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.

Purple text represents the responses from measure developers. Red text denotes developer information has changed since the last measure evaluation review. Some content in the document is from Measure Developers.

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 #: 3561

De.2. Measure Title: Medicare Spending Per Beneficiary – Post Acute Care Measure for Inpatient Rehabilitation Facilities

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

De.3. Brief Description of Measure: The Medicare Spending Per Beneficiary – Post Acute Care Measure for Inpatient Rehabilitation Facility (MSPB-PAC IRF) was developed to address the resource use domain of the Improving Medicare Post-Acute Care Transformation Act of 2014 (IMPACT Act). This resource use measure is intended to evaluate each IRF's efficiency relative to that of the national median IRF. Specifically, the measure assesses Medicare spending by the IRF and other healthcare providers during an MSPB episode. The measure reports the ratio of the payment-standardized, risk-adjusted MSPB-PAC Amount for each IRF divided by the episode-weighted median MSPB-PAC Amount across all IRFs. The MSPB-PAC Amount is the ratio of the observed episode spending to the expected episode spending, multiplied by the national average episode spending for all IRFs. The measure is calculated using two consecutive years of Medicare Fee-for-Service (FFS) claims data and was developed using calendar year (CY) 2015-2016 data. This submission is based on fiscal year (FY) 2016-2017 data; i.e., IRF admissions from October 1, 2015 through September 30, 2017.

Claims-based MSPB-PAC measures were developed in parallel for the IRF, long-term care hospital (LTCH), skilled nursing facility (SNF), and home health agency (HHA) settings to meet the mandate of the IMPACT Act. To align with the goals of standardized assessment across all settings in PAC, these measures were conceptualized uniformly across the four settings in terms of the construction logic, the approach to risk adjustment, and measure calculation. Clinically meaningful case-mix considerations were evaluated at the level of each setting. For example, clinicians with IRF experience evaluated IRF claims and then gave direction on how to adjust for specific patient and case-mix characteristics.

The MSPB-PAC IRF measure was adopted by the Centers for Medicare & Medicaid Services (CMS) for the IRF Quality Reporting Program (QRP) and finalized in the FY 2017 IRF Prospective Payment System (PPS) Final Rule.[1] Public reporting for the measure began in Fall 2018 through the IRF Compare website (https://www.medicare.gov/inpatientrehabilitationfacilitycompare/) using FY 2016-2017 data.

Notes:

[1] Medicare Program; Inpatient Rehabilitation Facility Prospective Payment System for Federal Fiscal Year 2017 Federal Register, Vol. 81, No. 151. https://www.gpo.gov/fdsys/pkg/FR-2016-08-05/pdf/2016-18196.pdfThe Medicare Spending Per Beneficiary – Post Acute Care Measure for Inpatient Rehabilitation Facility (MSPB-PAC IRF) was developed to address the resource use domain of the Improving Medicare Post-Acute Care Transformation Act of 2014 (IMPACT Act). This resource use measure is intended to evaluate each IRF's efficiency relative to that of the national median IRF. Specifically, the measure assesses Medicare spending by the IRF and other healthcare providers during an MSPB episode. The measure reports the ratio of the payment-standardized, risk-adjusted MSPB-PAC Amount for each IRF divided by the episode-weighted median MSPB-PAC Amount across all IRFs. The MSPB-PAC Amount is the ratio of the observed episode spending to the expected episode spending, multiplied by the national average episode spending for all IRFs. The measure is calculated using two consecutive years of Medicare Fee-for-Service (FFS) claims data and was developed using calendar year (CY) 2015-2016 data. This submission is based on fiscal year (FY) 2016-2017 data; i.e., IRF admissions from October 1, 2015 through September 30, 2017.

Claims-based MSPB-PAC measures were developed in parallel for the IRF, long-term care hospital (LTCH), skilled nursing facility (SNF), and home health agency (HHA) settings to meet the mandate of the IMPACT Act. To align with the goals of standardized assessment across all settings in PAC, these measures were conceptualized uniformly across the four settings in terms of the construction logic, the approach to risk adjustment, and measure calculation. Clinically meaningful case-mix considerations were evaluated at the level of each setting. For example, clinicians with IRF experience evaluated IRF claims and then gave direction on how to adjust for specific patient and case-mix characteristics.

The MSPB-PAC IRF measure was adopted by the Centers for Medicare & Medicaid Services (CMS) for the IRF Quality Reporting Program (QRP) and finalized in the FY 2017 IRF Prospective Payment System (PPS) Final Rule.[1] Public reporting for the measure began in Fall 2018 through the IRF Compare website (https://www.medicare.gov/inpatientrehabilitationfacilitycompare/) using FY 2016-2017 data.

Notes:

[1] Medicare Program; Inpatient Rehabilitation Facility Prospective Payment System for Federal Fiscal Year 2017 Federal Register, Vol. 81, No. 151. https://www.gpo.gov/fdsys/pkg/FR-2016-08-05/pdf/2016-18196.pdf

IM.1.1. Developer Rationale: MSPB-PAC IRF was developed to address the resource use domain of the Improving Medicare Post-Acute Care Transformation Act of 2014 (IMPACT Act). As part of the IMPACT Act, MSPB-PAC aims to achieve interoperability, data exchange, and standardized measurement among post-acute providers. The mandated use of MSPB-PAC measures is intended to allow for a greater ability to measure resource use and efficiency of care to improve outcomes, as well as encourage all PAC providers towards aligned incentives and care coordination.

Differences in post-acute care payments are a key driver of variation in Medicare spending overall.[1,2] There have been a number of studies demonstrating significant variability in IRF care and outcomes, links between facility characteristics and readmissions, and significant opportunities for improvement.[3,4,5] The cost and quality link is important, with this resource use measure playing an important role in discerning value of IRF care.

The MSPB-PAC IRF measure was adopted by CMS for the IRF Quality Reporting Program (QRP) and finalized in the FY 2017 IRF Prospective Payment System (PPS) Final Rule.[6] Public reporting for the measure began in Fall 2018 through the IRF Compare website.

[1] Institute of Medicine. (2013). Variation in Health Care Spending Assessing Geographic Variation. (July)

[2] Kahn, E. N., Ellimoottil, C., Dupree, J. M., Park, P., & Ryan, A. M. (2018). Variation in payments for spine surgery episodes of care: Implications for episode-based bundled payment. Journal of Neurosurgery: Spine, 29(2), 214–219.

[3] Middleton, A., Graham, J. E., Prvu Bettger, J., Haas, A., & Ottenbacher, K. J. (2018). Facility and Geographic Variation in Rates of Successful Community Discharge After Inpatient Rehabilitation Among Medicare Fee-for-Service Beneficiaries. JAMA Network Open, 1(7), e184332. https://doi.org/10.1001/jamanetworkopen.2018.4332

[4] Middleton A, Graham JE, Deutsch A, O. K. (2017). Potentially Preventable Within-Stay Readmissions Among Medicare Fee-for-Service Beneficiaries Receiving Inpatient Rehabilitation. PM R, 9(11), 1095–1105.

[5] Daras, L. C., Ingber, M. J., Deutsch, A., Hefele, J. G., & Perloff, J. (2018). Geographic Region and Profit Status Drive Variation in Hospital Readmission Outcomes Among Inpatient Rehabilitation Facilities in the United States. Archives of Physical Medicine and Rehabilitation, 99(6), 1060–1066.

[6] Medicare Program; Inpatient Rehabilitation Facility Prospective Payment System for Federal Fiscal Year 2017 Federal Register, Vol. 81, No. 151. https://www.gpo.gov/fdsys/pkg/FR-2016-08-05/pdf/2016-18196.pdf

De.1. Measure Type: Cost/Resource Use

S.5. Data Source: Assessment Data

Claims

Enrollment Data

Other

S.3. Level of Analysis: Facility

New Measure Submission

Preliminary Analysis: New Measure

Criteria 1: Importance to Measure and Report

1a. <u>High impact or high resource use</u>:
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 focus of this measure is to assess Medicare spending by inpatient rehabilitation facilities (IRFs) and other healthcare providers during an MSPB-PAC IRF episode, which includes all Medicare Part A and Part B services with a start date in the episode window, except for a limited set of services that are not clinically related to the episode.
- This measure was developed as part of the Improving Medicare Post-Acute Care Transformation Act of 2014 (IMPACT Act).
- The measure is a claims-based measure and is specified at the facility level and is calculated as the ratio of the payment-standardized, risk-adjusted MSPB-PAC Amount for each IRF divided by the episode-weighted median MSPB-PAC Amount across all IRFs.
- The developer cites data that show that the differences in post-acute care payments are a key driver of variation in Medicare spending overall. They cite a number of studies demonstrating significant variability in IRF care and outcomes, links between facility characteristics and readmissions, and significant opportunities for improvement.
- According to the March 2020 Medicare Payment Advisory Commission (MedPAC) report, Medicare spent $8 billion on IRF care provided to fee-for-service (FFS) beneficiaries in about 1,170 IRFs Nationwide in 2018. About 364,000 beneficiaries had 408,000 IRF stays. On average, the Medicare FFS program accounted for about 59 percent of IRF discharges.

1b. Opportunity for Improvement.

- The developer provided publicly reported measure score data for all US providers under Medicare's IRF Prospective Payment System (PPS) with 20 or more eligible episodes in the reporting period of FY 2016-2017. The data from 1,147 IRFs with 20 or more episodes in the reporting period of FY 2016-2017, which include measure scores from 618,123 patient episodes shows the mean of 1.00 with a standard deviation of 0.08. The developer also reported the interquarile range of 0.10 (min: 0.74 and max: 1.47).
- No disparities data provided.

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)?

Comments:

- **Yes**
- **The measure is part of a Congressionally mandated set of measures on post-acute care. Legislative requirement probably creates adequate impact**
- **Yes.**
- **there is substantial variation in post-acute care spending. Unclear what role disparities/social risk factors play in this.**
- **Yes**
- **Yes.**
- **Yes**
- **Yes (required by IMPACT Act)**
- **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?

Comments:

- **Yes**
- **About a 20% difference in 10th to 90th percentile. Reasonable to attack, not the largest differential.**
- **Data from FY 2016-2017 was used. The measure score varied between 0.74 and 1.47, with 30% IRFs having lower than national average, and 46% IRFs having higher than national average, respectively. Thus, the proposed measure can potentially flag very high-cost or very low-cost IRFs relative to the national average, and thus help minimize cost variations through different incentive programs.**
- **yes, there is variation on this measure in terms of spending**
- **Yes**

- **The developer pointed to overall variation in cost in post-acute care. I agree this is an area for improvement. However, the cost measures presented are for each of the different post-acute care settings. A more effective approach may be to allow providers to identify lower cost post-acute care settings to send a patient rather than trying to control costs within a specific setting. For example, the highest costs identified in the Medicare Comprehensive Joint Replacement (CJR) Model are for inpatient rehabilitation. To reduce spending, hospitals may explore outpatient rehabilitation. This approach to reducing in costs could not be captured in the current post-acute care cost measures. Further, identifying variation in cost does not necessarily mean that there is room for reducing costs. The developer does not identify appropriate care or the impact of high and low cost care on patient heath outcomes.**
- **Yes**
- **The data from 1,147 IRFs with 20 or more episodes in the reporting period of FY 2016-2017, which include measure scores from 618,123 patient episodes shows the mean of 1.00 with a standard deviation of 0.08. The developer also reported the interquarile range of 0.10 (min: 0.74 and max: 1.47).**
- **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: Bijan Borah, MSc, PhD, Jack Needleman, PhD, Jennifer Perloff, PhD, Zhenqiu Lin, PhD, Jeffrey Geppert, EdM, JD, Eugene Nuccio, PhD, Christie Teigland, PhD, Susan White, PhD, RHIA, CHDA, Ronald Walters, MD, MBA, MHA, MS **(Evaluation A: Methods Panel)**

Methods Panel Individual Reliability Ratings: H-3; M-4; L-0; I-0

Methods Panel Individual Validity Ratings: H-1; M-6; L-1; I-0

- The developer provided responses to the concerns raised by the SMP, which can be found in the SMP Spring 2020 Discussion Guide on page 84 – 85.

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.

2a2. Reliability Testing:

- The developers conducted reliability testing at the measure score level using signal-to-noise analysis to examine the measure score's ability to capture between-facility differences versus random error. They also assessed measure score reliability using split-sample method with intraclass correlation coefficients (ICCs) to examine agreement between two performance measure scores for a facility based on randomly-split, independent subsets of Inpatient Rehabilitation Facilities (IRF) episodes.
- The performance measure score reliability testing was based on 1,147 Inpatient Rehabilitation Facilities (IRFs) with 20 or more episodes in the measurement period of FY 2016-2017.
- The developers used signal-to-noise analysis using Adams' method and reported mean reliability score of 0.86 and median of 0.89. The results demonstrated that on average, 86 percent of the variation in the risk adjusted MSPB amount was associated with systematic differences between facilities, with a range of 70 to 96 percent (on average) among the smallest and largest facility quartiles, respectively.
- The results of reliability testing at the measure score level using split-sample method with correlations (ICCs) show mean score of 0.87 with 95% confidence interval of 0.85 to 0.88, with a range of 0.81 to 0.95 (on average) among the smallest and largest facility quartiles, respectively

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?*

Guidance from reliability algorithm:

(Box 1) Are specifications precise, unambiguous, and complete? YES → (Box 2) Was empirical testing conducted using statistical tests with the measure as specified? YES → (Box 4) Was reliability testing at the score level? YES → (Box 5) Was the method appropriate? YES → (Box 6) Moderate certainty of measure reliability → MODERATE

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:

Comments:

- **Why 30-day post-discharge window? How is death handled within episode definition?**
- **Exclusions of unrelated services is briefly described but Appendix D, with extensive description not provided to committee. Appendix D leaves some details unclear. Specifically, I am concerned with the statement:Limit to Services Representing Significant Cost Once integrated into clinically meaningful service categories, services that did not account for a sufficiently large share of payments within their respective clinical service category were not included in the review to allow clinicians to focus their review on services representing a higher percentage of overall Medicare spending within the episode window. Services representing insignificant cost are therefore automatically included in the MSPB-PAC measures and counted toward the attributed PAC provider's episode. This statement does not provide adequate description of "sufficiently large share." Given the measure is applied to practices with 20 patients, a large impact on one or a few cases could change the ranking of a smaller volume facility while the expenses as a percentage of total spending are small.**
- **I don't have any questions/concerns beyond what the Scientific Methods Panel has raised regarding the definitional transparency of this measure (e.g., attribution of services to specific episode in case of multiple episode during measurement period, winsorization of the outliers).**
- **It is odd to use the MDS to determine whether patient has been institutionalized. this is a complicated measure to construct. Issue of services being assigned to multiple epsisodes--not sure why this would be the case. Also, need to windsorize values before running models. Concur that extreme values should not be excluded by truncated/windorized.**
- **No concerns**
- **No concerns that were not raised by the Scientific Methods Panel.**
- **No concerns**
- **none**
- **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.

