used for assigning severity level and provide rationale for this methodology)*

The measure uses administrative claims data to risk-adjust for patient comorbidities but does not include adjustments for clinical severity. Our team has demonstrated the validity of claims-based measures for profiling hospitals for a number of prior measures by comparing either the measure results or the individual data elements against medical records. CMS validated the six NQF-endorsed claims-based measures currently in public reporting (i.e., mortality and readmission measures for AMI, heart failure, and pneumonia) with models that used medical record-abstracted data for risk adjustment. Specifically, claims model validation was conducted by building comparable models using abstracted medical record data for risk adjustment for heart failure patients (National Heart Failure data), AMI patients (Cooperative Cardiovascular Project data) and pneumonia patients (National Pneumonia Project dataset). When both models were applied to the same patient population, the hospital risk-standardized mortality and readmission rates estimated using the claims-based risk-adjustment models had a high level of agreement with the results based on the medical record model, thus supporting the use of the claims-based models for public reporting.

S.8.6. Comorbid and interactions *(Detail the treatment of co-morbidities and disease interactions and provide rationale for this methodology.)*

The goal of risk adjustment for this measure is to account for patient age, prior procedures (e.g., PCI and/or CABG), and comorbid conditions that are clinically relevant and have strong relationships with the outcome, while illuminating important payment differences between hospitals.

Comorbidities that are included in risk adjustment are identified in administrative claims during the 12 months prior to and including the index admission. To assemble the more than 70,000 ICD-10 codes into clinically coherent variables for risk adjustment, the measure primarily employs the publicly available CMS condition categories (CCs) to group ICD-10 codes into CCs, and selects comorbidities on the basis of both clinical relevance and statistical significance [1].

Reference

Pope G, Ellis R, Ash A, et al. Principal Inpatient Diagnostic Cost Group Models for Medicare Risk Adjustment. Health Care Financing Review. 2000;21(3):26.

Adjustments for Comparability

S.9.1. Inclusion and Exclusion Criteria *Detail initial inclusion/exclusion criteria and data preparation steps (related to clinical exclusions, claim-line or other data quality, data validation, e.g. truncation or removal of low or high dollar claim, exclusion of ESRD patients)*

:

The measure includes admissions for Medicare FFS beneficiaries aged 65 years and older who are discharged from non-federal short-term acute care hospitals (including Indian Health Service hospitals) and critical access hospitals, with an eligible principal discharge diagnosis of HF.

CMS FFS beneficiaries with an index hospitalization to an acute care non-federal hospital are included if they have been enrolled in Part A and Part B Medicare for the 12 months prior to the date of admission to ensure a full year of administrative data for risk-adjustment.

For patients with more than one admission in a given year for a given condition, only one admission is randomly selected to include in the cohort as an index HF. Additional eligible HF admission occurring within that given year are excluded.

The episode of care begins with an admission for HF to a short-term acute care hospital. The hospital that initially admits the patient is assigned all payments that occur during the episode of care. This includes payments for patients who are subsequently transferred to another hospital for further care of the index HF. Claims from an emergency department do not begin the episode of care because CMS does not classify emergency department care as an inpatient admission. If a patient is transferred from an emergency department to another hospital and then subsequently admitted, the episode of care begins with the inpatient admission at the receiving hospital.

See data dictionary for list of codes used to define measure specifications.

Exclusion Criteria for HF Payment Measure

1. Discharged against medical advice (AMA)

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

2. Incomplete administrative data in the 30 days following the index admission if discharged alive.

Rationale: This is necessary in order to identify the outcome (payments) in the sample over our analytic period.

3. Transferred to a federal hospital

Rationale: We do not have claims data for these hospitals; therefore, including these patients would systematically underestimate payments.

4. Discharged alive on day of admission or following day and not transferred to another acute care facility.

Rationale: This exclusion prevents inclusion of patients who likely did not have clinically significant HF.

5. Not matched to admission in the HF mortality measure

Rationale: As part of the current data processing, we match our index HF admissions to the HF mortality cohort to obtain the risk-adjustment variables. Patients are excluded if they cannot be matched between the HF payment and HF mortality cohorts.

6. Missing index DRG weight where provider received no payment

Rationale: With neither DRG weight nor payment data, we cannot calculate a payment for the patient's index admission; this would make the entire episode of care appear significantly less expensive

7. Patients with inconsistent or unknown vital status or other unreliable demographic data

Rationale: Reliable and consistent data are necessary for valid calculation of the measure.

8. Patients enrolled in the Medicare hospice program any time in the 12 months prior to the index admission, including on the first day of the index admission.

Rationale: These patients are excluded to align with the 30-Day HF Mortality measure.

9. With a procedure code for LVAD implantation or heart transplantation either during the index admission or in the 12 months prior to the index admission.

Rationale: For patients with more than one eligible admission for an HF in a single year, only one index admission for HF is randomly selected for inclusion in the cohort. Additional admissions within that year are excluded. When index admissions occur during the transition between two years within the measurement period (that is, June/July 2017 or June/July 2018) and both are randomly selected for inclusion in a measure, the measures include only the June admission. July admissions within the 30-day outcome window of the June admission are excluded to avoid assigning payments for the same claims to two admissions.

S.9.2. Risk Adjustment Type (Select type)

Statistical risk model

If other:

S.9.3. Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets)*

N/A

S.9.4 Costing method

Detail the costing method including the source of cost information, steps to capture, apply or estimate cost information, and provide rationale for this methodology.

Standardized pricing

Medicare pays for health care services using a number of different payment systems that are generally organized by delivery setting. These payment systems consider not only the products the Medicare patient is buying in each setting, but also the characteristics of the care provider, the extent to which the same product may be furnished in different settings, and the market circumstances that affect providers' costs. Payment amounts within each payment system are usually updated annually (for example, the IPPS) with some fee schedules having quarterly updates (for example, Durable Medical Equipment/Prosthetics Orthotics and Supplies [DME/POS]). Information on CMS reimbursement rates for each care setting are made publicly available through either Final Rules published in the Federal Register or fee schedules provided on the CMS website. A summary of Medicare's reimbursement system for most care settings is publicly available at the Medicare Payment Advisory Committee (MedPAC) website.

In the measure technical report we describe the key features of these payment systems and how we used these CMS payment algorithms to determine an episode-of-care payment for PN that isolates clinical care decisions. Please see Appendix C in the technical report for a full description of how we standardize payments for each care setting: https://qualitynet.cms.gov/files/5d0d3b4e764be766b0105350?filename=HF_Pymnt_MeasMeth_Rprt_092313.pdf

Details of CMS´s price standardization methodology are described here: https://qualitynet.cms.gov/files/5ea9c75a6a6cce001fbe04eb?filename=CMS_Price_Pymnt_Standardization.pdf

S.9.6b_Standardized_Pricing_Table-635222004979339249-637266117498760428.pdf

S.10. Type of score*(Select the most relevant):*

Continuous variable

If other:

Attachment:

S.11. Interpretation of Score *(Classifies interpretation of a ratio score(s) according to whether higher or lower resource use amounts is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score, etc.)*

Results of the measure alone do not necessarily reflect the quality of care provided by hospitals but simply whether the total episode payments are greater than or less than would be expected for an average hospital with a similar case mix. Hospitals are classified as having a less than average, no different than average, or greater than average payment as compared to national average payment for an episode. Accordingly, a classification of lower than average payment should not be interpreted as better care. The HF risk-standardized payment (RSP) is most meaningful when presented in the context of an HF outcome measure, such as the publicly reported HF mortality measure. This is because a measure of payments to hospitals that is aligned with a quality measure facilitates profiling hospital value (payments and quality).

S.12. Detail Score Estimation *(Detail steps to estimate measure score.)*

The RSP is calculated as the ratio of "predicted" payment to "expected" payment, multiplied by the national unadjusted average payment for the episode of care. The expected payment for each hospital is estimated using its patient mix and the average of the hospital-specific intercepts. The predicted payment for each hospital is estimated given the same patient mix but an estimated hospital-specific intercept. Operationally, the expected payment for each hospital is obtained by summing the expected payments for all patients in the hospital. The expected payment for each patient is calculated via the hierarchical model by applying the subsequent estimated regression coefficients to the observed patient characteristics and adding the average of the hospital-specific intercepts. The predicted payment for each hospital is calculated by summing the predicted payments for all patients in the hospital. The predicted payment for each patient is calculated through the hierarchical model by applying the estimated regression coefficients to the patient characteristics observed and adding the hospital-specific intercept.

Reporting Guidelines

This section is optional and will be available for users of the measure as guidance for implementation and reporting.

S.13.1. Describe discriminating results approach

Detail methods for discriminating differences (reporting with descriptive statistics--e.g., distribution, confidence intervals).

To categorize hospital payments, CMS estimates each hospital's RSP and the corresponding 95% interval estimate. CMS assigns hospitals to a payment category by comparing each hospital's RSP interval estimate to the national mean payment. Comparative payments for hospitals with 25 or more eligible cases are classified as follows:

- "No Different than the National Payment" if the 95% interval estimate surrounding the hospital's RSP includes the national mean payment.
- "Greater than the National Payment" if the entire 95% interval estimate surrounding the hospital's RSP is higher than the national mean payment.
- "Less than the National Payment" if the entire 95% interval estimate surrounding the hospital's RSP is lower than the national mean payment.

If a hospital has fewer than 25 eligible cases for a measure, CMS assigns the hospital to a separate category: "Number of Cases Too Small." This category is used when the number of cases is too small (fewer than 25) to reliably estimate the hospital’s RSP. If a hospital has fewer than 25 eligible cases, the hospital’s RSP and interval estimate will not be reported for the measure.

S.13.2. Detail attribution approach

Detail the attribution rules used for attributing resources/costs to providers (e.g., a proportion of total measure cost or frequency of visits during the measure's measurement period) and provide rationale for this methodology.

The measure attributes payments incurred during the 30-day episode to the original admitting hospital. We assign these payments to the admitting hospital because decisions made at the admitting hospital affect payments for care in the inpatient setting as well as the post-discharge and recovery periods for a HF. Furthermore, attributing payments for a continuous episode of care to admitting hospitals may reveal practice variations in the full care of the illness that can result in increased payments. For patients who are admitted and then transferred to another hospital during the original index admission, we assign all payments to the original admitting hospital since this hospital is responsible for the initial care decisions and the decision to transfer the patient.

S.13.3. Identify and define peer group

Identify the peer group and detail how peer group is identified and provide rationale for this methodology.

As part of the measure methodology we compare payments for a hospital with the expected payment amounts for an average hospital with the same case mix. While we include all hospitals when estimating the risk-adjustment model, we do not report RSPs for hospitals with fewer than 25 HF admissions, since estimates for hospitals with fewer procedures are less reliable and CMS’s past approach to public reporting has been not to report these results.