Comments:

- **Yes**
- **Data come from claims, not fully audited but the source has been accepted in the past.**
- **None.**
- **yes. could raise the minimum denominator of 20 a bit higher to ensure adequate reliablity of 0.7 for nearly all IRFs (would be helpful to see what a cut off of 50 or 75 episodes looks like).**
- **How was 20 episodes selected as the minimum reporting size?**
- **No concerns that were not raised by the Scientific Methods Panel.**
- **Yes**
- **none**
- **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 is attributed to the IRFs. This attribution approach was developed in order to encourage IRFs to facilitate care coordination, improve referral practices, and support their role in managing cost and resource use.
- Costing approach:
    - The costing approach is based on payments by Medicare for services within the identified resource use service categories. Payments are based on agreed upon fee schedules for each setting.

2b2. Validity Testing:

- The developer conducted three separate empirical tests on validity:
    - 1) Assessed how this measure correlates to resource /utilization such as hospitalization within the episode window and emergency room (ER) visits within the episode window.
    - 2) Correlated this measure with the Discharge to Community (DTC) rates for IRFs (measure endorsed by NQF (#3479).
    - 3) Correlated this measure with the Percent of Residents or Patients with Pressure Ulcers That Are New or Worsened (Short-Stay) measures (NQF #0678).
- The developers found a positive relationship between MSPB PAC-IRF and known indicators of resource or service utilization.
    - The mean observed to expected cost ratio for episodes without a hospital admission is 0.91, compared with 1.39 for episodes with at least one hospital admission during the episode period (p-value<0.0001).
    - The mean observed to expected cost ratio for episodes without an ER visit is 0.98, compared to 1.09 for episodes with at least one ER visits (p-value<0.0001).
    - These positive relationships between the mean observed to expected cost ratio and the number of hospitalizations/ER visits was as hypothesized.
    - The developer also found a small, significant negative association between MSPB PAC-IRF measure scores and the DTC measure scores as hypothesized and a small but statistically significant correlation (Pearson -0.193; Spearman -0.165) between MSPB PAC-IRF measure scores and DTC measure scores.
    - Lastly, the developer found a small, statistically significant correlation (both Pearson and Spearman) between MSPB PAC-IRF measure scores and Pressure Ulcers measure scores (Pearson 0.1207; Spearman 0.1198).

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

- No concerns

2b4/2c. Risk adjustment

- The developer uses a linear regression risk adjustment model with 146 risk factors.
- The developer reported results showing that spending was higher for patients who were dual eligible ($36,204 vs $33,489 for non-duals), non-white (e.g., $36,805 for Blacks compared to $33,469 for whites), and those beneficiaries with low SES based on AHRQ SES Index ($34,367 in lowest quantile vs $33,870 in highest quantile).
- The developer reported that each of the social risk factors was statistically significant in the risk adjustment model. However, the developer did not include them in the overall model, concluding that adding them, individually or together, did not substantially improve overall model fit. The developer reported that including social risk factors in the risk-adjustment model had minimal impact on measure scores.

2b5: Meaningful Differences

- Across 1,147 facilities, the mean score was 1.00 with a standard deviation of 0.08 and a range of 0.74 to 1.47. The 10th percentile had a score of 0.91 and the 90th percentile 1.10.

Questions for the Committee regarding validity:

- *Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?*
- *Do you agree with the developer's rationale for not included SES factors in the risk-adjustment model?*

Guidance from reliability algorithm:

(Box 1) Were threats to validity addressed? YES → (Box 2) Was empirical testing conducted using statistical tests with the measure as specified? YES → (Box 5) Was validity testing at the score level? YES → (Box 6) Was the method appropriate? YES → (Box 7) Moderate certainty of measure validity → MODERATE

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 and NQF-convened Clinical Technical Expert 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?

Comments:

- **None**
- **Basic testing is Signal to noise reliability Split sample reliability. Differences are tested across 4 quartiles of size, with most critical the 20-75 episodes quartile. Signal to noise and split sample scores high on average but low in 1st quartile (20-190 episodes) SN reliability: .64 in bottom 25% of smallest quartile, 0.73 at Median, split sample 0.81) SN may be too low for approval for 1st quartile, notwithstanding split sample ICC. I am concerned the measure is not reliable for the lowest volume providers and the thresholds were not chosen to assure reliability across the full set of providers.**
- **No concerns.**
- **While the developer looked at inclusion of social risk factors and reported that differences were not meaningful, this can be generally true for most providers but not the case for providers with large fractions of patients with social risk factors. I did not see information on the effects on providers with large #s of patients with social risk factors as I suspect there are some IRFs that predominantly care for patients with substantial social risk factors. low R2 suggests the developer hasn't accounted for the broader set of factors driving variation in spending.**
- **No concerns.**
- **No concerns not raised by the Scientific Methods Panel.**
- **No concerns**
- **none**
- **No concerns**

2b5a. Threats to Validity: Meaningful Differences: Describe any concerns with the analyses demonstrating meaningful differences among accountable units:

Comments:

- **Not clear how death is handled if poor care leads to increased death care post-discharge.**
- **Variation exists. 10th to 90th approximately $6K for average cost of $30K**
- **None.**
- **differences are notable**
- **No concerns.**

- **The developer says they used an OLS model. It is unclear how they accounted for multiple observations from each IRF where observations would not be independent. I am concerned with the inclusion of patients in hospice and SNF residents. The comorbid conditions, care, and cost profiles of these patients would be very different from those discharged to the community after an IRF stay. I am also concerned that there are important social factors missing from the model. Specifically, the home environment of the patient, their access to transportation to enable them to use other settings for rehabilitation, whether they have a caregiver who can aid them, as well as whether they live alone. These elements could be collected in the IRF patient assessment instrument which is already administered at admission to an IRF.**
- **No concerns**
- **none**
- **No concerns**

2b5b. 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 opulation still be valid?

Comments:

- **This measure excludes costs for "Clinically Unrelated Services." The process of aggregating costs to be considered for exclusion and the expert panel process are relatively clear. In the process, services that did not account for a sufficiently large share of payments within their respective clinical service category were not included in the review to allow clinicians to focus their review on services representing a higher percentage of overall Medicare spending within the episode window. The basis for determining which costs were not examined under this criterion is not specified. I am concerned that by looking at aggregates, costs that substantially and idiosyncratically affect individual patients may not be considered, and low volume providers will be penalized by random events unrelated to their care. I would like more description of this process than is provided in Appendix D. THIS IS A COMMON ISSUE ACROSS THE MEASURES BEING CONSIDERED IN THIS CYCLE.**
- **None.**
- **the lack of use of the Medicare administrative data indicating whether a beneficiary is dually eligible, low income supplement, and disabled for starters per social risk factors.**
- **No concerns.**
- **n/a**
- **No concerns**
- **none**
- **No concerns**

2b2. Additional threats to validity: attribution, the costing approach, or truncation: 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?

Comments:

- **None**
- **No concerns about attribution. Standard issues associated with CMS standardized prices. Standard winsorizing of outliers.**
- **Episodes with residuals below 0.1 percentile or above 0.99 percentile of the residual distribution are excluded. Scientific Methods Panel members #3 and #9's observations on handling of outliers by this measure need further discussion. Specifically, should the measure necessarily exclude outliers at all?**
- **outlier issue--shouldn't exlude, should cap/windsorize**
- **Not a concern since the SMP reviewed the measures, but two questions since I am less familiar with these methodologies. In winsorization of predicted costs, why was the decision made to bottom-code only and not top-code? Why were outliers defined based on the residuals?**
- **It is unclear whether the accountable entity is providing appropriate care although potentially tied to higher costs.**
- **No concerns**
- **Because the measure is episode based, there is potential bias against larger IRFs with more episodes than smaller IRFs with fewer episodes; SMP noted that the prediction model over-estimates observed costs of care for smaller IRFs and under-estimates the observed costs of larger IRFs resulting in smaller IRFs looking "better than predicted" and reverse for larger IRFs.**
- **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:

Comments:

- **See comments on data exclusions above**
- **None.**
- **Concern that the validation approach (correlating measures to the MSPB measure) focused on correlation with 2 variables that are part of the IRF episode that the measure is measuring. This is not an independent check of the measure's validity. Of course one would expect spending to be higher if someone was hospitalized or used the ER. This part of the validity check was sub-par.**

- **No concerns.**
- **There is no comparison to patients that may have been treated in an IRF but were treated in a different post-acute care setting.**
- **No concerns**
- **none**
- **No concerns**

2b4. 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?

Comments:

- **As suggested by reviewer, risk-adjustment models should be based on theory, not empirics. Dropping variables will bias the coefficient estimates of the remaining variables.**
- **Risk adjustment approach is standard approach for CMS measures. For a specialized service like IRF, this may not be an adequate approach. The history variables include HCC ranked comorbidities, but individuals come into IRF for specific conditions that should influence the course and cost of treatment, separate from overall health or significant morbidities. If the condition that determines the HCC is not the condition that brings an individual into IRF care, then this aspect is not included in the risk adjustment model, and may distort estimates of risk of downstream services and costs. I am not clear whether the initial cause of treatment in the period immediately before IRF is taken into account separately, and would like this clarified. The committee also needs to discuss whether the general approach of CMS to risk adjustment on these variables, to use their standard method with minimal tweaks to reflect the risk for costs associated with the reason the patient is in a specific PAC setting, should be accepted. Social risk factors analyzed, and as implemented in risk adjustment model, make small-negligible difference in score or adjusted R-square. We have no data on how much variability across facilities there is in social risk factors, and whether facilities with high proportion of disadvantaged SES patients regularly score worse, and whether the differences is scores are larger for these facilities.**
- **While the measure developer offers rationale for not including SES in the final measure, I agree with the Scientific Panel Member #2 as to why the measure should actually include SES. Exclusion of SES may potentially penalize IRFs that serve higher proportion of patients with these SES factors.**

- **Issue of not adjusting for social risk factors is problematic. The coefficients were significant on these variables and other clinical risk factors are soaking up a lot of the effects that go hand-in-hand with social risk factors. Unclear why the measure contractor/Acumen did not use beneficiary variables (rather than the AHRQ SES index) to adjust directly (e.g., dual status, disability). What is most important to examine is what are the effects of IRFs that disproportionately treat large numbers of high social risk factor patients--that is where risk adjustment will matter most, not the mean effect. Low R2 suggests other factors not accounted for that are driving epsiode spending. Validity not fully established--would rate that as low. Outlier removal is not warranted---better to winsorize rather than remove these cases.**
- **I was surprised that despite the conceptual and data suggested relationship between the SES Index that the decision was made to omit these from the risk adjustment model. I was interested in how the determination was made that they didn't have a substantive contribution to fit? Also, while the SES factors didn't affect scores on average, were there instances where the inclusion of SES did affect score materially?**
- **Although the developer tests social risk factors and there are statistically significant results, they exclude them from the model. There are also new measures of social risk being added to the IRF-PAI that could be meaningful in the model including: mental status, confusion, depression, nutrition, language, health literacy, transportation, and social isolation.**
- **No concerns**
- **Yes. Developer reported results showing spending higher for patients who were dual eligible ($36,204 vs $33,489 for non-duals), non-white (e.g., $36,805 for Blacks compared to $33,469 for whites), and those beneficiaries with low SES based on AHRQ SES Index ($34,367 in lowest quantile vs $33,870 in highest quantile). The SES factors seem to align, yet developer chose not to include them in the overall model."The developer reported that each of the social risk factors was statistically significant in the risk adjustment model. However, the developer did not include them in the overall model, concluding that adding them, individually or together, did not substantially improve overall model fit." Seems contradictory.**
- **No concerns**

Combined Scientific Methods Panel Preliminary Analysis of Scientific Acceptability

Measure Number: 3561

Measure Title: Medicare Spending Per Beneficiary - Post Acute Care Measure for Inpatient Rehabilitation Facilities

Type of measure:

☐ **Process** ☐ **Process: Appropriate Use** ☐ **Structure** ☐ **Efficiency** ☒ **Cost/Resource Use**
☐ **Outcome** ☐ **Outcome: PRO-PM** ☐ **Outcome: Intermediate Clinical Outcome** ☐ **Composite**

Data Source:

☒ **Claims** ☐ **Electronic Health Data** ☐ **Electronic Health Records** ☐ **Management Data**
☐☒ **Assessment Data** ☐ **Paper Medical Records** ☐ **Instrument-Based Data** ☐ **Registry Data**
☐☒ **Enrollment Data** ☒☐ **Other: Minimum Data Set (Panel Member #2) (MDS) (Panel Member #8)**

Level of Analysis:

☐ **Clinician: Group/Practice** ☐ **Clinician: Individual** ☒ **Facility** ☐ **Health Plan**

☐ **Population: Community, County or City** ☐ **Population: Regional and State**
☐ **Integrated Delivery System** ☐ **Other**

Measure is:

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

RELIABILITY: SPECIFICATIONS

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

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

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

   Panel Member #3 Description: "This resource use measure is intended to evaluate each IRF's efficiency relative to that of the national median IRF. Specifically, the measure assesses Medicare spending by the IRF and other healthcare providers during an MSPB episode. The measure reports the ratio of the payment-standardized, risk-adjusted MSPB-PAC Amount for each IRF divided by the episode-weighted median MSPB-PAC Amount across all IRFs. The MSPB-PAC Amount is the ratio of the observed episode spending to the expected episode spending, multiplied by the national average episode spending for all IRFs. The measure is calculated using two consecutive years of Medicare Fee-for-Service (FFS) claims data…"

2. **Briefly summarize any concerns about the measure specifications.** None

   Panel Member #2 This is a very complex measure to calculate requiring many steps, some of which seem a bit ambiguous, such as which events apply as exclusions, where episodes end, which events be included in 2 different overlapping IRF episodes, etc. While CMS may be able to utilize the developing set of codes and logic, I do not think this measure is easily replicable by others who may want to calculate and compare IRF spending. It likely has applicability only to the CMS IRF compare program for public reporting.

   Panel Member #3 Terms: 1) payment-standardized, risk-adjusted MSPB-PAC Amount for each IRF; 2) episode-weighted median MSPB-PAC Amount across all IRFs, and 3) MSPB-PAC Amount is the ratio of the observed episode spending to the expected episode spending, multiplied by the national average episode spending for all IRFs, will need to be specified operationally. There is already a lack of clarity due to redundancy of terminology. MSPB-PAC Amount is described as a "ratio of observed to expected times national average", but the term MSPB-PAC Amount is initially modified with the terminology of "payment-standardized, risk adjusted" and "episode-weighted median". There is a lack of transparency in what this measure is and how the measure is computed due to these redundancies.

   Measure score: If the MBSP-PAC Amount is "ratio of observed to expected times national average" and the measure is "the ratio of the payment-standardized, risk-adjusted MSPB-PAC Amount for each IRF divided by the episode-weighted median MSPB-PAC Amount across all IRFs", the the measure is a ratio of ratios. "An MSPB-PAC IRF measure score of less than 1 indicates that a given IRF's resource use is less than that of the national median IRF during a performance period. An MSPB-PAC IRF measure score of greater than 1 indicates that a given IRF's resource use is more than that of the national median IRF during a performance period."… "For example, a measure score of 1.1 indicates that the IRF had average risk-adjusted spending levels that are 10 percent higher than the median IRF. On the other hand, an MSPB-PAC IRF measure score of less than 1 indicates that an IRF had lower average risk-adjusted spending levels compared to those of the median IRF. For example, a measure score of 0.9 indicates that the IRF had average risk-adjusted spending levels that are 10 percent lower than the median IRF."

The measure already seems to be in use by CMS and is publicly reported: The MSPB-PAC IRF measure was adopted by the Centers for Medicare & Medicaid Services (CMS) for the IRF Quality Reporting Program (QRP) and finalized in the FY 2017 IRF Prospective Payment System (PPS) Final Rule.[1] Public reporting for the measure began in Fall 2018 through the IRF Compare website.

Appropriate use of data(?): The MDS is used to construct a risk adjustment variable, indicating beneficiaries who have been institutionalized for at least 90 days in a given year. Why use MDS data rather than claims data to identify if a patient has been institutionalized? Is this the only risk factor?

Time period for performance: The episode window is opened by a trigger event (i.e., admission to the IRF) and ends 30 days after the discharge from that IRF. The measure uses episodes across a two-year period. What if the patient goes back to the hospital after discharge from the IRF? How are all of those costs captured?

Standardizing the cost of care: The calculation of cost addresses differences due to geography or specifi factors by "eliminate variation in payments due to Medicare geographic adjustment factors and add-on payments for Medicare programs" as the first step. Is this done by identifying hours and clinicial specialty and then applying a standardized rate of pay for that specialty across all calculations? NOTE1: low predicted cost values are "winsorized" to the predicted value at the 0.5th percentile value if the predicted value is lower. NOTE2: No adjustment appears to be made for outrageously high predicted values. A statement later in the document indicates that episodes with residuals above the 99th percentile in value are excluded.

Incorporation of clinical differences among patients served by different IRFs: The calculation offered for the numerator (and by extension of the denominator which is the median national value), does not appear to address differences in clinical characteristics of the patients served. ""MSPB-PAC IRF Amount"_k=((1/n_k)*Sum_i_{i_k}[Y_ik/(Y_ik)^])*((1/N)*Sum_k[Sum_i_{i_k}[Y_ik]])

Where:

Y_ik is the attributed standardized spending for episode i and provider k.

(Y_ik)^ is the expected standardized spending for episode i and provider k, as predicted from risk adjustment, and resulting from Step 6 above

n_k is the number of episodes for provider k

N is the number of episodes nationally

i_{i_k} is all episodes i in the set of episodes attributed to provider k"

Are patient clinical differences only addressed by the risk adjustment model that creates the "expected standardized spending for episode"? NOTE: There is apparently a categorization based on the clinical characteristiics of the patient prior to patient's episode into one of seven categories ranging from "prior acute surgical" through "prior acute medical" to "Community". This categorization is based on a 90-day "look back" period.

PLEASE NOTE: GIVEN THAT THE BACKGROUND INFORMATION ON HOW MEASURES 3561, 3562, 3563, AND 3564 IS IDENTICAL, MY CONCERNS AND CITATIONS OF EVIDENCE IS IDENTICAL FOR EACH OF THESE **FOUR MEASURES. I WILL NOT REPEAT THE DETAILS PRESENTED HERE FOR MEASURES 3562 – 3564, BUT SIMPLE MAKE THE STATEMENT "SEE RELATED COMMENTS FROM MY REVIEW OF MEASURE 3561".**

Panel Member #6 I was curious about Winzorization after calculating predicted episode level costs – seems like the predicton model will be stronger if outliers are removed first.

Panel Member #8 General concern about the potential overlap in the various MSPB metrics submitted. It appears that each services may be attributed to multiple 'episodes' and providers. Unclear how improvements may be made when the attribution is so scattered.

Panel Member #9 My key concern is with the outlier exclusion (Step 6 – Contruction logic). Based on the measure specifications, episodes with residuals below the 1st percentile or above the 99th percentile of the residual distribution are excluded. Three factors make this approach particularly concerning: 1) This is a measure focusing on resource utilization, should episodes with very low or very high residuals be excluded? If the concern is with potentially undue influences of outliers, is exclusion the best available approach? 2) Winsorize predicted values: low predicted value below 0.5th percentile is already winsorized before calculation of residual. 3) Closing episodes: full payment for all claims that begin within the epidose window is counted toward the episode. This may make it likely that such episodes (with substantial claims at the end of episode window) become outleirs and be excluded. At minimum, the developer should report the distribution of outlier exclusion across facilities to ensure that they don't concentrate in a limited number of facilities.

RELIABILITY: TESTING

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

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

   ☐ **Yes** ☐☒ **No** **X—Not applicable (Panel Member #3)**
6. **Assess the method(s) used for reliability testing**

   Submission document: Testing attachment, section 2a2.2

   Panel Member #1 Signal-to-noise (reliability score) and split-sample reliability testing (Intraclass correlation or ICC) were conducted both for the overall sample and for the sample size quartiles.

   Panel Member #2 The developers used 2 different measures of reliability: 1) Reliability score (signal to noise) to evaluate the extent to which variation in the masure is due to true, underlying differences in provider performance (signal)rather than random variation (noise). 2) split-sample reliability testing to examine agreement between 2 scores for a facility based on randomly-spllit, independent subsets of IRF episodes. Good agreement indicates the performance score is more the result of facility characteristics (efficient care) than statistical noise due to random variation. They used 4 years of data to achieve #'s of episodes per facility comparable to the numbers used for actual measurement (at least 20 episodes per year) with episodes across years evenly distributed. They used the Shrout-Fleixx interclass correlation coefficient (ICC) between the split-half scores to measure reliability.

   Panel Member #3

   - Used a signal-to-noise approach comparing between facility vs. random error method
   - Used four years of data and split-half comparison by year
   - Calculated Shrout-Fleiss intraclass correlation coefficients ICC(2,1)[1] between the split-half scores to measure reliability… For these reasons, we view the Reliability Score as the preferred and more

relevant metric of reliability. We still present both reliability metrics for completeness. [Note: definition of "Reliability Score" not specified in text]

Panel Member #6 Appropriate methods

Panel Member #7 Split sample reliability testing was utilized with four years of data first split- into quartiles and both reliability scores and intraclass coefficients were calculated to measure reliability.

Panel Member #8 ICC using split half.

Panel Member #9 The developer used two approaches to calculate measure score reliability. One is the signal-to-noise reliability with reference to Adams' NEJM paper, another is the split-sample reliability based on Shrout-Fleiss' intraclass correlation coefficient. However, Adams obtained between variance from a two-level hierarchical linear model while this measure is not based on a linear hierarchical lear model, it is not completely clear how different variance components were obtainted to calculate the reliability scores.

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

Submission document: Testing attachment, section 2a2.3

Panel Member #1 Based on both the mean (SD) reliability score of 0.86 (0.12) and ICC of 0.87(95% CI: 0.85-0.88), the reliability of performance measure score can be termed as good.

Panel Member #2 Overall, reliability testing indicated good reliability, regardless of facility size. The average reliability score for all agencies was 0.86 with median of 0.89. When examined by facility size, the average reliability score ranged from 0.70 (Q1) to 0.96 (Q4). The ICC for the overall sample was 0.87 with 95% confidence interval of 0.85-0.88. The ICC was lowest for Q1 (0.81) and highest in Q4 (0.95)

Panel Member #3 Results as presented show strong overall reliability, but clear relationship to the size of the IRF based on the number of episodes.

Table 3. Facility Reliability Score Distribution of the Episode-Level MSPB Risk Adjusted Spending, overall IRF sample and by sample size quartile, with public reporting exclusions (k = 1,147)

Facility Sample	K	Mean (SD)	25th Pct*	Median	75th Pct
Overall	1,147	0.86 (0.12)	0.81	0.89	0.94
Quartile 1: 20-190 episodes	288	0.70 (0.12)	0.64	0.73	0.78
Quartile 2: 191-334 episodes	286	0.85 (0.04)	0.83	0.85	0.87
Quartile 3: 335-686 episodes	287	0.91 (0.02)	0.90	0.91	0.93
Quartile 4: 687-5,941 episodes	286	0.96 (0.01)	0.96	0.97	0.97

[1] Shrout, Patrick E., and Joseph L. Fleiss. "Intraclass correlations: uses in assessing rater reliability." *Psychological bulletin* 86, no. 2 (1979): 420.

Table 4. Split-sample reliability: Intraclass correlation coefficients between split-sample performance measure scores for the overall IRF sample and by sample size quartile, with public reporting exclusions (N = 1,147)

Facility Sample	K	ICC(2,1) (95% CI)
Overall	1,147	0.87 (0.85-0.88)
Quartile 1: 20-190 episodes	287	0.81 (0.76-0.84)

Quartile 2: 191-334 episodes	287	0.87 (0.84-0.90)
Quartile 3: 335-686 episodes	286	0.93 (0.91-0.94)
Quartile 4: 687-5,992 episodes	287	0.95 (0.94-0.96)

Panel Member #6 Strong reliability; minimum episode size of 20 per facility may be too small. The 25th percentile for the lowest volume quartile of facilities is .64. With 25% of cases having lower scores, there are facilities with relatively low reliablity.

Panel Member #7 The facility reliability score mean score was 0.89 overall , range 0.81 to 0.94 and varied from 0.73 for the first quartile, range of 0.64 to 0.78, to 0.97 for the fourth quartile, range of 0.96 to 0.97. The ICC split-sample reliability was 0.87 overall with a range of 0.81 for the first quartile to 0.95 for the fourth quartile. Thus, though there was variation in the reliability of the scores between quartiles, all were moderate to good.

Panel Member #8 Good performance – concern re: small number of episodes (20-190)

Panel Member #9 Results are very good based on two different approaches.

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 <u>all</u> testing results):

    ☐☒ **High** (NOTE: Can be HIGH <u>only if</u> score-level testing has been conducted)

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

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

    ☐ **Insufficient** (NOTE: Should rate <u>INSUFFICIENT</u> 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 Please see my rationale for this rating in 6 and 7 above.

    Panel Member #2 Based on reliability testing using two different approaches, both showing good to high reliability regardless of facility size.

    Panel Member #3 Variation across IRF sizes gave me some concerns—especially if the measure is being used to financially reward IRF facility performance.

    Panel Member #6 Strong scores on reliability measures.

Panel Member #7 86% of the variation in the risk-adjusted MSPB was associated with systemic differences between the facilities by the reliability score.

Panel Member #8 Developer may consider raising the minimum episode threshold to address the concern of lower reliability in the 20-190 range -

Panel Member #9 Although further clarification on the signal-to-noise approach will be helpful, the results based on split-sample reliability testing are very good. Since Shrout's ICC estimate tends to be low, this is very reassuring.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

Submission document: Testing attachment, section 2b2.

Panel Member #1 None

Panel Member #2 NONE

Panel Member #3 Exclusions: These seem reasonable (e.g., routine procedures such as dialysis, colonoscopy; and non-standard episodes (e.g., missing claims data, death)). "…clinically unrelated services built off the planned readmissions algorithm developed by the Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation."

Exclusion distribution tables presented in 2b2.2 and 2b2.3 were very interesting and should be used to monitor future changes in distribution to see if IRF's can improve their data submission quality.

Panel Member #6 No major concerns; exclusions are primarily driven by two criteria, each of which is needed to calculate the measure.

Panel Member #7 None. Appropirate exclusion criteria are given and were provided by expert opinion.

13. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.** None

Submission document: Testing attachment, section 2b4.

Panel Member #2 NONE

Panel Member #3

From MIF document:

MSPB-PAC IRF measure scores are reported publicly for providers with 20 or more eligible episodes, along with the national average score. The distribution of MSPB-PAC IRF measure scores (based on FY 2016-2017 data) that are statistically significantly different from the national average is as follows:

- Significantly lower than the national average: 30%
- Not statistically different from the national average: 24%
- Significantly higher than the national average: 46%

Inference about both measure performance of individual providers can be made based on the score value. The distribution of MSPB-PAC IRF measure score values based on FY 2016-2017 data is as follows:

- Minimum: 0.74
- 10th Percentile: 0.91
- 25th Percentile: 0.95
- 75th Percentile: 1.05

- 90th Percentile: 1.10
- Maximum: 1.47

MSPB-PAC IRF measure scores are publicly reported on IRF Compare for IRFs with 20 or more eligible episodes. Out of 1,161 IRFs with FY 2016-2017 episodes, only 14 did not meet this minimum threshold.

Panel Member #6 The authors show ability to identify facilities that are different from the national average. I'm not sure that's really tells us much about any type of clinically 'clinically meaningful' difference, but it may meet the goals of the measure sponsore (CMS).

Panel Member #7 Across 1,147 facilities across the country, The mean score was 1.00 with a standard deviation of 0.08 and a range of 0.74 to 1.47. The 10th percentile had a score of 0.91 and the 90th percentile 1.10.

Panel Member #9 No concern.

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
Submission document: Testing attachment, section 2b5.
Panel Member #1 N/A
Panel Member #2 NONE
Panel Member #6 NA
Panel Member #9 No concern.

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

Submission document: Testing attachment, section 2b6.

Panel Member #1 None

Panel Member #2 NONE

Panel Member #6 NA

Panel Member #7 For this claims-based measure, only 0.002% of episodes were excluded from the measure due to claims issues.

Panel Member #9 No concern.

16. **Risk Adjustment**

16a. **Risk-adjustment method** ☐ **None** ☐☒ **Statistical model** ☐ **Stratification**

16b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☒☐ Yes ☐ No ☐☒ Not applicable

16c. **Social risk adjustment:**

16c.1 Are social risk factors included in risk model? ☒☐ Yes ☐☒ No ☐ Not applicable

16c.2 Conceptual rationale for social risk factors included? ☒ Yes ☐ No

16c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☒ Yes ☐ No

16d.**Risk adjustment summary:**

16d.1 All of the risk-adjustment variables present at the start of care? ☐☒ Yes ☒☐ No – hospice during stay **(Panel Member #6)**

16d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐ Yes ☒☐ No X—Not applicable **(Panel Member #3)**

16d.3 Is the risk adjustment approach appropriately developed and assessed? ☒ Yes ☐ No

16d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration)

☐☒ Yes ☐ No

Panel Member #3

- Adjusted R-squared value = 0.1595
- Hosmer-Lemeshow 10-decile analysis was very strong

16d.5.Appropriate risk-adjustment strategy included in the measure? ☒ Yes ☐ No Yes applies to clinical risk factors, NO SES factors were included. **(Panel Member #2)**

16e**. Assess the risk-adjustment approach**

Panel Member #1 I am convinced with the empirical justification as to why social risk factors were not worth including in this measure.