S.13.4. Sample size

Detail the sample size requirements for reporting measure results.

In order for hospitals to be publicly reported, they must have at least 25 index HF admissions during the measurement period.

S.13.5. Define benchmarking and comparative estimates

Detail steps to produce benchmarking and comparative estimates and provide rationale for this methodology.

Comparative estimates are provided by classifying hospitals as less than average, no different than average, or greater than average payment depending on the span of their confidence interval in comparison with the national average payment amount (i.e., the benchmark). To categorize hospital payments, we estimate each hospital’s RSP and the corresponding 95% interval estimate. As with all estimates, there is a degree of uncertainty associated with the RSP. The interval estimate is a range of probable values around the RSP that characterizes the amount of uncertainty associated with the estimate. A 95% interval estimate indicates that there is 95% probability that the true value of the RSP lies between the lower limit and the upper limit of the interval. In an effort to provide fair comparisons, we provide three categories (less than, no different than, or greater than the national average payment amount), which allows for conservative discrimination of hospital RSPs.

Validity – See attached Measure Testing Submission Form

SA.1. Attach measure testing form

NQF_2436_HFpayment_Testing_Spring2021_010521_FINAL-637541841858856085.docx

Measure Number (*if previously endorsed*)**:** 2436

Measure Title: Hospital-level, risk-standardized payment associated with a 30-day episode-of-care for Heart Failure (HF)

Date of Submission: **1/5/2021**

Type of Measure:

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

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 Database (including the Master Beneficiary Summary File), Medicare Fee Schedules, CMS Wage Index Data.	☒ other: Census Data/American Community Survey, Medicare Enrollment Database (including the Master Beneficiary Summary File), Medicare Fee Schedules, CMS Wage Index Data.)

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

The datasets/data sources we used in testing include: Medicare administrative claims data, Medicare enrollment database (EDB), Medicare fee schedules, Federal Register Final Rules for Medicare PPS systems and payment policies, and CMS published wage index data.

To assess socioeconomic factors, we used census as well as Medicare enrollment data. Dual eligibility was obtained through enrollment data. The Agency for Healthcare Research and Quality (AHRQ) socioeconomic status (SES) index score was obtained using the American Community Survey (ACS), 2013-2017.

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. 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 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 datasets also contain price-standardized payments for Medicare patients across all Medicare settings, services, and supplies (that is, inpatient, outpatient, SNF, home health agency, hospice, physician/clinical laboratory/ambulance services, and durable medical equipment, prosthetics/orthotics, and supplies). The CMS Standardization Methodology for Allowed Amount for 2006 through 2019 was applied to the claims to calculate the measures.

Refer to the original methodology report for further descriptions of these data sources.

Federal Register Final Rules for Medicare Prospective Payment Systems and Payment Policies: Certain data necessary to calculate payments (e.g. annual base payments and conversion factors, DRG weights, wage indexes, and average length of stay) were taken from applicable Federal Register Final Rules. CMS-published Wage Index Data Wage index data not published in Federal Register Final Rules (such as the wage index data for Renal Dialysis Facilities) were obtained via the CMS website.

CMS-published Wage Index Data Wage index data not published in Federal Register Final Rules (such as the wage index data for Renal Dialysis Facilities) were obtained via the CMS website.

Table 1. Dataset Descriptions

Dataset	Applicable Section in the Testing Attachment	Description of Dataset
Original 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	Full 2008 Sample (**Sample "A3")** Dates of Data: January 1, 2008 – January 31, 2008 Number of admissions = 348,061 Number of measured hospitals: 4,579 Full 2009 Sample: This cohort was randomly split for initial model testing. First half of split sample (**Sample "A1"**) -Number of Admissions: 173,296 -Number of Measured Hospitals: 4,508 Second half of split sample (**Sample "A2")** -Number of Admissions: 173,296 -Number of Measured Hospitals: 4,493
EM Testing Dataset (Medicare Fee-For-Service Administrative Claims Data) (July 1, 2016 – June 30, 2019)	Section 2a2 Reliability Testing Section 2b1 Validity Testing Section 2b2 Testing of Measure Exclusion	Dates of Data: July 2016-June 2019. Number of admissions = 987,227. Number of measured hospitals: 4,502.

Dataset	Applicable Section in the Testing Attachment	Description of Dataset
	Section 2b3 Risk Adjustment/Stratification Section 2b3.6. Statistical Risk Model Discrimination Statistics Section 2b4 Meaningful Differences	This cohort was randomly split into two halves. First half of split sample - Number of Admissions: 492,506. - Number of measured hospitals: 4,457. Second half of split sample - Number of Admissions: 494,721. - Number of measured hospitals: 4,502. Patient Descriptive Characteristics:
The American Community Survey (ACS)	Section 2b3: Risk adjustment/Stratification for Outcome or Resource Use Measures	Dates of Data: 2013-2017 We used the AHRQ SES index score derived from the American Community Survey (2013-2017) to study the association between the 30-day readmission outcome and SRFs. The AHRQ SES index score is based on beneficiary 9-digit zip code level of residence and incorporates 7 census variables found in the American Community Survey.
Master Beneficiary Summary File (MBSF)	Section 2b3: Risk adjustment/Stratification for Outcome or Resource Use Measures	Dates of Data: July 2016 – June 2019 We used dual eligible status (for Medicare and Medicaid) derived from the MBSF to study the association between the 30-day measure outcome and dual-eligible status.

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

We selected social risk factor (SRF) variables to analyze after reviewing the literature and examining available national data sources. We sought to find variables that are consistently captured in a reliable fashion for all patients in this measure. There is a large body of literature linking various social risk factors to worse health status and higher mortality, readmissions, complications of care, and outcomes and cost of care more broadly. Overall, income, education, and occupation are the most commonly examined SRFs studied.

For heart failure specifically, studies have shown that black and Hispanic Medicare patients with heart failure were less likely to have a follow-up clinic visit but more likely to experience a more costly readmission within 30 days of hospital discharge, compared with white patients. In addition, patients with Medicare Advantage plans (considered by the authors as a possible proxy for income) were much less likely to have a follow-up visit and also more likely to experience a readmission within 30 days; the inverse was true for patients with commercial insurance. Medicare patients with additional Medicaid coverage were much less likely to have a follow-up visit but also less likely to experience a readmission.

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)

  We selected the AHRQ SES index score because it is a well-validated variable that describes the average SES of people living in defined geographic areas. Its value as a proxy for patient-level information is dependent on having the most granular-level data with respect to communities that patients live in. We considered the area deprivation index (ADI) among many other potential indicators when we initially evaluated the impact of 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 American Community Survey (ACS) data. A census block group is a geographical unit used by the US Census Bureau which is between the census tract and the census block. It is the smallest geographical unit for which the bureau publishes sample data. The target size for block groups is 1,500 and they typically have a population of 600 to 3,000 people. We used 2013-2017 ACS data and mapped patients’ 9-digit ZIP codes via vendor software to the census block group level. Given the variation in cost of living across the country, the median income and median property value components of the AHRQ SES Index were adjusted by regional price parity values published by the Bureau of Economic Analysis (BEA). This provides a better marker of low SES neighborhoods in high expense geographic areas. We then calculated an AHRQ SES Index score for census block groups that can be linked to 9-digit ZIP codes. We used the percentage of patients with an AHRQ SES index score equal to or below 42.7 to define the lowest quartile of the AHRQ SES Index.

References:

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

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

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

DeLia, D., Tong, J., Gaboda, D., & Casalino, L. P. (2014). Post-discharge follow-up visits and hospital utilization by Medicare patients, 2007-2010. Medicare & medicaid research review, 4(2), mmrr.004.02.a01.

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.

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

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 estimated the overall measure score reliability by calculating the intra-class correlation coefficient (ICC) using a split sample (i.e. test-retest) method.

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. Hospital performance is measured once using a random subset of patients from a defined dataset from a measurement period, and then measured again using a second random subset exclusive of the first from the same measurment period, 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 patients aged 65 years and older, we randomly sampled half of patients within each hospital from a one-year measurement period, calculated the measure for each hospital, and repeated the calculation using the second half of patients. Thus, each hospital is measured twice, but each measurement is made using an entirely distinct set of patients. To the extent that the calculated measures of these two subsets agree, we have evidence that the measure is assessing an attribute of the hospital, not of the patients. As a metric of agreement we calculated the intra-class correlation coefficient (Shrout & Fleiss, 1979), and assessed the values according to conventional standards. Specifically, we used the 2020 EM Testing Dataset and, randomly split it into two approximately equal subsets of patients, and calculated the RSRR for each hospital for each sample. The agreement of the two RSRRs was quantified for hospitals in each sample using the intra-class correlation as defined by ICC (2,1). (Shrout & Fleiss, 1979).

Using two non-overlapping random samples provides a conservative estimate of the measure's reliability, compared with using two random, but potentially overlapping samples which would exaggerate the agreement. Moreover, because our final measure is derived using hierarchical logistic regression, and a known property of hierarchical logistic regression models is that smaller volume hospitals contribute less 'signal', a split sample using a single measurement period would introduce extra noise. This leads to an underestimate in the actual 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., The reliability of provider profiling: A tutorial. RAND Health, 2009. https://www.rand.org/content/dam/rand/pubs/technical_reports/2009/RAND_TR653.pdf; accessed on September 13, 2020

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*)
Measure Score Reliability Results

In total, 987,227 admissions were included in the analysis, using 3 years of data. After randomly splitting the sample into two halves, there were 494,721 admissions from 4,502 hospitals in one half and 492,506 admissions from 4,457 hospitals in the other half.

As a metric of agreement, we calculated the ICC for hospitals with 25 admissions or more.

Using the Spearman-Brown prediction formula, the agreement between the two independent assessments of the risk-standardized payment (RSP) for each hospital was 0.781.

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

The split-sample reliability score of 0.781 discussed in the previous section, represents the lower bound of estimate of the true measure score reliability.

According to published interpretations of reliability, this is considered sufficiently high (Adams et al., 2010; Landis and Koch, 1977; Yu et al., 2013).

References:

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

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)*

Measure Score Validity-Validity Indicated by Established Measure Development Guidelines

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

Measure Score Validity-Validity as Assessed by External Groups

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

The health economist with whom we consulted had years of experience in economic analysis and working with claims data. We worked with the consultant to address key issues surrounding measure development, including detailed discussions regarding the appropriate cohort for inclusion in the measure. Having regular meetings with a consultant provided a forum for focused expert review and discussion of technical issues during measure development prior to consideration by the broader TEP.

In addition to consulting with a health economist, and in alignment with the CMS Measure Management System, we convened a TEP to provide input and feedback during measure development from a group of recognized experts in relevant fields. To convene the TEP, we released a public call for nominations and selected individuals who represent a range of perspectives including clinicians, consumers, and purchasers, as well as individuals with experience in quality improvement, performance measurement, and healthcare disparities. We convened two structured TEP conference calls consisting of presentation of key issues, our proposed approach, and relevant data, followed by open discussion among TEP members. We made modifications to the measure based on TEP feedback.