Panel Member #2 In spite of making a strong conceptual argument for including SES, and statistical results showing spending was higher for patients who were dual eligible ($36,204 vs $33,489 for non duals, or 8% higher which is a very large difference in the Medicare payment world, non-white (e.g., $36,805 for Blacks compared to $33,469 for whites, or 10% higher which can have very significant impact on a plan with many non-white members, and had low SES based on AHRQ SES index ($34,367 in lowest quantile vs $33,870 in highest quantile), and that each of those social risk factors was statistically significant in the risk adjustment model, they concluded that adding them, individually or together, did not substantially improve overall model fit. This is NOT surprising given there were already **146 variables** in the model, and patients with above SES characteristics tend to have more clinical risk factors as well; however the SES remained significant in the model. Further, they found including SES had minimal impact on average measure scores, so they chose to NOT adjust the scores for these social risk factors. I believe the small impact on overall scores was likely due to several factors, one that only 19% of all episodes had patients who were dual eligible and 16% involved patients who were non-white. Also the AHRQ SES index is based on wide geographic area and is not likely to fully capture the impact of SES on individuals who may live in an area with both poor and wealthy people thus the characteristics wash out when averaged. More granular data on patient's SES characteristics is needed. Still, given the empirical findings of 8-10% higher costs, which is significant for Medicare plans, I strongly believe the factors should have been kept in the model. In cases where an IRF serves a large proportion of patients with these SES factors, they will be penalized for having expected higher costs, which is not intent of the measure and could restrict access to high quality IRFs for these patients.

Panel Member #6 Relaitvely low R-squared (15%); do not find the argument that excluded costs like dialsys drive now model fit – if you have a near perfect predictor (ESRD status), why not include these costs? The use of prior utilization also drives up the R-squared – those with a past history of service use are likely to have more service use in the future regardless of clinical need.

Panel Member #7 The regression coefficients, standard errors, and p-values for each of the covariates are given with an adjusted R-squared of 0.1595. For deciles of spending, the observed/predicted costs ranged between 0.99 and 1.01.

Panel Member #9 Risk-adjustment approach is acceptable.

<u>For cost/resource use measures ONLY</u>:

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

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

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

Panel Member #1 None

Panel Member #3

- **See previously stated concern about smaller IRFs with fewer episodes**
- **Note—the prediction model over-estimates the observed cost of care for smaller IRFs and under-estimates the observed cost of care for the larger IRFs. If the prediction model were absolutely correct, this would mean that the smaller IRFs are doing "better than predicted", while the reverse is true for the larger IRFs!**

Panel Member #6 Attribution is very clean with this measure b/c it starts with an institutional stay; I'm suprized the measures is not more skewed, but that is a result of the truncation and renormalization. No major concerns on the design of the measure.

Panel Member #8 General concern about the potential overlap in the various MSPB metrics submitted. It appears that each services may be attributed to multiple 'episodes' and providers. Unclear how improvements may be made when the attribution is so scattered.

Panel Member #9 See my comments on outlier exclusion earlier.

VALIDITY: TESTING

19. **Validity testing level: ☒ Measure score ☐ Data element ☐ Both**
20. **Method of establishing validity of the measure score:**

    ☐ **Face validity**

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

    ☐ **N/A (score-level testing not conducted)**
21. **Assess the method(s) for establishing validity**

    Submission document: Testing attachment, section 2b2.2

    Panel Member #1 The directionality of the all the hypothesized correlations (see 21 above) were established and were statistically significant, implying that that the MSPB measure is sensitive to both the occurrence and the intensity of high cost events.

    Panel Member #2 The developers used the following methods to test reliability:1. Evaluated correlation with known indicators of resource or service utilization (hospital admissions and emergency room (ER) visits during the episode period). They compared the ratio of observed over expected spending for MSPB-PAC IRF episodes with and without hospital admissions occurring in the episode period. They also compared the observed over expected spending for episodes with and without ER visits. This analysis tested whether variation in service utilization is captured by the MSPB-PAC cost measure. 2. Examined the correlation between MSPB-PAC IRF scores and the Discharge to Community (DTC) rates for IRFs (measure endorsed by NQF (#3479) is publicly reported and also based on Medicare claims. 3.Examined the correlation between MSPB-PAC IRF scores and provider's scores on the Percent of Residents or Patients with Pressure Ulcers That Are New or Worsened (Short-Stay) measures (NQF #0678). Pressure ulcers are high-cost adverse events, and their prevalence is an important measure of quality of care. The measure reports the percent of Medicare patients with Stage 2-4 pressure ulcer(s) that are new or worsened since admission, is assessment-based, and is publicly reported on the IRF Compare website.

    Panel Member #3

    - Correlated MSPB measure score with three other related events:
        - Hospitalization (should increase MSPB values)
        - Discharge to community after IRF (should decrease MSPB values)

- Existence of pressure ulcers (should increase MSPB values)

- All relationships were demonstrated

Panel Member #6 I believe ED and inpatient vist costs within 30 days of discharge from the IRF are included in the measure. If this is true, comparing the costs of episodes with and without these high cost events is useful descriptive information, but not really validation of the measures. I like the use of IRF compare quality measures – even if these are claims based, they likey use different pieces of information (diagnosis, time relationship of events). It is not clear to me why worsening pressure ulcers during an episode would not be associated with cost of care – seems like a potentially cost increasing event.

Panel Member #7 Validity was assessed by the relationship of MSPB to other known indicators of resource/utilization, specifically number of hospitalizations and ER visits, discharge to community measures, and pressure ulcer measures.

Panel Member #8 Correlation with other measures

Panel Member #9 The developer conducted three separate empirical tests on validity: 1) Assess how this measure is related to resource /utilization such as hospitalization within the episode window and ER visit within the episode window. 2) Correlate this measure with the discharge to community rates measure. 3) Correlated this measure with the Pressure Ulcers measure. These are not unreasonable, but both hospitalization and. ER visit are functionally related to the measure. And DTC is also inherently related to this measure.

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

Submission document: Testing attachment, section 2b2.3

Panel Member #1 The directionality of the all the hypothesized correlations (see 21 above) were established and were statistically significant, implying that that the MSPB measure is sensitive to both the occurrence and the intensity of high cost events.

Panel Member #2 The developers found a positive relationship between MSPB and known indicators of resource or service utilization. The mean observed to expected cost ratio for episodes without a hospital admission is 0.91, compared with 1.39 for episodes with at least one hospital admission during the episode period (p-value<0.0001). The mean observed to expected cost ratio for episodes without an ER visit is 0.98, compared to 1.09 for episodes with at least one ER visits (p-value<0.0001). They also observed a positive relationship between the mean observed to expected cost ratio and the number of hospitalizations/ER visits as hypothesized. They also found a small, significant negative association between MSPB measure scores and the DTC measure scores as hypothesized and a very small but statistically significant correlation (Pearson -0.193; Spearman -0.1165) between MSPB measure scores and DTC measure scores. Lastly, they found very small, statistically significant correlation (both Pearson and Spearman) between MSPB measure scores and Pressure Ulcers measure scores (Pearson 0.1207; Spearman 0.1198) which may indicate pressure ulcrs are associated with somewhat higher spending in 30 days after discharge from IRF.

The positive relationship between MSPB and other indicators of resource/service utilization confirms that the MSPB measure is sensitive to both the occurrence and the intensity of high cost events. The small, significant negative correlation between MSPB and DTC measures confirms that, on average, more efficient IRFs are associated with better discharge to community rates and fewer unplanned hospitalizations.

Panel Member #3

- Relationships were all directionally appropriate

- However, this raises concerns that the MSPB measure may be more reflective of the 30-day period post discharge than the costs associated with IRF care during confinement to the IRF.

Panel Member #6 The .12 correlation between worsening pressure ulcers and IRF resource use is the most convincing evidence of validity the authors present (for me). I'm less clear on the .19 correlation between IRF resource use and community tenure – I guess healthier cases cost less and having longer community tenure.

Panel Member #7 The first was expected and found to have the highest correlative relationship, with a cost ratio from 0.91 to 1.81 for hospitalizations, and 0.98 to 1.17 for ER visits

Panel Member #8 Relationship with all but # of hospitalizations is relatively weak.

Panel Member #9 Given the nature of their relationship, the results are as expected.

23. **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)

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

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

    ☐ **Yes**

    ☐ **No**

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

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

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

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

    ☒☐ **Low** (NOTE: Should rate LOW if you believe that there <u>are</u> threats to validity and/or relevant threats to validity were <u>not assessed OR</u> 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 <u>is required</u>; 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 Please see my explanations in 21 and 22 above.

    Panel Member #2 Validity results were based on several approaches and all showed strong validity based on hypothesized relationships.

    Panel Member #3 See comment for item #22

    Panel Member #6 I think the evidence on validity is realitvely thin (correltions for two quality measures). Over time it would be helpful to examine more evidene, such IRF cost reports, to look at the relationship between reimbursements and cost of producing the care.

Panel Member #7 Medicare spending per beneficiary did positively correlate with expected cost drivers such as hospitalization and EC visits.

Panel Member #8 Low correlation with ER visits concerning. Correlations with hospitalizations, DTC and pressure ulcer rate statistically significant, but question the clinical relevance (need clinician advice on this).

Panel Member #9 Outlier exclusion is a key concern. And empirical testing results are to be expected given the functional relationship between the tested components.

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**

    Panel Member #6 Evidence is pointing in the right direction, but it is thin.

ADDITIONAL RECOMMENDATIONS

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

    Panel Member #9 Outlier exclusion should be discussed further. Based on the measure specifications, episodes with residuals below the 1st percentile or above the 99th percentile of the residual distribution are excluded. Three factors make this approach particularly concerning: 1) This is a measure focusing on resource utilization, should episodes with very low or very high residuals be excluded? If the concern is with potentially undue influences of outliers, is exclusion the best available approach? 2) Winsorize predicted values: low predicted value below 0.5th percentile is already winsorized before calculation of residual. 3) Closing episodes: full payment for all claims that begin within the epidose window is counted toward the episode. This may make it likely that such episodes (with substantial claims at the end of episode window) become outleirs and be excluded. At minimum, the developer should report the distribution of outlier exclusion across facilities to ensure that they don't concentrate in a limited number of facilities.

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.

- The developer indicated that all data elements are coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims) and are defined in fieleds in a combination of electronic sources

- The developer noted that the data used to calculate this measure are already collected as part of Medicare's payment process. Thus, the measure does not pose any additional data collection burden on providers

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):

Comments:

- **Approach appears feasible**
- **Measure is routinely compiled from claims and related data, so it is feasible.**
- **No concerns with feasibility.**
- **this measure is feasible. agree it is complicated to construct, but the data are available.**
- **No concerns with generating the measure, but it will be challenging for facilities to obtain timely feedback and track their performance outside of the reporting cycle.**
- **This measure is feasible to collect and calculate.**
- **No concerns**
- **Think the availability of social risk data that would improve risk adjustment not yet routinely captured**
- **Data routinely generated**

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

- The developer noted that this measure is publicly reported as part of the Center of Medicare & Medicaid Services' Inpatient Rehabilitation Facilities Quality Reporting Program (IRF QPS). The data are publicly reported on IRF Compare website, which in addition to tracking quality of care, it is are intended to help consumers make informed decisions when selecting healthcare providers.
- The IRF QRP includes all IRFs paid under the IRF PPS. MSPB-PAC IRF scores are publicly reported for active providers with 20 or more eligible episodes in the reporting period.

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

- The developer solicited feedback on this measure via a 24-day public comment period furing January-February 2016. Additionally, feedback was also sought through the pre-rulemaking process, and public comments were received after the release of the Measures Under Consideration (MUC) List on December 1, 2015.
- The developer noted that the feedback was received from a wide range of stakeholders and they covered a range of topics, including episode construction, exclusions, score calculation, risk adjustment, and reporting.
- The developers addressed the comments, either by revising the measure or by providing the rationale why revisions are not necessary or appropriate. They reviewed and considered all public comments during development and implementation, before finalizing the measure in the FY 2017 IRF PPS final rule

Additional Feedback:

- The developer also sought feedback on the measure through the pre-rulemaking process. Four public comments were received after the release of the Measures Under Consideration (MUC) List on December 1, 2015. The MAP PAC/LTC Workgroup met on December 14-15 to consider this measure, and provided the preliminary decision of "encourage continued development" for the MSPB-PAC IRF measure. Following the release of the MAP PAC Workgroup's preliminary recommendation, the report was open for a public comment period. Eight public comments on this measure were received in this time. The MAP Coordinating Committee considered these comments alongside the Workgroup recommendation and finalized the recommendation of "encourage continued development," releasing their final recommendations in February 2016. Members of the public could comment during both MAP meetings
- MAP noted that socioeconomic status is a particular concern for IRFs. Patients need social supports to be able to return to the community with a disability. Additionally, it was suggested that providers who

are at teaching facilities and serve low income patients may have higher costs than others. Members noted the importance of balancing cost measures with quality and access. Although the MAP encouraged continued development, they did note concerns about the potential for unintended consequences. In particular, the group raised concerns about issues of premature discharges and ability to make comparisons across providers. The group noted this could put a tremendous burden on family caregivers who may have to care for a patient they are not fully able to support. Members noted the need to consider risk adjustment for severity and socioeconomic status and urged CMS to incorporate functional status assessments into risk adjustment models to promote improvements. MAP requested consideration in the finalization of specifications to ensure costs are not double-counted between care settings; and recommended submission to NQF for endorsement. The MAP noted the measure was never fully specified before the PAC/LTC workgroup deliberations and the current specifications were released in mid-January with public comment period closing Jan 27th. It was noted that the measures double count costs between providers and is inconsistent with IMPACT act to develop comparable resource measures of PAC providers. While the MAP's final decision was to recommend continued development, there was a level of discomfort in this decision expressed by a number of Members.

***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 did not provide any improvement results. NQF does not require this for initial endorsement.

4b2. Unintended consequences

- The developer did not note any unexpected consequences

4b2.Potential harms

- The developer did not highlight any potential harms

4b3. Transparency

- This measure is publicly reported.
- Data and result detail are maintained such that the resource use measure, including the construction logic for a defined unit of measurement, can be deconstructed to facilitate transparency and understanding

***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?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

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?