Following completion of the measure, we solicited public comment on the measure through CMS, and the public comments were posted publicly for 60 days.

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.

Technical Expert Panel Members:

Ann-Marie Audet, MD Commonwealth Fund

Peter Bach, MD Memorial Sloan-Kettering Cancer Center

Richard Bankowitz, MD Premier Inc.

Donald Casey, MD New York University Langone Medical Center

Lesley Curtis, PhD Duke University

David Dunn, MD ZHealth LLC

Terri Golash, MD Aetna

Vivian Ho, PhD Rice University

David Hopkins, PhD Pacific Business Group on Health

Amanda Kowalski, PhD Yale University

Kavita Patel, MD Brookings Institute

Stephen Schmaltz, PhD Joint Commission

Measure Score Validity-Face Validity as Determined by TEP

To systematically assess face validity, we surveyed the Technical Expert Panel and asked each member to rate the following statement using a six-point scale (1=Strongly Disagree, 2=Moderately Disagree, 3=Somewhat Disagree, 4=Somewhat Agree, 5= Moderately Agree, and 6=Strongly Agree): “This is a measure of payments for Medicare patients for a 30-day HF episode of care. The measure removes policy adjustments that are independent of care decisions and risk-adjusts based on case mix. The measure is intended to provide CMS a tool to compare payments across hospitals nationally to identify hospitals that have notably higher or lower payments associated with HF care. To what extent does the committee agree that this measure accomplishes this purpose?”

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 payment measure, we identified and assessed the measure’s correlation with other measures that target the same domain (payment or utilization) for the same or similar populations. After literature review and consultations with measure experts in the field, there were very few measures identified. Given that challenge, we selected the hospital Medicare Spending per Beneficiary (MSPB) measure for comparison. We report an unweighted Pearson’s correlation coefficient for this analysis.

The hospital Medicare Spending per Beneficiary (MSPB) measure is a risk-adjusted, price-standardized measure that assesses Medicare Part A and Part B payments for services provided to Medicare beneficiaries for episodes that spanning from three days prior to an inpatient hospital admission through 30 days after discharge. More information about the hospital MSPB measure can be found here: https://qualitynet.cms.gov/inpatient/measures/mspb.

Because the MSPB measure assesses payments for all Medicare FFS patients for all conditions during the measurement period, and the HF payment measure is focused on a single diagnosis, we predicted that HF payment measure scores would be weakly-to-moderately, positively correlated with MSPB measure scores.

As additional evidence of measure score validity, we also present a measure of internal validity of the outcome by examining the distribution of payment types across the quartiles of risk-standardized payments. Our expectation would be that detailed level observed payments would be greater in hospitals with higher risk-standardized payments.

References:

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

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

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

Krumholz HM, Wang Y, Mattera JA, Wang Y-F, Han LF, Ingber MJ, Roman S, Normand SL. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. Circulation. 2006 Apr 4;113(13):1683-92.

Krumholz HM, Wang Y, Mattera JA, et al. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with heart failure. Circulation 2006;113(13):1693-1701.

Krumholz HM, Lin Z, Drye EE, Desai MM, Han LF, Rapp MT, Mattera JA, Normand SL. An administrative claims measure suitable for profiling hospital performance based on 30-day all-cause readmission rates among patients with acute myocardial infarction. Circulation: Cardiovascular Quality and Outcomes. 2011 Mar 1;4(2):243-52.

Lindenauer PK, Normand SL, Drye EE, Lin Z, Goodrich K, Desai MM, Bratzler DW, O'Donnell WJ, Metersky ML, Krumholz HM. Development, validation, and results of a measure of 30-day readmission following hospitalization for pneumonia. Journal of Hospital Medicine. 2011 Mar;6(3):142-50.

National Quality Forum. National voluntary consensus standards for patient outcomes, first report for phases 1 and 2: A consensus report http://www.nysna.org/images/pdfs/practice/nqf_ana_outcomes_draft10.pdf. Accessed August 19, 2010.

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

Measure Score Validity-Validity as Assessed by External Groups

Among the 8 TEP members who provided a response, 1 responded “Somewhat Agree,” 3 responded “Moderately Agree,” and 4 reported “Strongly Agree” that this measure accomplished the purposes of measuring payments for Medicare patients for a 30-day HF episode of care, removing policy adjustments unrelated to care decisions, risk-adjusting based upon case mix, and providing CMS with a tool that it can use to compare payments across hospitals and identify hospitals with notably higher and lower payments.

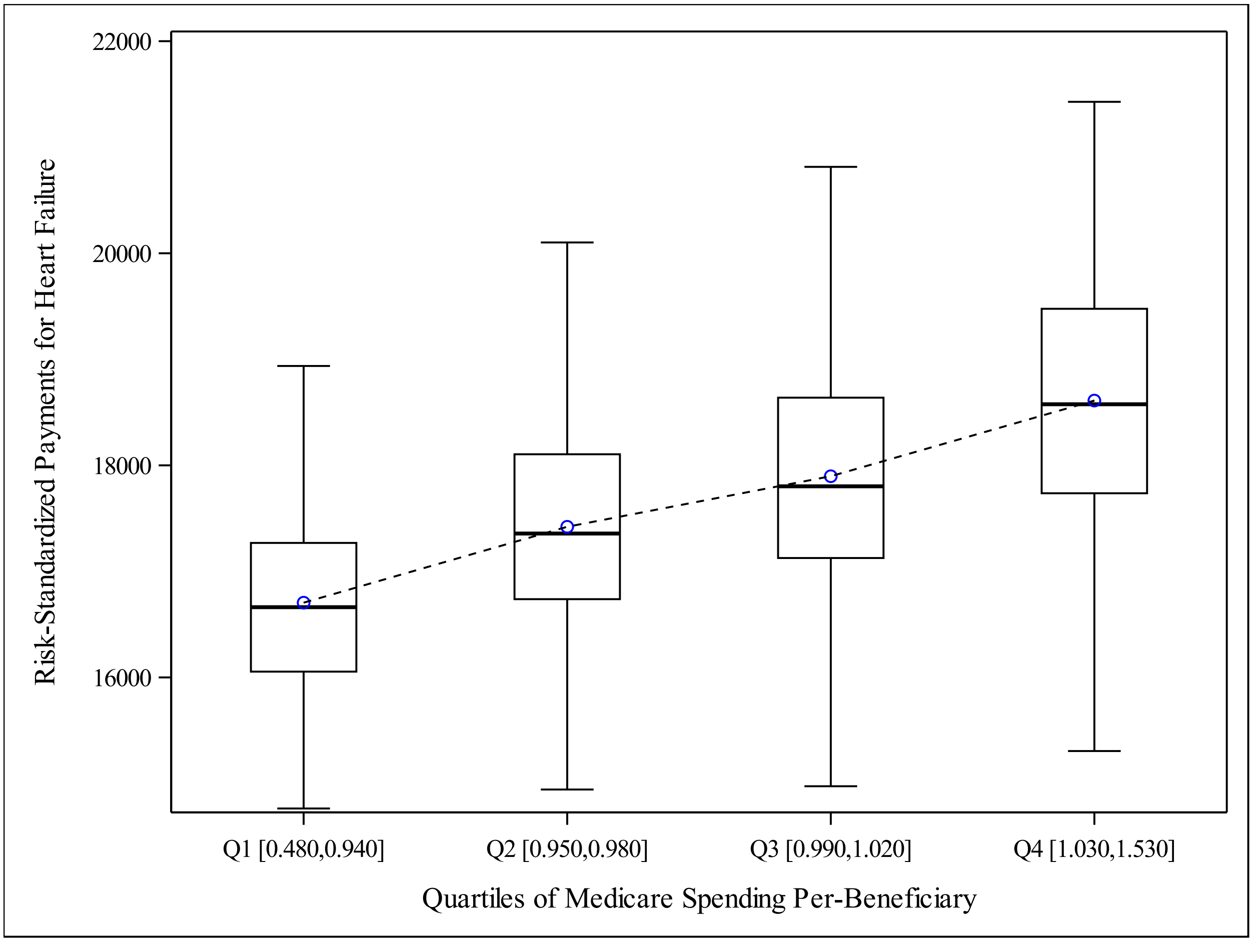
External Empiric Validity

Correlations with Medicare Spending per Beneficiary

The results of the correlation analyses described in section 2b1.2 are shown below in Figure 1. As expected, the HF payment measure score was positively correlated with the Medicare Spending Per Beneficiary (MSPB) measure, with a correlation coefficient of 0.543 meaning that higher spending across all Medicare FFS beneficiaries correlated with higher spending on patients hospitalized with HF.

Figure 1 shows the box-whisker plots of the HF risk-standardized payment (RSP) within each quartile of the MSPB measure score. The blue circles represent the mean RSPs of HF payment score quartiles. The correlation between HF RSPs and the MSPB score is 0.543, which suggests that hospitals with higher RSPs are more likely to have higher MSPB measure scores.

Figure 1. Box-whisker plots of HF payment RSPs within each quartile of the Medicare Savings Per Beneficiary (MSPB) measure score



Disposition of Payments

Below we show the disposition of payments for the observed outcomes (inpatient and post-acute care), within quartiles of the provider RSP (Table 2).

Table 2. Subcategories of observed payments within the HF Payment Quartiles of Provider RSPs

Description	1st Quartile of RSPs	2nd Quartile of RSPs	3rd Quartile of RSPs	4th Quartile of RSPs
Total Number of Patients in Each RSP Quartile	192,503	215,538	270,009	309,177
Total Observed Episode Payment per Patient	$15,410	$16,768	$17,802	$19,592
Index Inpatient Payment/Patient	$9,688	$10,211	$10,755	$11,609
Index Inpatient Facility Payment/Patient	$8,586	$8,943	$9,287	$9,844
Index Inpatient Physician Payment/Patient	$1,103	$1,268	$1,468	$1,765
Patient with PAC%	95.9	96.3	96.4	96.3
PAC Payment/Patient	$6,036	$6,910	$7,426	$8,497

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 Payment measure is supported by three types of evidence: face validity results derived from a systematic survey of a Technical Expert Panel (TEP), empiric validity demonstrated by correlations, and internal consistency.

The validity of the HF Payment measure is supported by face validity as indicated by the Technical Expert Panel (TEP) vote. There was unanimous TEP support for the face validity of the measure: 8 of 8 TEP members strongly, mostly, or somewhat agreed with the validity statement.

The validity of the measure is further supported by the empiric evidence that shows a correlation in the expected strength and direction with a related and valid payment measure.

Finally, the observed payment breakdowns appropriately align with the distribution of the provider-level risk-standardized payments.