Comments:

- **Unclear**
- **Current use is public reporting for accountability, not yet used for payment.**
- **No additional comments beyond what NQF has already stated.**
- **CMS posts the measure publicly**
- **No comment.**
- **This measure has been publicly reported since Fall 2018.**
- **No concerns**
- **Data is publicly available, though unclear whether the measure is easily replicable by others who may want to calculate and compare IRF spending. It likely has applicability only to the CMS IRF compare program for public reporting.**
- **Planned use**

4a2. Use – Feedback: 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?

Comments:

- **yes.**
- **None.**
- **Complicated measure to unpack (for the IRFs) and to then figure out how to make changes that would influence the measure. Inadequate validity testing based on not testing against an external standard.**
- **No concerns.**
- **Although the measure has been publicly reported since Fall 2018, since this is the measure's initial endorsement, no improvement data was provided. It is unclear whether public reporting has provided IRFs with sufficient information to reduce costs.**
- **No concerns**
- **Seems so**

- **Planned use**

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?

Comments:

- **I am not clear on what is being reported as part of the measure back to providers, including whether individual cost components and relative performance by cost component is provided.**
- **Yes, potential for improvement was demonstrated by incentivizing high-cost IRFs to the national mean.**
- **moderage to low. not super actionable for IRFs absent a better understanding of the factors that drive variations in spending**
- **Yes.**
- **Although the measure has been publicly reported since Fall 2018, since this is the measure's initial endorsement, no improvement data was provided. It is unclear whether public reporting has impacted IRF costs.**
- **No concerns**
- **Seems to assume it can be used for improvement**
- **Planned use**

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

Comments:

- **None**
- **None.**
- **measures is still relatively new to identify harms. biggest potential harm is for those IRFs with high presence of individuals with social risk factors. This measure could create negative incentives for IRFs to take these types of patients.**
- **The measure could be more beneficial is the developers explored appropriate resource use and its relationship to health outcomes.**
- **No concerns**

- **It seems that by not including SES factors in risk adjustment there is a real concern that for IRFs treating a large proportion of patients with SES factors will risk being penalized for having expected higher costs, which goes against the intent of the measure and could limit access to IRFs for these patients.**

None

Criterion 5: Related and Competing Measures

- There are no competing measures for #3561 (i.e. same measure focus and target population)
- The developer identified the following NQF endorsed measures as related measures:
  - 2158 : Medicare Spending Per Beneficiary (MSPB) - Hospital

Harmonization

- The developer stated that the MSPB-PAC measures are harmonized across post-acute care (PAC) settings as well as with MSPB-Hospital. MSPB-PAC measures were developed in parallel for all PAC settings to meet the mandate of the IMPACT Act. To align with the goals of standardized assessment across PAC settings, these measures were conceptualized uniformly across the four settings in terms of the construction logic, the approach to risk adjustment, and measure calculation. The measures mirror the general construction of MSPB-Hospital. The developer noted that aligning the MSPB-Hospital and MSPB-PAC measures in this way creates continuous accountability and aligns incentives to improve care planning and coordination across inpatient and PAC settings.

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?

Comments:

- **This is part of a package of PAC measures, and some patients will have multiple PAC services, each generating an episode. Not clear how the different experiences should be understood at the patient level or how provider attribution is to be interpreted.**
- **No competing measure identified. Related NQF-endorsed measure identified by the developer is: 2158: Medicare Spending Per Beneficiary (MSPB) – Hospital. The developer stated that the MSPB-PAC measures are harmonized across post-acute care (PAC) settings (Inpatient Rehabilitation Facility, Long Term Care Hospital, Skilled Nursing Facility and Home Health Agency) as well as with MSPB-Hospital.**
- **No concerns**
- **Agree with SMP commenter that there is potential overlap in the various MSPB metrics submitted. It appears that each service may be attributed to multiple 'episodes' and providers. Unclear how improvements may be made when the attribution is so scattered. Developer believes the PAC MSPB measures are in alignment with hospital MSBP measure.**

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: July 1, 2020

- Comment by Federation of American Hospitals
  Member Vote N/A
  8390 :
  The Federation of American Hospitals (FAH) appreciates the opportunity to comment on this measure prior to the Standing Committee's evaluation. The FAH requests that the committee carefully consider whether the measure as specified produces performance scores that are reliable and valid for facility-level reporting.

  Specifically, the FAH is concerned to see that reliability at the 25th percentile for 20-190 episodes was 0.64, which leads us to question what result was produced at the minimum level. We believe that the results currently provided indicate that the measure as specified may not produce scores that yield acceptable minimum thresholds for reliability.

  The scientific acceptability of the measure is further called into question on review of the risk model's fit with the overall adjusted R-squared as 0.1595. While the developer provides some explanation on why the result is low, the FAH does not believe that the reasons for this result are adequately addressed and risk adjustment must be improved prior to endorsement.

  In addition, while the FAH appreciates that social risk factors were reviewed, we believe that the risk adjustment approach should not consider the identification and testing of social risk factors as supplementary to clinical risk factors. This approach was identified as a concern by the NQF Disparities Standing Committee and developers must begin to include these factors within the testing of the model rather than the approach of "adding on" factors after the model is developed. This type of analysis would assist facilities and others in understanding how their inclusion could impact the model and provide additional information for groups examining this issue such as the NQF and Office of the Assistant Secretary for Planning and Evaluation. As a result, the FAH believes that this measure lacks sufficient information on the potential impact these social risk variables have on the risk adjustment model.

  Furthermore, while the developer provides information on the changes in performance scores that result from social risk factor adjustment (tables 2b3.4b_4a and 2b3.4b_4b), it is not necessarily clear on the degree to which these changes would result in a facility's score being statistically significantly different from the national average. If the interpretation of the results under meaningful differences leads to a "conclusion that even small difference between facility scores can be treated as meaningful"(response to 2b4.3 in the testing form), to what extent would changes in performance as a result of adjustment for social risk factors also lead to different but meaningful 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)

MSPB-PAC IRF was developed to address the resource use domain of the Improving Medicare Post-Acute Care Transformation Act of 2014 (IMPACT Act). As part of the IMPACT Act, MSPB-PAC aims to achieve interoperability, data exchange, and standardized measurement among post-acute providers. The mandated use of MSPB-PAC measures is intended to allow for a greater ability to measure resource use and efficiency of care to improve outcomes, as well as encourage all PAC providers towards aligned incentives and care coordination.

Differences in post-acute care payments are a key driver of variation in Medicare spending overall.[1,2] There have been a number of studies demonstrating significant variability in IRF care and outcomes, links between facility characteristics and readmissions, and significant opportunities for improvement.[3,4,5] The cost and quality link is important, with this resource use measure playing an important role in discerning value of IRF care.

The MSPB-PAC IRF measure was adopted by CMS for the IRF Quality Reporting Program (QRP) and finalized in the FY 2017 IRF Prospective Payment System (PPS) Final Rule.[6] Public reporting for the measure began in Fall 2018 through the IRF Compare website.

[1] Institute of Medicine. (2013). Variation in Health Care Spending Assessing Geographic Variation. (July)

[2] Kahn, E. N., Ellimoottil, C., Dupree, J. M., Park, P., & Ryan, A. M. (2018). Variation in payments for spine surgery episodes of care: Implications for episode-based bundled payment. Journal of Neurosurgery: Spine, 29(2), 214–219.

[3] Middleton, A., Graham, J. E., Prvu Bettger, J., Haas, A., & Ottenbacher, K. J. (2018). Facility and Geographic Variation in Rates of Successful Community Discharge After Inpatient Rehabilitation Among Medicare Fee-for-Service Beneficiaries. JAMA Network Open, 1(7), e184332. https://doi.org/10.1001/jamanetworkopen.2018.4332

[4] Middleton A, Graham JE, Deutsch A, O. K. (2017). Potentially Preventable Within-Stay Readmissions Among Medicare Fee-for-Service Beneficiaries Receiving Inpatient Rehabilitation. PM R, 9(11), 1095–1105.

[5] Daras, L. C., Ingber, M. J., Deutsch, A., Hefele, J. G., & Perloff, J. (2018). Geographic Region and Profit Status Drive Variation in Hospital Readmission Outcomes Among Inpatient Rehabilitation Facilities in the United States. Archives of Physical Medicine and Rehabilitation, 99(6), 1060–1066.

[6] Medicare Program; Inpatient Rehabilitation Facility Prospective Payment System for Federal Fiscal Year 2017 Federal Register, Vol. 81, No. 151. https://www.gpo.gov/fdsys/pkg/FR-2016-08-05/pdf/2016-18196.pdf

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.**

MSPB-PAC IRF measure scores are reported publicly for all US providers paid under Medicare's IRF Prospective Payment System (PPS) with 20 or more eligible episodes in the reporting period. There were a total of 1,147 IRFs with 20 or more episodes in FY 2016-2017. Their scores represent 618,123 patient episodes, after all exclusions were applied. The scores show a good deal of variability – the descriptive statistics are provided below.

MSPB-PAC IRF score descriptive statistics:

Mean: 1.00

Standard Deviation: 0.08

Min: 0.74

Max: 1.47

Interquartile range: 0.10

Score Percentiles

10: 0.91

20: 0.94

30: 0.96

40: 0.98

50: 1.00

60: 1.02

70: 1.04

80: 1.06

90: 1.10

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.

Not applicable

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.**

Not applicable

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.

Not applicable

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.

MSPB-PAC IRF is intended to allow for a greater ability to measure resource use and efficiency of care to improve outcomes, as well as encourage all PAC providers towards aligned incentives and care coordination. The measure assesses Medicare spending by IRFs and other healthcare providers during an MSPB-PAC IRF episode. An MSPB-PAC IRF episode includes all Medicare Part A and Part B services with a start date in the episode window, except for a limited set of services that are not clinically related to the episode. The episode window is opened by a trigger event (i.e., admission to the IRF) and ends 30 days after the discharge from that IRF. The measure is calculated as the ratio of the payment-standardized, risk-adjusted MSPB-PAC Amount for each IRF divided by the episode-weighted median MSPB-PAC Amount across all IRFs. The MSPB-PAC Amount is the ratio of the observed episode spending to the expected episode spending, multiplied by the national average episode spending for all IRFs.

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):*

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

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

Post-Acute Care

S.1. Measure-specific Web Page *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/Downloads/2016_07_20_mspb_pac_ltch_irf_snf_measure_specs.pdf

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.

Assessment Data

Claims

Enrollment Data

Other

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.)*

This measure is based on Medicare FFS administrative claims and uses data from the Medicare enrollment database and Minimum Data Set (MDS). The enrollment database provides information such as date of birth, date of death, sex, reasons for Medicare eligibility, periods of Part A and Part B coverage, and periods in the Medicare FFS program. The MDS is used to construct a risk adjustment variable, indicating beneficiaries who have been institutionalized for at least 90 days in a given year. The data elements from the Medicare FFS claims are those basic to the operation of the Medicare payment systems and include data such as date of admission, date of discharge, diagnoses, procedures, and revenue center codes. The Medicare FFS claims data files are used to identify Medicare services from IRFs and other settings (e.g., the outpatient setting) within the episode window. No data beyond the claims submitted in the normal course of business are required from providers for the calculation of this measure.

This measure submission is based on FY 2016-2017 data, which were the most recent data available at the time of our analyses. We used the data sources listed below to develop the analytic file for measure specification and testing:

- Medicare Fee-For-Services claims and enrollment data: We accessed inpatient, outpatient, carrier, skilled nursing facility, home health, durable medical equipment, and hospice claims through the Centers for Medicare & Medicaid Services (CMS) Common Working File (CWF). The data dictionary for all Medicare FFS claims, demographic, and enrollment data are available at: https://www.resdac.org/cms-data?tid%5B%5D=4931&tid_1%5B%5D=1&=Find+Data+Files. General information about the CWF is available at: https://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/Downloads/clm104c27.pdf.
- Minimum Data Set (MDS): Acumen obtains the MDS through the Quality Improvement and Evaluation System (QIES). The data dictionary for the MDS data is available at: https://www.resdac.org/cms-data/files/mds-3.0/data-documentation.

We used two mappings to group diagnosis and procedure codes for use in identifying clinical events, implementing exclusions and applying risk adjustment:

- Agency for Healthcare Research and Quality (AHRQ) Clinical Classifications Software (CCS) groupings for Services and Procedures: Software is available for download at: https://www.hcup-us.ahrq.gov/toolssoftware/ccs_svcsproc/ccssvcproc.jsp
- CMS-Hierarchical Condition Category (HCC) mappings of ICD-9 and ICD-10 codes: We used the Version 22 CMS-HCC mapping, which is included in the software available at: https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Risk-Adjustors.html.

We used five additional data sources for measure testing purposes only and not for measure specification:

- 2017 American Community Survey (ACS) 5-year estimate: We used the ACS to obtain the ZIP Code Tabulation Area (ZCTA) level measures needed to compute the Agency for Healthcare Research and Quality (AHRQ) Socioeconomic Status (SES) index score for use in social risk factor testing. This information is downloadable at the US Census website: http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t.
- Rural-Urban Continuum Codes 2013: We used this data source to construct rural-urban identifiers for social risk factor testing. These codes include county FIPS indicators, which are then merged onto our episode file. More information on this data source can be found at: https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/.
- Provider of Services Current Files (POS File): We used this data source to describe the characteristics of IRFs included in measure specification and testing, such as census region, ownership type, and rurality, as reported in Table 1. The POS file contains data on characteristics of hospitals and other types of healthcare facilities, including the name and address of the facility and the type of Medicare services the facility provides, among other information. The data are collected through the CMS Regional Offices. General information about

the POS Files is available at: https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/index.html.

- IRF Compare data: We used this data source to examine the relationship between MSPB and assessment-based quality measures. The IRF Compare data include publicly reported IRF quality measures. The data are available at https://data.medicare.gov/data/inpatient-rehabilitation-facility-compare

- Common Medicare Environment (CME) database: We extracted patient-level dual eligibility information from the CME database for social risk factor testing. CMS has designated the CME database as the single, enterprise-wide authoritative source for Medicare beneficiary enrollment and demographic data. The CME database integrates and standardizes different types of beneficiary data from CMS legacy systems. The CME database receives information from the EDB and also contains additional information not available in the EDB. A description of the CME is available at: https://www.ccwdata.org/documents/10280/19002256/medicare-enrollment-impact-of-conversion-from-edb-to-cme.pdf.

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)*

<SamplingMethodologySpecificDataSourceAttachment nodeType="0" />

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: See section S.5.1

Please supply the username and password:

Attachment:

Code Table:

URL:

Please supply the username and password:

Attachment: S_6_Code_Table.xlsx

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.