____

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 impact of exclusions on the cohort, we examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion (EM

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

2b2.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Table 3 shows the distribution of exclusions (based on the EM Testing Dataset), among hospitals with 25 or more admissions.

Table 3. HF Payment: Number, percent, and distribution of exclusions among hospitals with 25 or more admissions.

Exclusion	N	%	Distribution across hospitals (N=2,311 prior to applying exclusion criteria: Min, 25th, 50th, 75th percentile, Max)
1. Discharged alive on the day of admission or the following day who were not transferred to another	59,929	4.24	(0.00,2.38,4.30,7.35,35.0)
2. Unreliable data	25	<0.01	(0.00,0.00,0.00,0.00,3.13)
3. Incomplete administrative data in the 30 days following the start of the index admission if discharged alive	58,127	4.11	(0.00,3.07,4.17,5.63,24.0)
4. Enrolled in the Medicare hospice program any time in the 12 months prior to the index	17,779	1.26	(0.00,0.22,1.01,1.89,18.8)
5. Discharged against medical advice (AMA)	7,686	0.54	(0.00,0.00,0.25,0.77,9.42)
6. Transferred to a federal hospital	610	0.04	(0.00,0.00,0.00,0.00,17.0)
7. not matched to admission in the HF mortality measure	18,064	1.28	(0.00,0.00,0.78,1.77,93.0)
8. Missing index DRG weight where provider received no payment	0	0	n/a
9. LVAD or transplant	4,625	0.33	(0.00,0.00,0.00,0.14,8.50)

After exclusions #1-9 are applied, the measure randomly selects one index admission per patient per year for inclusion in the cohort so that each episode of care is mutually independent. Additional admissions within that year are excluded; 267,798 admissions, or 18.95% of the cohort, were excluded in this step. For the three-year combined data, when index admissions occur during the transition between measure reporting periods (June and July of each year) and both are randomly selected for inclusion in the measure, the measure includes only the June admission. July admissions within the 30-day outcome window of the June admission are excluded to avoid assigning payments for the same claims to two admissions. There were 1,707 admissions in July, representing 0.12% of the cohort.

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data* *collection and analysis.* ***Note****:* ***If patient preference is an exclusion****, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Exclusion 1: Patients who were discharged alive on the day of admission or the following day who were not transferred to another acute care facility. This exclusion accounts for 4.24% of all index admissions excluded from the initial index cohort. This exclusion represents the majority of all exclusions, and is meant to ensure a clinically coherent cohort. This exclusion prevents inclusion of patients who likely did not have clinically significant HF.

Exclusion 2: Patients with inconsistent or unknown vital status or other unreliable demographic [age and gender] data. This exclusion accounts for <0.01% of all index admissions excluded from the initial index cohort. We do not include stays for patients where the age is greater than 115, where the gender is neither male nor female, where the admission date is after the date of death in the Medicare Enrollment Database, or where the date of death occurs before the date of discharge but the patient was discharged alive.

Exclusion 3: Patients with incomplete administrative data in the 30 days following start of index admission if discharged alive. This exclusion accounts for 4.11% of all index admissions excluded from the initial index cohort. This is necessary in order to identify the outcome (payments) in the sample over our analytic period.

Exclusion 4: Patients enrolled in the Medicare hospice program or used VA hospice services any time in the 12 months prior to the index admission, including the first day of the index admission. This exclusion accounts for 1.26% of all index admissions excluded from the initial index cohort. These patients are likely continuing to seek comfort measures only; thus, mortality is not necessarily an adverse outcome or signal of poor quality care.

Exclusion 5: Patients who are discharged AMA. This exclusion accounts for 0.6% 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.

Exclusion 6: Patients transferred to a federal hospital. This exclusion accounts for 0.04% of all index admissions excluded from the initial index cohort. We do not have claims data for these hospitals; therefore, including these patients would systematically underestimate payments.

Exclusion 7: Patients whose claims are not matched to admission in the HF mortality measure. This exclusion accounts for 1.2% of all index admissions excluded from the initial index cohort. As part of the current data processing, we match our index HF admissions to the HF mortality cohort to obtain the risk-adjustment variables. Patients are excluded if they cannot be matched between the HF payment and HF mortality cohorts.

Exclusion 8: (patients whose claims are missing index DRG weight where provider received no payment) accounts for 0.00% of all index admissions excluded from the initial index cohort. With neither DRG weight nor payment data, we cannot calculate a payment for the patient's index admission; this would make the entire episode of care appear significantly less expensive.

Exclusion 9: Patients with a procedure code for left ventricular assist device (LVAD) implementation or heart transplant either during the index admission or in the 12 months prior to the index admission. This exclusion accounts for 0.33% of all index admission excluded from the initial index cohort. These patients are excluded since 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 30 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?

The goal of risk adjustment for this measure is to account for patient age and comorbid conditions that are clinically relevant and have strong relationships with the outcome while illuminating important payment differences between hospitals. The measure adjusts for case-mix differences based on the comorbidities of the patient at the time of index admission. Conditions that may represent adverse outcomes due to care received during the index admission are not considered for inclusion in risk adjustment. Although they may increase the risk of mortality and complications, including them as covariates in risk adjustment could attenuate the measure's ability to characterize payments influenced by care delivered by hospitals.

The candidate variables for the model are derived from secondary diagnoses of the index hospital stay (excluding potential complications), inpatient Part A data, outpatient hospital data, and Part B carrier files for physician, radiology and laboratory services during the 12 months prior to the index hospital stay.

For candidate variable selection using the development sample (A1; random 50% of 2009 data), we started with the 189 Condition Categories (CCs). We used the ICD-9-to-CC assignment map, which is maintained by CMS and posted at www.qualitynet.org. To select candidate variables, a team of clinicians reviewed all 189 CCs and excluded those that were not relevant to the Medicare population or that were not clinically relevant to the HF payment outcome (e.g., attention deficit disorder, female infertility). Clinically relevant CCs were selected as candidate variables; some of these CCs were combined into clinically coherent groups. We also adjusted for age.

To inform variable selection, we performed a modified approach to stepwise generalized linear model regression. We used sample A1 to create 1,000 bootstrap samples. For each sample, we ran a generalized linear model that included all candidate variables. The results were summarized to show the percentage of times that each of the candidate variables was significantly associated with HF payment (at the p<0.05 level) in the 1,000 bootstrap samples (e.g., 70% would mean that the candidate variable was significant at p<0.05 in 70% of the bootstrap).

The working group reviewed these results and decided to retain all risk-adjustment variables above a 90% cutoff (i.e., to retain variables that were significant at the p<0.05 level in at least 90% of the bootstrap samples). We chose the 90% cutoff because variables above this threshold demonstrated a relatively strong association with HF payment and were clinically relevant.

Causal Pathways for Social Risk Variable Selection

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, (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) as well as race. 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 (Roshanghalb et al., 2019; Alghanem et al., 2020). 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; Jha et al., 2013).

Conceptual Framework

The relationship between social risk factors and episode-of-care payment is complex and not well understood. While patients with social risk factors might have have lower utilization because of reduced access to care, they may also have higher overall utilization of care due to worse outcomes.

Additionally, it is important to consider whether costs associated with patients with social risk factors influences outcomes. For example, if there is a pattern showing that increased spending results in better outcomes, it might be appropriate to risk adjust. However, given the complex relationships between social risk, costs, and outcomes, if there is no consistent association between payment and quality then it may not be appropriate to risk adjust.

Potential Mechanisms by which Social Risk Factors Affect Costs

Potential causal mechanisms by which social risk factors influence costs following an admission for heart failure are varied and complex. Few studies have assesed the relationship between patient social risk factors (e.g., gender, SES and race) and payment associated with heart failure, and few studies directly address the complex causal pathways. Our literature review has identified four potential mechanisms at the patient- and hospital-level: (1) Health at admission and other patient characteristics, (2) selection of patients into different quality hospitals, (3) care within the hospital, and (4) post-discharge care.

1. **Health at admission and other patient characteristics**

Patients with social risk factors such as low SES may have more comorbid conditions at the time of admission related to historical or lifelong social disadvantage. For example, research shows that patients with social risk factors can have worse health overall, and therefore at the time of admission. For heart failure specifically, Medicare patients with more comorbidities have higher rates of follow-up visits and readmission, and there are known differences in comorbidities between different races (Grahm, 2015). However this measure risk adjusts for comorbidities to account for health at admission.

2. **Selection of patients into different quality hospitals**

Some studies examining the link between social risk factors and costs suggest that the relationship can be mediated by hospital quality. Patients with social risk factors may be more likely to live near to and be admitted to lower quality hospitals. For heart failure, however, a recent study found that between-hospital differences readmission rates were small at hospitals treating a minimum volume of patients within different race and neighborhood-income subgroups (Downing et al., 2018).

Low- and high-quality hospitals can both contribute to increased episode-based costs. For example, care at low-quality hospitals may be associated with higher costs because lower-quality of care may require more frequent and intense follow-up care (such as a readmission). But care delivered at high-quality hospitals could be more costly, for example when high-quality, evidence-based care involves expensive treatments or procedures. In addition, both high- and low-quality hospitals also have the potential to deliver higher-cost but low-value care. In general the relationship between care quality and cost has been show to be inconsistent (Hussey, 2013), and in the case of heart failure more recent sudies have shown either a weak inverse association or no relationship between cost and quality (Desai et al., 2018; Krumholz et al., 2019).

It has been demonstrated, however, that hospital performance related to adverse events was associated with hospital-specific risk-standardized 30-day episode-of-care expenditures for patients with heart failure (Wang et al., 2020). In addition, a 2019 study authored by the developer showed that differences in hospital-level payments for heart failure were associated with hospital characteristics independently from patient characteristics (Krumholz et al., 2019). In this study the authors compared payments for the same Medicare patient for two admissions for the same condition – one admission to a low-payment hospital and one admission to a high-payment hospital and found that patients who were admitted to hospitals with the highest payment profiles incurred higher costs than when they were admitted to hospitals with the lowest payment profiles. The findings suggest that that variations in payments to hospitals are, at least in part, associated with the hospitals independently of non–time-varying patient characteristics.

3. Care within the hospital

Social risk factors can contribute to costs if patients do not receive equivalent or patient-centered care within a facility. For example, a study using linked hospital and census data found that low income or minority patients may experience differential, lower quality, or discriminatory care within a given facility (Trivedi 2014). Alternatively, patients with social risk factors may require and not necessarily receive differentiated care, such as provision of lower literacy information. For example, hospitals may provide the same care for all patients (e.g. the same discharge instructions) but this care might be insufficient for patients with social risk factors (e.g. due to low literacy). Failure to meet the needs of socially disadvantaged patients can lead to costly complications requiring readmission.