The MSPB-PAC IRF measure assesses Medicare spending by IRFs and other healthcare providers during an MSPB-PAC IRF episode. An MSPB-PAC IRF episode includes all Medicare Part A and Part B services with a start date in the episode window, except for a limited set of services that are not clinically related to the episode. The episode window is opened by a trigger event (i.e., admission to the IRF) and ends 30 days after the discharge from that IRF. The measure is calculated as the ratio of the payment-standardized, risk-adjusted MSPB-PAC Amount for each IRF divided by the episode-weighted median MSPB-PAC Amount across all IRFs. The MSPB-PAC Amount is the ratio of the observed episode spending to the expected episode spending, multiplied by the national average episode spending for all IRFs.

An MSPB-PAC IRF measure score of less than 1 indicates that a given IRF's resource use is less than that of the national median IRF during a performance period. An MSPB-PAC IRF measure score of greater than 1 indicates that a given IRF's resource use is more than that of the national median IRF during a performance period.

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

See supplemental documentation (S_7_2_Construction_Logic) for a version of text provided below with standard formulas and additional graphics.

Episode Construction

MSPB-PAC IRF episodes assess all Medicare Part A and Part B claims for services delivered to a beneficiary during the episode window, subject to exclusions for particular services that are clinically unrelated to PAC treatment. Constructing an MSPB-PAC IRF episode involves the following steps:

(1) Defining the episode trigger, episode window, treatment period, and associated services period;

(2) Excluding certain services from the episode that are clinically unrelated to PAC treatment and closing the episode.

Episode Trigger. Each episode is opened by an episode trigger. The trigger for IRF episodes is admission to the IRF, except for readmissions occurring within 7 days to the same provider. The IRF that triggers the episode is the provider to whom the episode is attributed for the purpose of calculating the MSPB-PAC IRF measure (attributed provider). We identify an admission to an IRF based on the inpatient claims with the third character of CMS Certification Number (CCN) equal to "T" or "R," or with last four digits of CCN that fall within the range of 3025-3099.[1]

Episode Window. The episode window consists of a treatment period and an associated services period.

Treatment Period. The treatment period of an MSPB-PAC IRF episode begins on the day of the trigger and ends at discharge. Readmissions of the same patient to the same provider within 7 or fewer days after discharge do not trigger a new episode and instead are included in the treatment period of the original episode to reflect the likelihood that these closely adjacent stays are related. For gaps of 7 or fewer days, stays in the same setting with the same patient and provider are collapsed into one treatment period. For instance, when two sequential stays at the same IRF occur within 7 or fewer days of one another, the treatment period ends on the day of discharge for the latest IRF stay. The treatment period includes Medicare Part A and Part B services delivered to a beneficiary that are provided directly or could reasonably have been managed by the attributed IRF provider, and that are related to the beneficiary's care plan. Treatment services occurring on the first day of MSPB-PAC IRF episodes are subject to exclusions related to prior institutional care, including ambulance transport to the attributed IRF facility and durable medical equipment, prosthetics, orthotics, and supplies (DMEPOS) orders preceding the patient's admission to the IRF. Treatment services are also subject to exclusions for particular services that are clinically unrelated to PAC treatment, as described in section S.9.1, below.

Associated Services Period. The associated services period starts at the trigger event for each of the MSPB-PAC IRF episodes, and ends 30 days after the end of the treatment period. The associated services period is the time during which all non-treatment services are counted towards the episode (associated services). Such services are associated with IRF care but are not provided directly, or could not reasonably have been managed by the attributed provider. For instance, an associated service includes an acute inpatient hospital admission for a complication arising during or after IRF treatment. The Medicare spending for all Part A and Part B services during the associated services period are counted toward the episode, with exceptions for clinically unrelated services, as described in section S.9.1, below.

Closing Episodes. MSPB-PAC IRF episodes end 30 days after discharge from the facility. The full payment for all claims that begin within the episode window is counted toward the episode; this is done to maintain consistency with the MSPB-Hospital measure (NQF #2158) and to fairly assign payment to the episode for Medicare claims paid on a prospective payment system, regardless of episode length.

An MSPB-PAC IRF episode may begin during the associated services period of another MSPB-PAC IRF episode in the 30 days post-treatment. See section S.7.3 for examples of situations in which this scenario occurs and how it is handled in the MSPB-PAC IRF measure.

Measure Calculation

Certain episodes are excluded from the MSPB-PAC IRF calculation to ensure that the measure facilitates meaningful comparisons between IRF providers. These exclusions are distinct from the exclusions for clinically unrelated services discussed above, which exclude a limited set of services from MSPB-PAC IRF episodes. In contrast, episode-level exclusions, discussed in section S.9.1, remove entire episodes from measure calculation when certain criteria are met.

After applying the episode-level exclusions, the measure can be calculated in the following steps:

Step 1: Standardize Claim Payments.

The first step in calculating the standardized payment for a claim is to eliminate variation in payments due to Medicare geographic adjustment factors and add-on payments for Medicare programs, such as indirect medical education (IME) and disproportionate share hospitals (DSH). The goal of this step is to remove sources of variation not directly related to decisions to provide clinical services. Payment standardization controls for geographic variation in healthcare payments, such as the hospital wage index and geographic practice cost index (GPCI).[2] All payment data shown in the MSPB-PAC IRF measure and supporting documentation reflect allowed amounts, which include both Medicare trust fund payments and beneficiary deductible and coinsurance. Bonus or penalty amounts due to Medicare quality reporting or other special programs are not included.

Step 2: Calculate Standardized Episode Payments.

Next, to prepare claims data for calculating risk-adjusted payments, standardized episode payments are calculated. For each episode, standardized payments include all standardized Medicare claims payments for services in the episode window, as detailed in previous paragraphs.

Step 3: Calculate Predicted Episode Payments.

The third step calculates predicted payments for each episode. This step estimates the relationship between the independent variables and standardized episode payments using an ordinary least squares (OLS) regression. See Appendix C of the Measure Specifications document provided in section S.1 for a full list of the independent variables used in the risk adjustment model.[3]

Step 4: Winsorize (Bottom Code) Predicted Values.

Next, the distribution of predicted values is examined. If the distribution of predicted values includes extremely low values (defined as below the 0.5th percentile), winsorization is performed at the low end of the distribution (i.e., “bottom coding”). The resultant values are renormalized to maintain a consistent average episode payment. If the distribution of predicted values does not include extremely low values, winsorization is not required to ensure meaningful ratios of observed to predicted spending (see below). In accordance with the MSPB-Hospital measure (NQF #2158) calculation, renormalization multiplies the winsorized predicted values by the ratio of the average original predicted payment and the average winsorized predicted payment. For example, suppose an episode’s predicted value (PREDICTED_VALUE) is $1,000, but the 0.5th percentile of predicted values is $1,500. Then, that episode’s “winsorized” predicted value (WINS_PREDICTED_VALUE) would be $1,500. The “renormalized” winsorized predicted value would be:

($1500×mean(PREDICTED_VALUE))/mean(WINS_PREDICTED_VALUE)

where the mean is taken over the entire national sample of the MSPB-PAC IRF episodes. This re-normalization ensures that the average of the resulting winsorized predicted values is equal to the average of the original predicted values.

Step 5: Calculate Residuals.

The residuals for each episode are calculated as the difference between the standardized episode spending and the standardized predicted spending for episode i and IRF k.

Residual_ik = Y_ik - (Y_ik)^

where:

Y_ik is the attributed standardized spending for episode i and provider k.

(Y_ik)^ is the standardized predicted spending for episode i and provider k, as predicted from risk adjustment.

Step 6: Exclude Episodes with Outlier Residuals.

The next step excludes outliers from the calculation and renormalizes the resultant predicted values to maintain a consistent average episode payment level. Episodes with residuals below the 1st percentile or above the 99th percentile of the residual distribution are excluded, reducing the impact of high- and low-payment outliers on a PAC provider's measure. Predicted values after outlier exclusion are renormalized by multiplying each value by the ratio of the average standardized un-risk-adjusted payments to the average of the standardized predicted payments remaining after exclusion of episodes with outlier residuals.

Step 7: Calculate MSPB-PAC IRF Measure Score.

The MSPB-PAC IRF measure score is calculated for individual providers, allowing them to be compared relative to other IRF providers nationally. Mathematically, MSPB-PAC IRF for individual provider k is:

"MSPB-PAC IRF Amount"_k/"National Median MSPB-PAC IRF Amount"

The numerator is the MSPB-PAC IRF Amount, or the average risk-adjusted episode spending across all episodes for the attributed provider. This is then multiplied by the national average episode spending level for all IRF providers nationally. Mathematically, the MSPB-PAC IRF Amount numerator is calculated as:

"MSPB-PAC IRF Amount"_k=((1/n_k)*Sum_i_{i_k}[Y_ik/(Y_ik)^])*((1/N)*Sum_k[Sum_i_{i_k}[Y_ik]])

Where:

Y_ik is the attributed standardized spending for episode i and provider k.

(Y_ik)^ is the expected standardized spending for episode i and provider k, as predicted from risk adjustment, and resulting from Step 6 above

n_k is the number of episodes for provider k

N is the number of episodes nationally

i_{i_k} is all episodes i in the set of episodes attributed to provider k

The denominator is the episode-weighted national median of the MSPB-PAC IRF Amounts for all IRFs nationally.

The MSPB-PAC IRF measure score is calculated for each provider. An MSPB-PAC IRF measure score with a value less than 1 indicates that a given IRF's resource use is less, after risk-adjustment, than the resource use of the national median MSPB-PAC IRF Amount across all IRFs nationally in the given performance period.

Notes:

[1] CMS Manual Transmittal 29 – New Number Series and State Codes for CMS Certification Numbers (formerly OSCAR Provider Numbers) at: https://www.cms.gov/Regulations-and-Guidance/Guidance/Transmittals/downloads/R29SOMA.pdf

[2] QualityNet, "CMS Price (Payment) Standardization – Detailed Methods" (Revised April 2019) https://www.qualitynet.org/inpatient/measures/payment-standardization

[3] Centers for Medicare and Medicaid Services. Measure Specifications: Medicare Spending Per Beneficiary-Post-Acute Care Resource Use Measures. April 2016. Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/2016_04_06_mspb_pac_measure_specifications_for_rulemaking.pdf

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:

Please supply the username and password:

Attachment: S_7_2_Construction_Logic.docx

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.)*

We do not provide specifications for concurrency of clinical events.

The MSPB-PAC IRF measure methodology does not separate concurrent events. The MSPB-PAC IRF measure methodology defines an MSPB episode as all claims from the start of admission to the IRF to 30 days post IRF discharge. Please refer to section S.8.4., which details the rationale for the construction of the MSPB-PAC IRF episode, for a discussion of the advantages of this approach.

The definition of MSPB-PAC IRF episodes allow one episode to overlap with other episodes. One possible scenario occurs where an IRF provider discharges a beneficiary who is then admitted to another IRF within 30 days. In this case, the second episode begins in the associated services period of the first episode in the 30 day post-treatment. The IRF stay will be included once as an associated service for the attributed provider of the first MSPB-PAC IRF episode and once as a treatment service for the attributed provider of the second MSPB-PAC IRF episode. This overlap is necessary to ensure continuous accountability between providers throughout a beneficiary's trajectory of care, as both providers share incentives to deliver high quality care at a lower cost to Medicare and engage in patient-focused care planning and coordination.

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

We do not provide specifications for linking complementary services.

An MSPB-PAC IRF episode includes all Medicare Part A and Part B services that fall within the episode window that starts at the IRF index admission and ends at 30-days post IRF discharge, except for a limited set of services that are excluded for being clinically unrelated to IRF treatment (as described in section S.9.1).

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

Clinical Classification Software (CCS). We use CCS for Services and Procedures to group HCPCS codes on outpatient (OP) claims that occur during the episode into clinically meaningful categories. This grouping is used to identify clinically unrelated events which are then excluded from episode calculation (as detailed in section S.9.1).

Hierarchical Condition Categories (HCC). Hierarchical Condition Categories with a 90-day lookback period are included as covariates in the risk-adjustment model. The MSPB-PAC IRF risk adjustment methodology is discussed in additional detail in section S.12.

Clinical Case-Mix Category. The clinical case-mix category variables used in the MSPB-PAC IRF risk-adjustment model are included to account for beneficiary characteristics prior to the start of an MSPB-PAC IRF episode that may influence the type and intensity of care. Taking the most recent institutional claim (by end date) in the 60 days prior to the start of an MSPB-PAC IRF episode, the episode is assigned to one of the following mutually exclusive and exhaustive clinical case-mix categories:

(1) Prior Acute Surgical IP – Orthopedic – beneficiaries who have most recently undergone orthopedic surgery in an acute inpatient hospital

(2) Prior Acute Surgical IP – Non-Orthopedic – beneficiaries who have most recently undergone a non-orthopedic surgery in an acute inpatient hospital

(3) Prior Acute Medical IP with Intensive Care Unit (ICU) – beneficiaries who have most recently stayed in an acute inpatient hospital for non-surgical reasons and had a stay in the ICU

(4) Prior Acute Medical IP without ICU – beneficiaries who have most recently stayed in an acute inpatient hospital for non-surgical reasons but did not have a stay in the ICU

(5) Prior PAC - Institutional – beneficiaries who are continuing PAC from an institutional PAC setting (i.e., coming from an LTCH, IRF, or SNF)

(6) Prior PAC - HHA – beneficiaries who are continuing PAC from a HHA

(7) Community – all other beneficiaries

In the event that there are multiple prior claims with the same end date in the 60 days prior to the start of a PAC episode, additional logic is employed to determine the episodes' clinical case-mix category. For conflicts occurring between two IP claims, the clinical case-mix category corresponding to the claim with the longest length of stay (LOS) is assigned. For all other types of conflicts including those where the LOS is the same between two IP claims, the clinical case-mix category is assigned using a hierarchy in the order of the categories listed above.

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 provide measure specifications or guidelines for missing data :

Accurate and complete Part A and Part B claims are necessary for physician and hospital billing. Thus, missing data on Medicare enrollment and claims are very rare. All the data used to calculate MSPB-PAC IRF measure values are included on Medicare claims and enrollment data. The data fields used to calculate the MSPB-PAC IRF measure (e.g., payment amounts, DRGs, diagnosis and procedure codes, etc.) are included in all Medicare claims because IRFs only receive payments for complete claims. We do have complete data for each beneficiary who has an MSPB-PAC IRF episode since beneficiaries are excluded if they are not continuously enrolled in only Medicare Parts A and B or if Medicare is not the primary payer during an episode, as described in section S.9.1. This ensures that we have all claims data for beneficiaries included in the MSPB-PAC IRF measure calculation.

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

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

Durable Medical Equipment (DME)

Other services not listed

All services covered by Medicare Part A and B (Hospice, SNF, Home Health, and services captured in carrier claims.)

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.)

The MSPB-PAC IRF measure assesses the standardized allowed amounts of services performed by IRFs and other healthcare providers during an MSPB-PAC IRF episode, which includes all Part A and Part B Medicare claims that occur at the IRF admission through 30 days after discharge from the IRF stay. As a result, costs from all Part A and Part B claim types (i.e., inpatient, outpatient, home health agency, hospice, skilled nursing facility, durable medical equipment, and carrier) are included. Note that costs of Part B drugs are included, but costs of Part D drugs are not included since Part D is not used to calculate the MSPB-PAC IRF measure. The methodology used to standardize payment for these claims is available for download from the URL provided in section S.7.8a ("CMS Price (Payment) Standardization").

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.org/inpatient/measures/payment-standardization

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.)

The MSPB-PAC IRF measure aims to calculate resource use in the period between the start of the treatment period and the end of the associated services period. The clinical topic area includes all IRF admissions in the United States. To adjust for beneficiary characteristics that are out of the influence of the attributed IRF and may affect resource use, we risk-adjust the total observed episode spending (described in section S.12) using CMS-HCC indicators and interactions between selected comorbidities. In addition to comorbidities, we also include indicators for clinical case-mix based on diagnosis and procedural information on the most recent institutional claim (by end date) in the 60 days prior to the start of an MSPB-PAC IRF episode. Finally, we account for Rehabilitation Impairment Categories (RICs) identified in the IRF admission claim. The MSPB-PAC IRF episode encompasses all procedures and clinical events that occur between the start of the treatment period and 30 days post IRF discharge.

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.)*

In order to create a resource use measure that is clinically valid, there were multiple steps involved in excluding the least clinically relevant codes. Using an episode window, we organized claims into clinically meaningful service categories or settings. For example, Medicare Severity-Diagnosis Related Groups (MS-DRGs) noted after an IRF discharge were evaluated as medical or surgical admissions post-discharge. Clinical Classifications Software (CCS) and Current Procedural Terminology/Healthcare Common Procedure Coding System (CPT/HCPCS) services were organized into outpatient services, emergency department (ER) services, and durable medical equipment claims and evaluated for their relevance or relatedness to IRF care.

Extensive clinical review was performed by clinicians with experience providing care in IRF settings, as well as in collaboration with Medical Officers at CMS. The hospitalizations and outpatient services least clinically related to the IRF care were excluded from resource use calculation. For instance, it was not felt that an IRF could influence a beneficiary's rehospitalization for nervous system neoplasms (DRG 054), post-discharge outpatient services for kidney transplant (CCS 105), or routine fecal occult blood testing (CPT 82270). Therefore, these types of services were excluded. Services were only added to the exclusions list if there was consensus across IRF and CMS clinicians. Please see section S.9.1 for overall clinical consensus regarding the types of exclusions.

To account for the association between clinical severity and resource use, we risk adjust the total observed episode spending (described in section S.12) using CMS-HCC indicators and interactions between selected comorbidities. Diagnosis codes on claims that occur during the 90-day period prior to the start of an MSPB-PAC IRF episode (90-day "look back") are used to create HCC indicators. The MSPB-PAC IRF measure accounts for comorbid conditions and interactions by broadly following the CMS-HCC risk adjustment methodology, which is derived from Medicare Part A and B claims and is used in the Medicare Advantage (MA) program. For example, the measure accounts for interactions between disability and selected HCC groups (e.g., Cystic Fibrosis, Severe Hematological Disorders, Opportunistic Infections, among others). Given the fact that beneficiaries often have more than one comorbidity, the model also includes commonly observed paired condition interactions, (e.g., chronic obstructive pulmonary disease [COPD] and congestive heart failure [CHF]) and commonly observed triple-interactions (e.g., diabetes mellitus, congestive heart failure, and renal failure). The full list of variables used in the risk adjustment model can be found in the Measure Specifications document provided in section S.1.

In addition to comorbidities, the MSPB-PAC IRF measure utilizes clinical case-mix categories to create clinically meaningful subgroups that influence the type of services a beneficiary will receive in an IRF. To create these subgroups, information was derived from the institutional claim of the most recent hospitalization. The clinical case-mix category variables used in the MSPB-PAC IRF risk-adjustment model are included to account for differences in intensity and type of care received by beneficiaries prior to the start of an MSPB-PAC IRF episode. Taking the most recent institutional claim (by end date) in the 60 days prior to the start of an MSPB-PAC IRF episode, the episode is assigned to one of the following mutually exclusive and exhaustive clinical case-mix categories:

1) Prior Acute Surgical IP – Orthopedic – beneficiaries who have most recently undergone orthopedic surgery in an acute inpatient hospital

2) Prior Acute Surgical IP – Non-Orthopedic – beneficiaries who have most recently undergone a non-orthopedic surgery in an acute inpatient hospital

3) Prior Acute Medical IP with ICU – beneficiaries who have most recently stayed in an acute inpatient hospital for non-surgical reasons and had a stay in the ICU

4) Prior Acute Medical IP without ICU – beneficiaries who have most recently stayed in an acute inpatient hospital for non-surgical reasons but did not have a stay in the ICU

5) Prior PAC - Institutional – beneficiaries who are continuing PAC from an institutional PAC setting (i.e., coming from an LTCH, IRF, or SNF)

6) Prior PAC - HHA – beneficiaries who are continuing PAC from a HHA

7) Community – all other beneficiaries

Finally, the MSPB-PAC IRF measure uses RICs from the IRF admission. A full list of the RICs used in the risk adjustment model is included in Appendix C of the Measure Specifications document provided in section S.1.

To simplify the clinical logic and avoid the issue of attributing claims to MSPB-PAC IRF episodes in the case of concurrent clinical events, all claims that begin within the episode window (treatment period and associated services period) are included in the MSPB-PAC IRF measure. An MSPB-PAC IRF episode is assigned to the rehabilitation facility of the index admission. A new episode may begin during the associated services period of a previous MSPB-PAC IRF episode in the 30 days post-discharge from the IRF.

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)*

As part of the IMPACT Act, the goals of MSPB-PAC were to standardize assessment data to allow for interoperability, data exchange, and standardized measurement among post-acute providers. It also mandated the use of quality measures for PAC.[1] This would ultimately allow for greater ability to measure resource use and efficiency of care to improve outcomes, as well as encourage all PAC providers towards aligned incentives and care coordination.

There have been a number of studies demonstrating significant variability in IRF care and outcomes, links between facility characteristics and readmissions, and significant opportunities for improvement.[2,3,4] The cost and quality link is important, with this resource use measure playing an important role in discerning value of IRF care. Indeed, post-acute care is a significant cause of variation in Medicare spending.[5,6]

Within PAC, accounting for the beneficiary's clinical severity is crucial for accurately predicting resource use within the PAC episode. There is ample evidence in both inpatient and post-acute settings that Medicare episode payments are associated with clinical case-mix or severity. For example, Vertrees et al. (2013) tested the relationship between case-mix and Medicare payments for acute hospitalization episodes using MS-DRGs and Clinical Risk Groups (CRGS), which are similar to the HCC groupings. This analysis found that Medicare costs of acute-hospitalization episodes that use 30, 60, and 90-days post-discharge windows can be predicted in part by the use of clinical severity groupings and case mix indicators as risk adjustors.[7] Clinical severity groupings, such as those defined in section S.8.2, have been shown to be associated with resource use in the post-acute settings.

Within this framework, the MSPB-PAC IRF measure was created to assess care provided by inpatient rehabilitation facilities. Prior literature has shown that patient severity is also an important consideration in this setting. Ramey et. al (2016) noted that there is significant variation of readmission across IRFs. They found that 41.4% of this variation was due to patient and facility characteristics and that IRFs that had high risk of readmission were noted to care for the more medically complex patients.[8] Accounting for clinical severity and case-mix is crucial to estimating the expected Medicare cost of the episode and thus supports the intent of the MSPB-PAC IRF quality measure.

The MSPB-PAC IRF measure methodology defines an MSPB-PAC IRF episode as all claims with start dates falling between the treatment period start date (date of admission to IRF) and the associated services period end date (30 days post IRF discharge). This episode definition is consistent with NQF's theoretical definition of an episode of care in that it is "…a series of temporally contiguous healthcare services related to the treatment of a given spell of illness or provided in response to a specific request by the patient or other relevant entity."[9] Moreover, NQF has endorsed multiple episode-based measure of resource use and cost, including

the MSPB-Hospital Measure (NQF #2158) and hospital-level measures for episodes of care for pneumonia (NQF #2579), acute myocardial infarction (NQF #2431), and heart failure (NWF #2436). Each of these measures estimate the risk-adjusted cost of a hospital based episode of care covering 30 days post-admission or post-discharge.

Notes:

[1] Medicare Program; Inpatient Rehabilitation Facility Prospective Payment System for Federal Fiscal Year 2017 Federal Register, Vol. 81, No. 151. https://www.gpo.gov/fdsys/pkg/FR-2016-08-05/pdf/2016-18196.pdf

[2] Middleton, A., Graham, J. E., Prvu Bettger, J., Haas, A., & Ottenbacher, K. J. (2018). Facility and Geographic Variation in Rates of Successful Community Discharge After Inpatient Rehabilitation Among Medicare Fee-for-Service Beneficiaries. JAMA Network Open, 1(7), e184332. https://doi.org/10.1001/jamanetworkopen.2018.4332

[3] Middleton A, Graham JE, Deutsch A, O. K. (2017). Potentially Preventable Within-Stay Readmissions Among Medicare Fee-for-Service Beneficiaries Receiving Inpatient Rehabilitation. PM R, 9(11), 1095–1105.

[4] Daras, L. C., Ingber, M. J., Deutsch, A., Hefele, J. G., & Perloff, J. (2018). Geographic Region and Profit Status Drive Variation in Hospital Readmission Outcomes Among Inpatient Rehabilitation Facilities in the United States. Archives of Physical Medicine and Rehabilitation, 99(6), 1060–1066.

[5] Institute of Medicine. (2013). Variation in Health Care Spending Assessing Geographic Variation. (July)

[6] Kahn, E. N., Ellimoottil, C., Dupree, J. M., Park, P., & Ryan, A. M. (2018). Variation in payments for spine surgery episodes of care: Implications for episode-based bundled payment. Journal of Neurosurgery: Spine, 29(2), 214–219.

[7] Vertrees, J. C., Richard, A. F., Eisenhandler, J, Quain, A., & Switalski, J. (2013). Bundling post-acute care services into MS-DRG payments. Medicare & Medicaid Research Review 3, no. 3

[8] Ramey, L., Goldstein, R., Zafonte, R., Ryan, C., Kazis, L., & Schneider, J. (2016). Variation in 30-Day Readmission Rates Among Medically Complex Patients at Inpatient Rehabilitation Facilities and Contributing Factors. Journal of the American Medical Directors Association, 17(8), 730–736.

[9] National Quality Forum. (2010). Measurement framework: Evaluating efficiency across patient-focused episodes of care. In Patient-Focused Episodes of Care. Retrieved from http://www.qualityforum.org/Publications/2010/01/Measurement_Framework__Evaluating_Efficiency_Across_Patient-Focused_Episodes_of_Care.aspx

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: MSPB-PAC_IRF_--_NQF_Testing_Attachment_-_Appendix_Tables.xlsx

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

Trigger Event: IRF admission

Start Date: Date of the IRF admission

End Date: 30 days after discharge from the IRF stay.

MSPB-PAC IRF episodes include services that take place during the time period 30 days post-IRF discharge in order to emphasize the importance of care transitions and care coordination in improving post-acute care and reducing unnecessary service use. As a result, services with claim start dates on or between the IRF admission date and 30 days after IRF discharge are attributed to the IRF episode. This timeframe was selected to align with other measures and was felt to be long enough to capture clinically relevant events, but not so long as to reduce the attributed facility's influence in future events.

The advantages of this measure trigger and end mechanism are twofold. First, this approach is simple and easily implementable since it includes all claims, except those described in section S.9.1, during the MSPB-PAC IRF episode. Second, the MSPB-PAC IRF approach incorporates costs due to care complications unrelated to the original reason for rehabilitation services, encouraging IRF care coordination.

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

Clinical Severity levels are embedded in the risk-adjustment model, as described in section S.12.

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

Co-morbidities and disease interactions are accounted for in the MSPB-PAC IRF measure risk adjustment methodology, as discussed in sections S.8.2 and S.12. Conditions which most directly impact beneficiaries' health status at the time of the IRF admission are captured in the risk-adjustment via the 90-day look-back period prior to the start of an episode. Because the relationship between comorbidities and episode cost may be non-linear in some cases (i.e., beneficiaries may have more than one disease during an IRF episode), the model also takes into account a limited set of interactions between HCCs and/or enrollment status variables. Example variable interaction terms include Diabetes Mellitus/Congestive Heart Failure, Renal Failure/Congestive Heart Failure, and Disability/Opportunistic Infections. (For a complete list of these variable interaction terms and other risk adjustment variables, please refer to Appendix C of the Measure Specifications document provided in section S.1). The MSPB-PAC IRF measure risk adjustment methodology includes only a limited set of interaction terms for two reasons. First, inclusion of too many interaction terms will over-fit the model. Second, the MSPB-PAC IRF measure risk adjustment methodology broadly follows the established CMS-HCC risk adjustment methodology, which uses similar interaction terms.

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)*

:

Exclusion of clinically unrelated services. Certain services are excluded from the MSPB-PAC IRF episodes because they are clinically unrelated to IRF care and/or because IRF providers may have limited influence over certain Medicare services delivered by other providers during the episode window. These limited service-level exclusions are not counted towards a given IRF provider's Medicare spending to ensure that beneficiaries with certain conditions and complex care needs receive the necessary care. The list of excluded services was developed by obtaining consensus on the exclusion of each service from CMS clinicians, eight independently contracted clinicians (including two TEP members) with expertise in each of the PAC settings, and the measure developer's clinicians. Feedback from the TEP provided through the in-person meeting and follow-up email survey was also taken into consideration. Additional information on the process for developing the list of clinically unrelated services is available in Appendix D of the Measure Specifications document provided in section S.1. The specialties of the non-CMS clinicians with whom we consulted during the measure

development process are provided in Appendix F of the Measure Specifications document provided in section S.1. Services that were determined by clinical consensus to be outside of the control of PAC providers include:

- Planned hospital admissions[1]
- Routine management of certain preexisting chronic conditions (e.g., dialysis for end-stage renal disease (ESRD), enzyme treatments for genetic conditions, treatment for preexisting cancers, and treatment for organ transplants)
- Some routine screening and health care maintenance (e.g., colonoscopy and mammograms)
- Immune modulating medications (e.g., immunosuppressants for organ transplant or rheumatoid arthritis)

Other Exclusions. Once clinically unrelated services are excluded at the claim line level, we exclude episodes based on several other characteristics, such as:

1) Any episode that is triggered by a PAC claim outside the 50 states, D.C., Puerto Rico, and U.S. Territories.