Specifically for heart failure, a recent study found no evidence of significant within-hospital differences in in utilization as measured by readmission rates for patients from lower-income neighborhoods compared with those from higher-income neighborhoods (Downing et al., 2018). This study suggests that any differences in readmission rates by race and neighborhood income may be systemic, rather than localized within particular hospitals.

4. Post-discharge care

Social risk factors can contribute to costs if patients receive or have access to more or less high-value post-discharge care. As mentioned in section 1.8 above, studies have shown that black and Hispanic Medicare patients with heart failure were less likely to have a follow-up clinic visit but more likely to experience a readmission within 30 days of hospital discharge, compared with white patients (DeLia et al., 2014). In addition, patients with Medicare Advantage plans (considered by the study authors as a possible proxy for income) were much less likely to have a follow-up visit and also more likely to experience a readmission within 30 days; the inverse was true for patients with commercial insurance. Medicare patients with additional Medicaid coverage were much less likely to have a follow-up visit but also less likely to experience a readmission. However, a recent study found that readmission rates within hospitals were only slightly higher for patients in lower-income neighborhoods compared with higher-income neighborhoods (Downing et al., 2018).

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 are 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.

Note that while race has been used in studies of heart failure outcomes and utilization, NQF has discouraged the use of race variables in social risk factor testing. Therefore, we do not present those results.

Based on this model and the considerations outlined above and 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

References

Alghanem F, Clements JM. Narrowing performance gap between rural and urban hospitals for acute myocardial infarction care. Am J Emerg Med. 2020 Jan;38(1):89-94.

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.

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.

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)

DeLia, D., Tong, J., Gaboda, D., & Casalino, L. P. 2014. Post-discharge follow-up visits and hospital utilization by Medicare patients, 2007-2010. Medicare & Medicaid research review, 4(2).

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.

Desai, N. R., Ott, L. S., George, E. J., Xu, X., Kim, N., Zhou, S., Hsieh, A., Nuti, S. V., Lin, Z., Bernheim, S. M., & Krumholz, H. M. (2018). Variation in and Hospital Characteristics Associated With the Value of Care for Medicare Beneficiaries With Acute Myocardial Infarction, Heart Failure, and Pneumonia. JAMA network open, 1(6), e183519.

Downing, N. S., Wang, C., Gupta, A., Wang, Y., Nuti, S. V., Ross, J. S., Bernheim, S. M., Lin, Z., Normand, S. T., & Krumholz, H. M. (2018). Association of Racial and Socioeconomic Disparities With Outcomes Among Patients

Hospitalized With Acute Myocardial Infarction, Heart Failure, and Pneumonia: An Analysis of Within- and Between-Hospital Variation. JAMA network open, 1(5), e182044.

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.

Graham, Garth. "Disparities in cardiovascular disease risk in the United States." Current cardiology reviews vol. 11,3 (2015): 238-45.

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.

Hayes, B. H., Haberling, D. L., Kennedy, J. L., Varma, J. K., Fry, A. M., & Vora, N. M. (2018). Burden of Pneumonia-Associated Hospitalizations: United States, 2001-2014. Chest, 153(2), 427–437.

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

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

Hussey PS, Wertheimer S, Mehrotra A. The association between health care quality and cost: a systematic review. Ann Intern Med. 2013;158(1):27-34.

Huckfeldt, P. J., Mehrotra, A., & Hussey, P. S. (2016). The Relative Importance of Post-Acute Care and Readmissions for Post-Discharge Spending. Health services research, 51(5), 1919–1938.

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.

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.

Krumholz, H. M., Wang, Y., Wang, K., Lin, Z., Bernheim, S. M., Xu, X., Desai, N. R., & Normand, S.T. 2019. Association of Hospital Payment Profiles With Variation in 30-Day Medicare Cost for Inpatients With Heart Failure or Pneumonia. JAMA network open, 2(11), e1915604.

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

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

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

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

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

Trivedi AN, Nsa W, Hausmann LRM, et al. Quality and Equity of Care in U.S. Hospitals. New England Journal of Medicine. 2014;371(24):2298-2308.

Wang Y, Eldridge N, Metersky ML, Sonnenfeld N, Rodrick D, Fine JM, Eckenrode S, Galusha DH, Tasimi A, Hunt DR, Bernheim SM, Normand ST, Krumholz HM. Association Between Medicare Expenditures and Adverse Events for Patients With Acute Myocardial Infarction, Heart Failure, or Pneumonia in the United States. JAMA Netw Open. 2020 Apr 1;3(4):e202142.

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

☒ **Published literature**

☐ **Internal data analysis**

☐ **Other (please describe)**

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

The table below shows the final variables in the model in the testing dataset with associated payment ratios and 95 percent confidence intervals (CI) calculated with the EM Testing Dataset.

Table 4. Payment Ratios and 95% Confidence intervals for risk variables in the HF Payment measure model

Risk Variable	Payment Ratios (95% CI) 07/2016-06/2019
Age (>=85)	Reference
Age (65 - 74)	1.06 (1.06 - 1.07)
Age (75 - 84)	1.05 (1.05 - 1.05)
Severe infection (CC 1, 3-6)	1.05 (1.03 - 1.06)
Other infectious diseases (CC 7)	1.02 (1.02 - 1.02)
Protein-calorie malnutrition (CC 21)	1.11 (1.10 - 1.11)
Morbid obesity; other endocrine/metabolic/nutritional disorders (CC 22, 25-26)	1.00 (1.00 - 1.01)
Other significant endocrine and metabolic disorders (CC 23)	1.05 (1.04 - 1.05)
Other gastrointestinal disorders (CC 38)	1.01 (1.00 - 1.01)
Bone/joint/muscle infections/necrosis (CC 39)	1.05 (1.04 - 1.06)
Other musculoskeletal and connective tissue disorders (CC 45)	1.00 (1.00 - 1.01)
Delirium and encephalopathy (CC 50)	1.02 (1.01 - 1.02)

Risk Variable	Payment Ratios (95% CI) 07/2016-06/2019
Dementia or other specified brain disorders (CC 51-53)	1.03 (1.03 - 1.03)
Severe mental illness (CC 57-58)	1.03 (1.02 - 1.03)
Other psychiatric disorders (CC 63)	1.01 (1.01 - 1.01)
Respiratory arrest/cardiorespiratory failure/respirator dependence (CC 82-84 plus ICD-10-CM codes R09.01 and R09.02, for discharges on or after October 1, 2015; CC 82-84 plus ICD-9-CM diagnosis codes 799.01 and 799.02, for discharges prior to October 1, 2015)	1.01 (1.01 - 1.02)
Coronary atherosclerosis or angina (CC 88-89)	1.03 (1.03 - 1.03)
Heart infection/inflammation, except rheumatic (CC 90)	1.07 (1.06 - 1.08)
Major congenital cardiac/circulatory defect (CC 92)	1.08 (1.03 - 1.13)
Hypertension (CC 95)	0.98 (0.97 - 0.98)
Specified arrhythmias and other heart rhythm disorders (CC 96-97)	0.97 (0.96 - 0.97)
Precerebral arterial occlusion and transient cerebral ischemia; cerebral atherosclerosis and aneurysm; cerebrovascular disease, unspecified (CC 101-102)	1.01 (1.01 - 1.02)
Vascular or circulatory disease (CC 106-109)	1.01 (1.01 - 1.02)
Pneumonia (CC 114-116)	1.06 (1.06 - 1.06)
Other ear, nose, throat, and mouth disorders (CC 131)	0.99 (0.98 - 0.99)
Dialysis status (CC 134)	1.09 (1.09 - 1.10)
Renal failure (CC 135-140)	1.11 (1.11 - 1.12)
Decubitus ulcer of skin (CC 157-160)	1.04 (1.03 - 1.05)
Chronic ulcer of skin, except pressure (CC 161)	1.06 (1.06 - 1.07)
Cellulitis, local skin infection (CC 164)	1.01 (1.01 - 1.01)
Hip fracture/dislocation (CC 170)	1.03 (1.02 - 1.03)
Internal injuries (CC 172)	1.04 (1.03 - 1.05)

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.**

To understand the impact of adjusting for social risk factors we examined the following: the prevalence of the social risk factor in the patient cohort, the observed outcome for patients with and without social risk factors, and the incremental effects of adding social risk variables. These analyses and their results are described below.

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 social risk factors in the cohort

The prevalence of social risk factors in the HF cohort varies across measured entities (Table 5). The median percentage of dual-eligible patients was 16.2% (interquartile range of 10.0%-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 [lowest quartile]) was 18.3% (interquartile range of 6.30%-35.9%).

Table 4. Variation in prevalence of each social risk factor across measured entities

Social Risk Factors	Median hospital prevalence of the social risk factor (IQR)
Dual Eligible	16.2% (10.0%-25.1%)
Low AHRQ SES	18.3% (6.30%-35.9%)

Observed outcome rates in patients with and without social risk factors

Mean observed patient-level HF payments are similar for patients with and without social risk factors (Table 6). Mean observed payments are $930 higher for dual-eligible patients compared with non-dual enrolled patients ($18,440 vs. $17,511) and mean observed payments were $96 lower for patients with low AHRQ SES compared with patients without low AHRQ SES Index.

Table 5. Observed payments for patients with and without social risk

Social Risk Factors	Mean observed payments with (and without) the social risk factor
Dual Eligible	$18,440 (vs. $17,511)
Low AHRQ SES	$17,610 (vs. $17,705)

Incremental effects of SRF variables in a multivariable model

We then examined the strength and significance of the two SRFs, patients' low SES and dual-eligibility, in the context of a multivariable model. Table 7 shows the payment ratios for each social risk factor when the two

SRFs are added one at a time and together, along with the clinical risk factors included in the original risk-adjustment model. (Note that a payment ratio of less than one means lower payments.) The payment ratio for the low AHRQ SES variable is 0.98 when added to the model with the clinical risk factors; the payment ratio for the dual eligibility variable is 1.01. When both variables are added to the model, the payment ratio for the low AHRQ SES variable is 0.98 and payment ratio for the dual eligibility variable is unchanged.

Table 6. Strength and significance of social risk factor variables

Model	Variable	Payment Ratio	p-values
Base model plus Low AHRQ SES variable	Low AHRQ SES	0.98	<0.0001
Base model plus Dual Eligibility variable	Dual Eligibility	1.01	0.0055
Base Model plus Low AHRQ SES and Dual Eligibility variables	Low AHRQ SES	0.98	<0.0001
Base Model plus Low AHRQ SES and Dual Eligibility variables	Dual Eligibility	1.01	<0.0001

We also evaluated the impact on model performance (Table 8) and find that the R-squared values for each version of the model (with each social risk factor separately, and then together) were similar.

Table 7. Model performance with and without social risk factors

Model	Quasi-R-square
Base Model	0.031
Base Model plus Low AHRQ SES	0.032
Base Model plus Dual Eligibility	0.031
Base Model plus both Low AHRQ SES and Dual Eligibility	0.032

Impact on measure scores

We then examined the impact of adding each social risk factor separately on measure scores. As shown in Table 9, we found that when adding the low AHRQ SES variable to the model, the median change in measure scores (risk-standardized payments or RSPs) was very small, and in a negative direction: -$7.55 (interquartile range [IQR] (-$97.80 – $77.50). When the dual eligibility variable was added to the model the median change in hospitals' RSPs was also small: $1.30 (interquartile range [IQR] (-$2.50 – $5.00).

To further characterize the impact on measure scores, we examined the correlation between measure scores (risk-standardized payments or RSPs) calculated with the baseline model and with either social risk factor included in the model. The results show that measures scores were highly correlated (Table 9): the correlation coefficient between RSPs for each hospital with and without the low AHRQ SES variable is 0.989; the correlation coefficient between RSPs for each hospital with and without the dual eligibility variable is >0.999.

These results demonstrate that overall, risk adjustment for either social risk variable has a small impact on measure scores.

Table 8. Distributions of changes in measure scores and correlations between the measures scores based on models with and without adjustment for each social risk factor

Metric	**Change in measure score ($)**	**Change in measure score ($)**	**Measure Score Correlation**
Social Risk Factor	**Median**	**IQR**	**Pearson Correlation Coefficient**
Low AHRQ SES	-$7.55	(-$97.80 – $77.50)	0.989
Dual Eligibility	$1.30	(-$2.50 – $5.00)	>0.999

Social Risk Factor Summary

The analyses presented above show that patients with either of two social risk factors (low AHRQ SES Index or dual eligibility) have slightly higher payments, in the case of dual eligibility, or slightly lower payments, in the case of low AHRQ SES, after adjusting for other risk factors in a multivariable model. However, adding the social risk variables results in little impact on model performance, little change in measure scores, and measure scores estimated for hospitals with and without dual eligibility are highly correlated.

As presented in the conceptual model (section 2b3.3a), the relationship between social risk and payment may reflect that patients with social risk factors are receiving differential care within hospitals (for example, fewer treatments or interventions), 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.

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 and the use of the measure (in a payment program or for public reporting). The HF Payment measure is not in a payment program; the measure is used only in public reporting. In addition, the Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation's (ASPE's) recommends that resource use measures that are used for public reporting should not be adjusted for social risk (ASPE 2020).

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 (such as reduced delivery of evidence-based care), adjusting will mask the disparity in care. In contrast, an unadjusted measure will illuminate payment differences and create an incentive to mitigate them (although they need to be interpreted in context and with additional information, such as clinical quality). Not adjusting, however may disadvantage providers who care for dual eligible patients, and unintentionally create an incentive for hospitals to care for fewer patients with social risk factors, potentially reducing access care. However, not adjusting for low AHRQ SES Index may disadvantage hospitals who care for fewer patients with low AHRQ SES. CMS considers these risks 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. In addition, the paired HF mortality measure, which can be used in conjunction with the HF payment measure to elucidate value, is also not adjusted for social risk. Please note that the HF Payment measure was part of NQF's 2015 SDS Trial Period and was last re-endorsed in 2016 without adjustment for social risk factors.

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*)

Risk-Adjustment Model Development and Validation in Medicare FFS:

Our goal was to develop a parsimonious model that accounted for differences in patient case mix at the time of index admission that were strongly associated with total payment for a HF 30-day episode of care. The candidate variables for the model were derived from secondary diagnoses of the index hospital stay (excluding potential complications), inpatient data, outpatient hospital data, and carrier files for physician, radiology and laboratory services during the 12 months prior to the index hospital stay. To select candidate variables, we started with the 189 CCs. We used the ICD-9-to-CC assignment map, which is maintained by CMS. A team of clinicians reviewed all 189 CCs and excluded those that were not relevant to the Medicare population or not clinically relevant to the HF payment outcome (for example, attention deficit disorder and female infertility). Some of these CCs were combined into clinically coherent groups. The remaining clinically relevant CCs, along with age were selected as candidate comorbid risk variables.

As is typical with data for healthcare payments, our dependent variable – total payment for a HF 30-day episode of care – is both right-skewed and leptokurtotic (skewness= 2.9; kurtosis = 15.2). We Winsorized payments at the 99.5th percentile to improve model performance and prevent drastically altering the performance of hospitals with an unrepresentative and expensive outlier patient. This reduced the skewness (2.2) and kurtosis (6.1) of the data. To address estimation problems that can arise with non-normally distributed data, we employed the algorithm suggested by Manning & Mullahy. Using this algorithm, we compared several alternative models in order to determine the best estimation approach. Based on these assessments, we chose to estimate a generalized linear model with a log link and a Gamma distribution.

Approach to Assessing Model Performance: During model development, we computed four summary statistics for assessing model performance using the development (A1; random 50% sample of 2008) and validation (A2; remaining 50% of 2008) cohorts:

(1) R-squared

(2) Over-fitting indices (Calibration γ0, γ1)

(3) Distribution of Standardized Pearson Residuals

(4) Predictive ratios

Approach to Annual Model Validation

CORE's measures undergo an annual measure reevaluation process, which ensures that the risk-standardized payment 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. Below we describe, for 2020 public reporting, the modifications to the payment measure:

- Updated the ICD-10 code-based specifications used in the measures. 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 the risk models; 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;
- Updated the measures' SAS analytic packages (SAS packs) and documentation;
- Evaluated and validated model performance for the three years combined; and,
- Evaluated the stability of the risk-adjustment model 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.

References:

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

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b3.9

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

R^2 results

Sample A1 – 0.035

Sample A2 – 0.034

Sample A3 – 0.027

The updated R^2 results calculated with the EM testing dataset were: 0.031

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

Over-fitting indices

Sample A1 – 0,1

Sample A2 – 0.10,0.99

Sample A3 – 1.24, 0.87

Standardized Pearson Residuals lack of fit

<-2 = A1 0.00%; A2 0.00%; A3 0.00%

[-2, 0) = A1 64.47%; A2 64.52%; A3 64.68%

[0, 2) = A1 32.87%; A2 32.80%; A3 32.62%

[2+ = A1 2.68%; A2 2.68%; A3 2.70%

2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

Below is the plot of predictive ratios by decile and top 1% of predicted payment for the development and validation samples:

Bottom Decile:	A1 1.01; A2 1.01; A3 1.02
First Decile:	A1 1.02; A2 1.03; A3 1.02
Second Decile:	A1 1.02; A2 1.02; A3 1.01
Third Decile:	A1 1.00; A2 1.01; A3 1.00
Fourth Decile:	A1 0.99; A2 0.99; A3 1.00
Fifth Decile:	A1 1.00; A2 0.99; A3 1.00
Sixth Decile:	A1 0.98; A2 0.98; A3 0.99
Seventh Decile:	A1 0.99; A2 0.99; A3 0.99
Eighth Decile:	A1 0.99; A2 0.98; A3 1.00
Ninth Decile:	A1 1.00; A2 1.00; A3 0.99
Tenth Decile:	A1 1.02; A2 1.03; A3 1.02
Top 1%:	A1 1.09; A2 1.07; A3 1.06

Table 9. Distribution of predictive ratios by decile and top and bottom 1% (EM Testing Dataset).

Decile	Predictive Ratio
Decile1	1.00
Decile2	0.99
Decile3	0.99
Decile4	0.99
Decile5	0.99
Decile6	1.00
Decile7	1.00
Decile8	1.02
Decile9	1.02
Decile10	1.03
Bottom 1%	1.01
Top 1%	1.00

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*)

R^2

For a traditional linear model (i.e. ordinary least squares regression) R^2 is interpreted as the amount of variation in the observed outcome that is explained by the predictor variables (patient-level risk factors). Generalized linear models (GLMs), however, do not output an R^2 that is akin to the R^2 of a traditional linear model. In order to provide the NQF Committee with a statistic that is conceptually similar, we produced a “quasi- R^2” by regressing the total payment outcome on the predicted outcome. 1 Specifically, we regressed the total payment on the payment predicted by the patient-level risk factors. This regression produced a quasi-R^2 of 0.035, suggesting that approximately three and a half percent of the variation in payment can be explained by patient-level risk factors. For this endorsement maintenance submission, the quasi-R^2 slightly decreased down to 0.031, suggesting that about three percent of the variation in payment could be explained by patient-level risk factors. This quasi- R^2 is in-line with R^2 from other patient-level risk adjustment models for health care payment (Pope et al., 2011)

References

Jones AM. Models for Health Care. Health, Econometrics and Data Group (HEDG) Working Papers. 2010.

Pope, G. C., Kautter, J., Ingber, M. J., Freeman, S., Sekar, R., & Newhart, C. RTI International, (2011). Evaluation of the CMS-HCC risk adjustment model (Final Report). pp.6.

Over-fitting (Calibration γ0, γ1)

Over-fitting can result in the phenomenon in which a model describes the relationship between predictor variables and the outcome well in the development sample, but fails to provide valid predictions in new patients. 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.

Standardized Pearson Residuals

Standardized Pearson residuals also assess model fit. If a substantial number of standardized Pearson residuals exceed 2 in absolute value, lack of fit may be indicated.

Predictive Ratios

A predictive ratio is an estimator's ratio of predicted outcome to observed outcome (Ash et al., 1998). A predictive ratio close to 1.0 indicates an accurate prediction. A ratio substantially greater than 1.0 indicates overprediction, and a ratio substantially less than 1.0 indicates underprediction.

Overall Interpretation

Interpreted together, our diagnostic results demonstrate the risk-adjustment model adequately controls for differences in patient characteristics (case mix). The prior results together with updated evidence provided for endorsement maintenance (quasi R^2 and predictive ratios) supports that the models continue to be valid for use with current data.

References:

Ash AS, Byrne-Logan S. How Well Do Models Work? Predicting Health Care Costs. Proceedings of the Section on Statistics in Epidemiology. American Statistical Association. 1998.

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 episode-of-care RSP for each measure is estimated using a hierarchical generalized linear model (HGLM). In brief, the approach simultaneously models data at the patient and hospital levels to account for the variance in patient outcomes within and between hospitals. At the patient level, the measures use a generalized linear model to model the total episode-of-care payment using age, selected clinical covariates, and a hospital-specific effect. For the HF Payment measure, the RSPs are estimated using a log link and inverse Gaussian distribution.

At the hospital level, the approach models the hospital-specific effects as arising from a normal distribution. The hospital effect represents the underlying episode-of-care payment at the hospital, after accounting for patient risk. The hospital-specific effects are given a distribution to account for the clustering (non-independence) of patients within the same hospital. If there were no differences among hospitals, then after adjusting for patient risk, the hospital effects should be identical across all hospitals.

The RSP is calculated as the ratio of the "predicted" payment to the "expected" payment at a given hospital, multiplied by the national mean payment. For each hospital, the numerator of the ratio is the payment predicted based on the specific hospital and its observed case mix; the denominator is the payment expected based on the nation and the specific hospital's case mix. This approach is analogous to a ratio of "observed" to "expected" used in other types of statistical analyses. It conceptually allows a particular hospital's payment, given its case mix, to be compared to an average hospital's payment for the same case mix. Thus, a ratio lower than one indicates a lower-than-expected episode-of-care payment, while a ratio higher than one indicates a higher-than-expected episode-of-care payment.

The "predicted" episode-of-care payment (the numerator) is calculated using the coefficients estimated by regressing the risk factors (found in the data dictionary) and the hospital-specific effect on the payment outcome. The estimated hospital-specific effect is added to the sum of the estimated regression coefficients multiplied by the patient characteristics. The results are summed over all patients attributed to a hospital to calculate a predicted value. The "expected" episode-of-care payment (the denominator) is obtained in the same manner, except that a common effect using all hospitals in our sample is added in place of the hospital-specific effect. The results are summed over all patients attributed to a hospital to calculate an expected value. To assess hospital payments for each reporting period, we re-estimate the model coefficients using the years of data in that period.

Multiplying the predicted over expected ratio by the national mean payment transforms the ratio into a payment amount that can be compared to the national mean payment. The HGLMs are described fully in Appendix A of the 2020 Measure Updates and Specification report, and in the original methodology report.

We characterize the degree of variation in the measure score by:

1. Reporting the distribution of the measure score and describing the variation, and
2. Presenting performance categories.

To categorize hospital payments, CMS estimates each hospital's RSP and the corresponding 95% interval estimate. CMS assigns hospitals to a payment category by comparing each hospital's RSP interval estimate to the national mean payment. Comparative payments for hospitals with 25 or more eligible cases are classified as follows:

- "Less than the National Average Payment" if the entire 95% interval estimate surrounding the hospital's RSP is lower than the national mean payment.

- "No Different than the National Average Payment" if the 95% interval estimate surrounding the hospital's RSP includes the national mean payment.
- "Greater than the National Average Payment" if the entire 95% interval estimate surrounding the hospital's RSP is higher than the national mean payment.
- If a hospital has fewer than 25 eligible cases for a measure, CMS assigns the hospital to a separate category: "Number of Cases Too Small." This category is used when the number of cases is too small (fewer than 25) to reliably estimate the hospital's RSP. If a hospital has fewer than 25 eligible cases, the hospital's RSP and interval estimate will not be publicly reported for the measure.

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*)

The distribution of measure scores across hospitals shows that there are meaningful differences. The range of risk-standardized payments across the 4,502 hospitals with a measure score $13,171-$27,996. Hospitals in the 10th percentile have risk-standardized payments that are about 8.5% lower than the median; hospitals in the 90th percentile have payments that are about 10.6% higher than the median.

Figure 2: Distribution of HF Risk-Standardized Payment Measure Scores

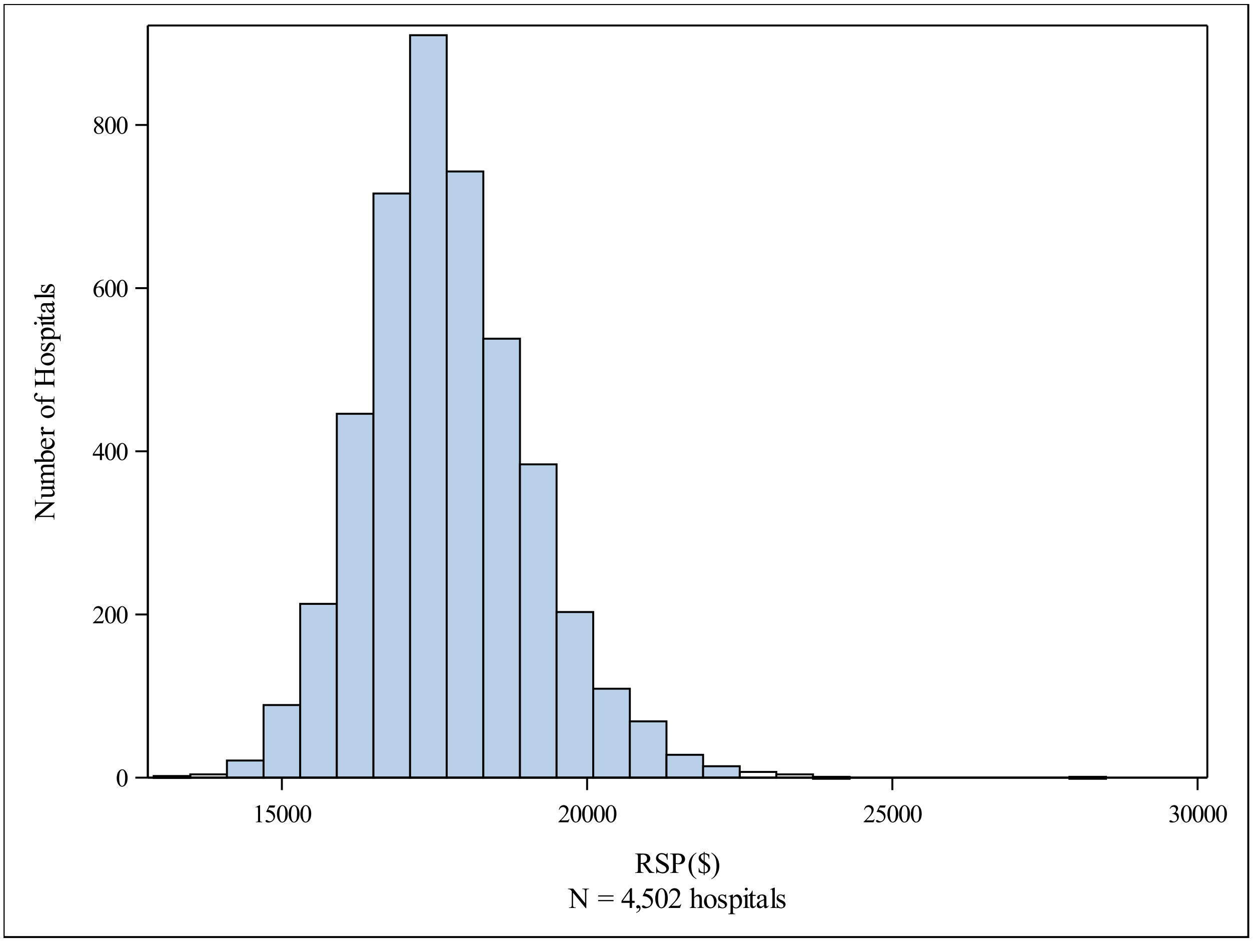


Table 10. Distribution of HF Risk-Standardized Payment Measure Scores

Characteristic	07/2016-06/2019
Number of hospitals	4,502
Mean (SD)	$17,722 ($1,368)
Range (min. – max.)	$13,171-$27,996
10th percentile	$16,106
25th percentile	$16,817
50th percentile	$17,607
75th percentile	$18,513
90th percentile	$19,482

Performance Categories

Of 4,502 hospitals in the study cohort, 409 had a payment "Less than the National Average Payment," 2,515 had a payment "No Different than the National Average Payment," and 542 had a payment "Greater than the National Average Payment." 1,036 were classified as "Number of Cases Too Small" (fewer than 25) to reliably estimate the hospital's RSP.

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 rates and the proportion of outliers suggests that there are meaningful differences across hospitals in risk-standardized payments associated with a 30-day episode of care for patients with HF.

____

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

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

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

2b5.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

N/A

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

N/A

2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i*.e., what do the results mean and what are the norms for the test conducted*)

N/A

____

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

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

2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse;* ***if no empirical sensitivity analysis****, identify the approaches for handling missing data that were considered and pros and cons of each*)

N/A

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders) 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

Feasibility

F.1. 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).

F.1.1. 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:

F.2. 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.

F.2.1. To what extent are the specified data elements available electronically in defined fields (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*)

ALL data elements are in defined fields in electronic claims

F.2.1a. 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.

F.2.2. If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.

Attachment:

F.3. 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.

F.3.1. Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

Using administrative claims variables for risk adjustment

This measure uses variables from claims data submitted by hospitals for payment, data from Medicare fee schedules, data from Final Rules for Medicare prospective payment systems and payment policies, and CMS-published wage index data. Prior research has demonstrated that administrative claims data can be used to develop risk-adjusted outcomes measures for both mortality and readmission following hospitalization for acute myocardial infarction[1,2], heart failure[3,4] and pneumonia[5,6] and that the models produce estimates of risk-standardized rates that are very similar to rates estimated by models based on medical record data. This high level of agreement supports the use of the claims-based risk-adjusted models for public reporting. The models have also demonstrated consistent performance across years of claims data.

The approach to gathering risk factors for patients also mitigates the potential limitations of claims data. Because not every diagnosis is coded at every visit; for Medicare FFS patients we use inpatient, outpatient, and physician claims data for the year prior to admission, and diagnosis codes during the index admission, for risk-adjustment. The 1-year time frame provides a more comprehensive view of patients' medical histories than is provided by the secondary diagnosis codes from the index hospitalization alone. If a diagnosis appears in some visits and not others, it is included, minimizing the effect of incomplete coding. We were careful, however, to include information about each patient's status at admission and not to adjust for possible complications of the admission. Although some codes, by definition, represent conditions that are present before admission (e.g. cancer), other codes and conditions cannot be differentiated from complications during the hospitalization (e.g. infection or shock). If these are secondary diagnoses coded only in the index admission, then they are not adjusted for in the analysis.

References:

1. Krumholz HM, Wang Y, Mattera JA, Wang Y-F, Han LF, Ingber MJ, Roman S, Normand SL. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. Circulation. 2006 Apr 4;113(13):1683-92.

2. Krumholz HM, Lin Z, Drye EE, Desai MM, Han LF, Rapp MT, Mattera JA, Normand SL. An administrative claims measure suitable for profiling hospital performance based on 30-day all-cause readmission rates among patients with acute myocardial infarction. Circulation: Cardiovascular Quality and Outcomes. 2011 Mar 1;4(2):243-52.

3. Krumholz HM, Wang Y, Mattera JA, Wang Y-F, Han LF, Ingber MJ, Roman S, Normand SL. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with heart failure. Circulation. 2006 Apr 4;113(13):1693-701.

4. Keenan PS, Normand SL, Lin Z, Drye EE, Bhat KR, Ross JS, Schuur JD, Stauffer BD, Bernheim SM, Epstein AJ, Wang Y-F, Herrin J, Chen J, Federer JJ, Mattera JA, Wang Y, Krumholz HM. An administrative claims measure suitable for profiling hospital performance on the basis of 30-day all-cause readmission rates among patients with heart failure. Circulation: Cardiovascular Quality and Outcomes. 2008 Sep;1(1):29-37.

5. Bratzler DW, Normand SL, Wang Y, O´Donnell WJ, Metersky M, Han LF, Rapp MT, Krumholz HM. An administrative claims model for profiling hospital 30-day mortality rates for pneumonia patients. Public Library of Science One. 2011 Apr 12;6(4):e17401.

6. Lindenauer PK, Normand SL, Drye EE, Lin Z, Goodrich K, Desai MM, Bratzler DW, O´Donnell WJ, Metersky ML, Krumholz HM. Development, validation, and results of a measure of 30-day readmission following hospitalization for pneumonia. Journal of Hospital Medicine. 2011 Mar;6(3):142-50.

F.3.2. **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, and algorithm)?**

There are no fees associated with the use of claims-based measures.

F.3.3. **If there are any fees associated with the use of this measure as specified, attach the fee schedule here. (Save file as: F3_3_FeeSchedule)**

Usability and Use

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making.

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.

U.1.1. **Current and Planned Use**

Specific Plan for Use	Current Use (for current use provide URL)
NA	Public Reporting Care Compare https://www.medicare.gov/care-compare/ Payment Program Hospital Inpatient Quality Reporting https://qualitynet.cms.gov/inpatient/iqr

U.1.2. **For each CURRENT use, checked above, provide:**

- Name of program and sponsor

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

Public Reporting
Program Name, Sponsor: 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.

U.1.3. 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.

U.1.4. 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.

U.2.1.1. 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.

All non-federal short-term acute care hospitals (including Indian Health Service hospitals) and critical access hospitals are included in the measure calculation. However, only those hospitals with at least 25 HF admissions are 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 payments] and post-discharge costs). These reports facilitate quality improvement activities such as review of patterns of care and make visible to hospitals post-discharge costs 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 frequently asked questions (FAQs), webinars, and provides 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. For the payment measures, the hospital-specific reports additionally include a national assessment of the value of care where the distributions of hospitals in each performance category for the mortality measure are evaluated against the distributions in each of the payment measure performance categories.

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.

U.2.1.2. **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; includes 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: posted in April/May of each calendar year on QualityNet; provide a brief overview of measures and measure updates.

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.

U.2.2.1. Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1. Describe how feedback was obtained.

Feedback on Measure performance: Hospital and Stakeholder comment

The measured entities (acute care hospitals) and other stakeholders or interested parties submit questions or comments measure through an online portal on QualityNet. Experts on measure specifications, calculation, or implementation, prepare responses to those inquiries and reply directly to the query. We consider issues raised through this process about measure specifications or measure calculation in measure reevaluation.

Feedback on Measure performance: Literature Reviews

In addition, we continually 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.

U.2.2.2. Summarize the feedback obtained from those being measured.

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

For the HF payment measure, we have received the following inquiries from hospitals since the last endorsement cycle:

1. Requests for the SAS code used to calculate measure results.
2. Questions about preview reports.
3. Questions about how to interpret the outcome and how performance categories are calculated.
4. Questions about the reporting period.
5. Questions about measure specifications, including risk adjustment, outcome definition, and inclusion/exclusion criteria.

U.2.2.3. Summarize the feedback obtained from other users.

Summary of Question and Comments from Other Stakeholders

For the HF payment measure, we have received the following inquiries from other stakeholders since the last endorsement cycle:

1. Requests for the SAS code used to calculate measure results.
2. Questions about what measures are publicly reported for Critical Acccess Hospitals.
3. Request for measure specifications.
4. Questions on the reporting period.
5. Questions about measure specifications, including inclusion and exclusion criteria, outcome definition and attribution.

6. A question from a Quality Improvement Organization (QIO) about where to find information in the Hospital Specific Report.

7. Questions on the CMS program that implements the measure.

8. Questions on how the payment measure is different from the Medicare Spending Per Beneficiary (MSPB) measure.

9. Request for SDS variable analysis results.

10. Questions on inflation adjustments used in the measures.

Summary of Relevant Publications from the Literature Review

Since the last endorsement maintenance cycle, we have reviewed several articles related to HF payment. Relevant articles covered the following topics: the association between payment and mortality or adverse events and impact of case mix and race/ethnicity on hospital costs. Additional details of these studies are provided below.

Desai et al. investigated the association between 30-day mortality and 30-day risk-standardized payments (RSPs) for AMI, HF, and pneumonia, in an effort to highlight patterns of value in care, using Medicare data from 2011-2014. They found considerable variation in both mortality rates and RSPs, though only a weak inverse correlation between the two outcomes; about 25% of hospitals had lower mortality and lower RSPs, illustrating that it is possible to improve the value of care for AMI, HF, and pneumonia patients while also lowering both mortality and readmission rates. Note: among the authors of this article are individuals who are employed by or who have affiliations with CORE.

Wang et al. investigated the association between 30-day episode of care spending and inpatient adverse events. The authors found that hospital-level adverse events were associated with hospital-specific 30-day episode of care spending for AMI, HF, and pneumonia patients, illustrating that hospitals who can reduce the number of adverse events seem to be able to reduce costs as a result. These findings provide validity for measuring episodic costs in tandem with outcomes such as mortality, readmissions, and complications for AMI, HF, and pneumonia patients. Note: among the authors of this article are individuals who are employed by or who have affiliations with CORE.

Krumholz et al. studied whether variations in hospital costs are due to patient case mix or other factors. The study cohort included Medicare patients who had admissions to two different hospitals for the same principal discharge diagnosis (HF, or pneumonia). Overall, they found that patients had different costs at each hospital, demonstrating that hospitals play a significant role in influencing episodic costs of care. Note: among the authors of this article are individuals who are employed by or who have affiliations with CORE.

Using data from 2006-2014, Ziaeian et al. studied how acute care costs differ for Medicare HF patients by race/ethnicity. Although they did not find any significant differences in index admission costs among different racial/ethnic groups, 30-day readmission costs were 9% higher among black patients compared to whites. These findings highlight increased acute care costs associated with black patients - further research should examine why this might be the case and what aspects of acute care are driving this association. Reevaluation analyses could be conducted to determine if this relationship is also present in the HF payment measure specifically.

References:

Desai NR, Ott LS, George EJ, et al. Variation in and Hospital Characteristics Associated With the Value of Care for Medicare Beneficiaries With Acute Myocardial Infarction, Heart Failure, and Pneumonia. JAMA Netw Open. 2018;1(6):e183519.

Wang Y, Eldridge N, Metersky ML, et al. Association Between Medicare Expenditures and Adverse Events for Patients With Acute Myocardial Infarction, Heart Failure, or Pneumonia in the United States. JAMA Netw Open. 2020;3(4):e202142.

Krumholz HM, Wang Y, Wang K, et al. Association of Hospital Payment Profiles With Variation in 30-Day Medicare Cost for Inpatients With Heart Failure or Pneumonia. JAMA Netw Open. 2019;2(11):e1915604.

Ziaeian B, Heidenreich PA, Xu H, et al. Medicare Expenditures by Race/Ethnicity After Hospitalization for Heart Failure With Preserved Ejection Fraction. JACC Heart Fail. 2018;6(5):388-397.

U.2.3. Describe how the feedback described in 4a2.2 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. There were no questions or issues raised by stakeholders requiring additional analysis or changes to the measure since the last endorsement maintenance cycle.

U.3.1. Progress on Improvement. (Not required for initial endorsement unless available.) Performance results on this measure (current and over time) should be provided in IM.1.2 and IM.1.4.

Discuss:

- **Purpose Progress (trends in performance results)**
- **Geographic area and number and percentage of accountable entities and patients included**

The median hospital 30-day RSP for the HF payment measure for the 3-year period between July 1, 2016 and June 30, 2019 was $17,607. The median RSP decreased by 2.6% from July 2017-June 2018 (median RSP: $17,781) to July 2018-June 2019 (median RSP: $17,310). The goal of this measure, however, is not necessarily to reduce costs, but to provide a value signal when combined with quality information.

Distribution of hospital-level RSPs for each individual year:

Periods//YEAR1617//YEAR1718//YEAR1819

Number of Hospitals//4,380//4,351//4,344

Number of Admissions//322,586//333,072//331,569

Mean(SD)//17,774(985)//17,891(962)//17,427(968)

Range(Min-Max)//14,567-23,164//14,748-22,846//13,929-22,393

Minimum//14,567//14,747//13,929

10th percentile//16,654//16,786//16,313

20th percentile//17,024//17,140//16,670

30th percentile//17,277//17,408//16,947

40th percentile//17,493//17,607//17,141

50th percentile//17,670//17,781//17,310

60th percentile//17,879//18,000//17,520

70th percentile//18,131//18,245//17,795

80th percentile//18,501//18,576//18,141

90th percentile//19,029//19,169//18,675

Maximum//23,164//22,846//22,393

U.3.2. If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

N/A

U.4.1. 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 and testing. We are committed to monitoring this measure's use and assessing potential unintended consequences over time, such as the inappropriate shifting of care or coding/billing practices, increased patient morbidity and mortality, and other negative unintended consequences for patients.

U.4.2. Please explain any unexpected benefits from implementation of this measure.

N/A

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.

H.1. Relation to Other NQF-endorsed Measures

If there are 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.

H.1.1. 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

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

2158 : Medicare Spending Per Beneficiary (MSPB) Hospital

2431 : Hospital-level, risk-standardized payment associated with a 30-day episode-of-care for Acute Myocardial Infarction (AMI)

2579 : Hospital-level, risk-standardized payment associated with a 30-day episode of care for pneumonia (PN)

3474 : Hospital-level, risk-standardized payment associated with a 90-day episode of care for elective primary total hip and/or total knee arthroplasty (THA/TKA)

H.1.2. If related or competing measures are not NQF endorsed please indicate measure title and steward.

N/A

H.2. Harmonization

H.2.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):

Are the measure specifications completely harmonized?

Yes

H.2.2. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

H.3. Competing Measure(s)

H.3.1. 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

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 Center for Outcomes Research and Evaluation

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

List the workgroup/panel members' names and organizations.

Describe the members' role in measure development.

Technical Expert Panel Members:

Amanda Kowalski, PhD	Yale University
Anne-Marie Audet, MD, MSc, SM	Commonwealth Fund
David S. P. Hopkins, PhD	Pacific Business Group on Health
Donald Casey, MD, MPH, MBA	NYU Langone Medical Center
Kavita Patel, MD, MS	Engelberg Center for Health Care Reform
Lesley Curtis, PhD, MS	Duke University
Peter Bach, MD, MAPP	Memorial Sloan-Kettering Cancer Center
Peter Lindenauer, MD, MSc	Tufts University; Baystate Medical Center; Center for Quality of Care Research
Scott Flanders, MD	University of Michigan
Stephen Schmaltz, PhD, MS, MPH	Joint Commission

Terry Golash, MD	Aetna
Vivian Ho, PhD	Rice University

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? Yearly

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