Rationale: This exclusion ensures that complete claims data are available for each provider.

2) Any episode where the claim(s) constituting the attributed PAC provider's treatment have a standard allowed amount of zero or where the standard allowed amount cannot be calculated.

Rationale: Episodes where the claim(s) constituting the attributed PAC provider's treatment are zero or have unknown allowed payment do not reflect the cost to Medicare. Including these episodes in the calculation of MSPB-PAC IRF measure could potentially misrepresent a providers' resource use.

3) Any episode in which a patient is not enrolled in Medicare FFS for the entirety of a 90-day lookback period (i.e., a 90-day period prior to the episode trigger) plus episode window (including where a beneficiary dies) or is enrolled in Part C for any part of the lookback period plus episode window.

Rationale: Episodes meeting this criteria do not have complete claims information that is needed for risk-adjustment and the measure calculation as there may be other claims (e.g., for services provided under Medicare Advantage [Part C]) that we do not observe in the Medicare Part A and B claims data. Similarly, episodes in which the patient dies are, by definition, truncated episodes and do not have a complete episode window. Including these episodes in the MSPB-PAC IRF measure could potentially misrepresent a provider's resource use. This exclusion also allows us to faithfully construct Hierarchical Condition Categories (HCCs) for each episode by scanning the lookback period prior to its start without missing claims.

4) Any episode in which a patient has a primary payer other than Medicare for any part of the 90-day lookback period plus episode window.

Rationale: When a patient has a primary payer other than Medicare, complete claims data may not be observable. These episodes are removed to ensure that the measures are accurately calculated using complete data.

5) Any episode where the claim(s) constituting the attributed PAC provider's treatment include at least one related condition code indicating that it is not a prospective payment system bill.

Rationale: Claims that are not a prospective payment system bill may not report sufficient information to allow for payment standardization.

6) Any episode with problematic claims data (e.g., anomalous records for stays that overlap wholly or in part, or are otherwise erroneous or contradictory)

Rationale: The episode with the most recent processing date is kept to ensure the accuracy of data elements.

Finally, as part of the measure construction process described in section S.7.2, episodes with residuals below the 1st or above the 99th percentile of the residual distribution are excluded, reducing the impact of high- and low-payment outliers.

Notes:

[1] The lists of clinically unrelated services built off the planned readmissions algorithm developed by the Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation, as well as the expansions to the Yale algorithm by RTI. Clinicians reviewed the list of exclusions from that algorithm in the context of PAC treatment. During the review process, clinicians reviewed admissions observed in MSPB-PAC episodes and created exclusions that overlap with the Yale algorithm. Details on the Yale and RTI algorithms are available here: "Hospital-Wide All-Cause Unplanned Readmission Measure - Version 4.0," in 2015 Measure Updates and Specifications Report, ed. Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation (2015). 10-11. Laura Smith, West, S., Coots, L., Ingber, M., "Skilled Nursing Facility Readmission Measure (SNFRM) NQF #2510: All-Cause Risk-Standardized Readmission Measure," (Centers for Medicare & Medicaid Services, 2015). 5-6

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)*

Not applicable: the MSBP-PAC IRF measure is not stratified.

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

As discussed in section S.7.2, the MSPB-PAC IRF measure removes sources of variation which are not directly related to decisions to utilize care, such as local or regional price differences, to capture differences in beneficiary resource use that an IRF can influence through appropriate practices and care coordination. The MSPB-PAC IRF measure relies on a detailed price standardization methodology to exclude geographic payment rate differences; in other words, the MSPB-PAC IRF measure adjusts observed payments for Medicare geographic adjustment factors.[1]

Notes:

[1] QualityNet, “CMS Price (Payment) Standardization – Detailed Methods” (Revised April 2019) https://www.qualitynet.org/inpatient/measures/payment-standardization

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

Ratio

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.)*

An MSPB-PAC IRF measure score of 1 indicates that an IRF had an average MSPB-PAC Amount (i.e., risk-adjusted spending level) which is equal to the national episode-weighted median MSPB-PAC Amount across all

IRFs during a given performance period. An MSPB-PAC IRF measure score of greater than 1 indicates that an IRF had higher average risk-adjusted spending levels compared to those of the national median IRF. For example, a measure score of 1.1 indicates that the IRF had average risk-adjusted spending levels that are 10 percent higher than the median IRF. On the other hand, an MSPB-PAC IRF measure score of less than 1 indicates that an IRF had lower average risk-adjusted spending levels compared to those of the median IRF. For example, a measure score of 0.9 indicates that the IRF had average risk-adjusted spending levels that are 10 percent lower than the median IRF.

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

The detailed steps to computing the measure score are described in section S.7.2. Risk-adjustment is applied in "Step 3: Calculate Predicted Episode Payments." The purpose of risk adjustment is to compensate for patient health circumstances and demographic factors that affect resource use but are beyond the influence of the attributed provider. The MSPB-PAC IRF measure risk adjustment model is adapted from the model used in the NQF-endorsed MSPB-Hospital measure, which itself is an adaptation of the standard CMS-HCC risk-adjustment model.[1,2] The MSPB-PAC IRF model uses a linear regression framework and a 90-day HCC lookback period. The risk adjustment model is estimated on all MSPB-PAC IRF episodes that meet the exclusion criteria. Each provider's MSPB-PAC IRF measure score is calculated as a provider's average MSPB-PAC Amount divided by the median MSPB-PAC Amount across all providers. A provider's MSPB-PAC IRF Amount is defined as the sum of standardized, risk-adjusted spending across all of a provider's eligible episodes divided by the number of episodes for that provider. Below is a description of the risk adjustment variables.

Risk-Adjustment Variables

The following beneficiary health status indicators are included as covariates in each MSPB-PAC IRF risk adjustment model and to the greatest extent possible are consistent across PAC settings (see Appendix C of the Measure Specifications document provided in section S.1 for a comprehensive list of independent variables used in the risk adjustment model):

- 70 HCCs
- 11 HCC interactions
- 11 brackets for age at the start of the episode
- Original entitlement to Medicare through disability
- ESRD status
- Long-term care institutionalization at start of episode.[3]
- Six clinical case-mix categories reflecting recent prior care (described further below).[4]
- Hospice utilization during the episode
- Prior acute ICU utilization day categories
- Prior acute length of stay categories
- Rehabilitation Impairment Categories (RICs)

The clinical case-mix category variables used in the MSPB-PAC IRF risk adjustment model are included to account for differences in intensity and type of care received by beneficiaries prior to the start of an MSPB-PAC IRF episode. See section S.7.5 for more details on the methodology of assigning clinical case-mix categories to each episode.

Notes:

[1] QualityNet, "CMS Price (Payment) Standardization – Detailed Methods" (Revised April 2019) https://www.qualitynet.org/inpatient/measures/payment-standardization

[2] CMS, “Medicare Risk Adjustment Information” (2016) https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Risk-Adjustors.html

[3] Identifies beneficiaries who have been institutionalized for at least 90 days in a given year. The indicator is based on 90-day assessments from the Minimum Data Set (MDS) and is calculated based on CMS’ definition of institutionalized individuals.

[4] There are 7 case-mix categories as described above, but one category is removed to prevent collinearity.

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).

MSPB-PAC IRF measure scores are reported publicly for providers with 20 or more eligible episodes, along with the national average score. The distribution of MSPB-PAC IRF measure scores (based on FY 2016-2017 data) that are statistically significantly different from the national average is as follows:

- Significantly lower than the national average: 30%
- Not statistically different from the national average: 24%
- Significantly higher than the national average: 46%

Inference about both measure performance of individual providers can be made based on the score value. The distribution of MSPB-PAC IRF measure score values based on FY 2016-2017 data is as follows:

- Minimum: 0.74
- 10th Percentile: 0.91
- 25th Percentile: 0.95
- 75th Percentile: 1.05
- 90th Percentile: 1.10
- Maximum: 1.47

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.

Each MSPB-PAC IRF episode is attributed to the IRF whose inpatient admission claim triggers the episode. Adjacent readmissions for the same patient and provider are treated as part of the same treatment period to reflect the likelihood that these closely adjacent stays are related. For gaps of 7 or fewer days, stays in the same setting with the same patient and provider are collapsed into one treatment period. Stays with a gap of 8 or more days trigger separate episodes.

The definition of MSPB-PAC IRF episodes allows episodes to overlap with hospital and other MSPB-PAC episodes. MSPB-PAC IRF episodes may begin within 30 days of discharge from an inpatient hospital as part of a patient’s trajectory from an acute to a PAC setting. An MSPB-PAC IRF stay beginning within 30 days of discharge from an inpatient hospital will therefore be included once in the hospital’s MSPB-Hospital measure and once in the PAC provider’s MSPB-PAC measure. Aligning the MSPB-Hospital and MSPB-PAC measures in this way creates continuous accountability and aligns incentives to improve care planning and coordination across inpatient and PAC settings.

Additionally, an MSPB-PAC episode may begin during the associated services period of another MSPB-PAC episode in the 30 days post-discharge. One possible scenario occurs where, for example, an IRF provider discharges a beneficiary who is then admitted to another IRF within 30 days. The IRF claim would be included once as an associated service for the attributed provider of the first MSPB-PAC IRF episode and once as treatment services for the attributed provider of the second MSPB-PAC IRF episode.

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.

The peer group for this measure includes all inpatient rehabilitation facilities in the United States that are Medicare-certified. Any Medicare-certified IRF that submits a PAC claim during the measure performance period can be included in this measure. The rationale for identifying this peer group is that under the Improving Medicare Post-Acute Care Transformation Act (IMPACT) of 2014, IRFs (and other PAC providers) are required to report data on quality, resource use, and other measures. The MSPB-PAC IRF measure was created to fulfill the statutory requirement for IRFs to submit measures of resource use for public reporting. As such, IRFs reporting the MSPB-PAC IRF measure can compare their performance relative to all other Medicare-certified IRFs in the United States.

S.13.4. Sample size

Detail the sample size requirements for reporting measure results.

MSPB-PAC IRF measure scores are publicly reported on IRF Compare for IRFs with 20 or more eligible episodes. Out of 1,161 IRFs with FY 2016-2017 episodes, only 14 did not meet this minimum threshold.

S.13.5. Define benchmarking and comparative estimates

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

The MSPB-PAC IRF measure itself is not calculated using benchmarks but rather is a comparison between a given IRF's MSPB-PAC IRF Amount and national episode-weighted median MSPB-PAC IRF Amount. The measure score is expressed as a ratio to that national amount, wherein a measure ratio of less than one indicates lower Medicare spending than the national median, a ratio of one indicates spending that is equivalent to the national median, and a ratio of greater than one indicates spending that is greater than the national median.

Validity – See attached Measure Testing Submission Form

SA.1. Attach measure testing form

MSPB-PAC_IRF_--_NQF_Testing_Attachment_2020_04_27.docx

Measure Number (*if previously endorsed*)**:** 3561
Measure Title: Medicare Spending per Beneficiary Post-Acute Care Measure for Inpatient Rehabilitation Facilities
Date of Submission: 1/2/2020

Type of Measure:

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

Instructions

- **Measures must be tested for all the data sources and levels of analyses that are specified. *If there is more than one set of data specifications or more than one level of analysis, contact NQF staff* about how to present all the testing information in one form.**
- **For all measures, sections 1, 2a2, 2b1, 2b2, and 2b4 must be completed.**
- **For outcome and resource use measures, section 2b3 also must be completed.**
- **If specified for multiple data sources/sets of specificaitons (e.g., claims and EHRs), section 2b5 also must be completed.**
- **Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b1-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.**
- **If you are unable to check a box, please highlight or shade the box for your response.**
- **Maximum of 25 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). *Contact NQF staff if more pages are needed.***
- **Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.**
- **For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for version 7.1 of the Measure Testing Attachment.**

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.

2b1. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument-based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b3. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration**

OR

- **rationale/data support no risk adjustment/ stratification.**

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful [16] differences in performance;

OR

there is evidence of overall less-than-optimal performance.

2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b6. Analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions.

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of

care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

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; Minimum Data Set (MDS).	☒ other: Medicare enrollment database; Minimum Data Set (MDS); Provider of Services File; American Community Survey; Rural Urban Continuum Codes; IRF Compare; and Common Medicare Environment (CME) database.

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 Medicare Spending per Beneficiary Post-Acute Care for Inpatient Rehabilitation Facilities (MSPB-PAC IRF) measure is based on Medicare fee-for service (FFS) administrative claims and uses data in the Medicare enrollment database and Minimum Data Set (MDS). The enrollment files provide information such as date of birth, date of death, sex, reason for Medicare eligibility, and enrollment in Medicare FFS. The MDS is used to construct a risk adjustment variable, indicating beneficiaries who have been institutionalized for at least 90 days in a given year. The data elements from the Medicare FFS claims are those basic to the operation of the Medicare payment systems and include data such as date of service, date of admission, date of discharge, diagnoses, procedures, and revenue center codes. The Medicare FFS claims data files are used to identify Medicare services from IRF and other settings (e.g., inpatient and outpatient hospitals) within the episode

window. No data beyond what agencies submit in of the normal course of business are required to calculate this measure.

This measure submission is based on fiscal year (FY) 2016-2017 data, which were the most recent data available at the time of our analyses. We used the data sources listed below to develop the claims analytic file for measure specification and testing:

- **Medicare Fee-For-Services claims and enrollment data**: We access inpatient, outpatient, carrier, skilled nursing facility, home health, durable medical equipment, and hospice claims through the Centers for Medicare & Medicaid Services (CMS) Common Working File (CWF). The data dictionary for all Medicare FFS claims, demographic, and enrollment data are available at: https://www.resdac.org/cms-data?tid%5B%5D=4931&tid_1%5B%5D=1&=Find+Data+Files. General information about the CWF is available at: https://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/Downloads/clm104c27.pdf.

- **Minimum Data Set (MDS)**: Acumen obtains the MDS through the Quality Improvement and Evaluation System (QIES). The data dictionary for the MDS data is available at: https://www.resdac.org/cms-data/files/mds-3.0/data-documentation.

We used two mappings to group diagnosis and procedure codes for use in identifying clinical events, implementing exclusions and applying risk adjustment:

- **Agency for Healthcare Research and Quality (AHRQ) Clinical Classifications Software (CCS) groupings for Services and Procedures**: Software is available for download at: https://www.hcup-us.ahrq.gov/toolssoftware/ccs_svcsproc/ccssvcproc.jsp

- **CMS-Hierarchical Condition Category (HCC) mappings of ICD-9 and ICD-10 codes**: We used the Version 22 CMS-HCC mapping, which is included in the software available at: https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Risk-Adjustors.html.

We used five additional data sources for measure testing only, not for specification:

- **2017 American Community Survey (ACS) 5-year estimate**: We used the ACS to obtain the ZIP Code Tabulation Area (ZCTA) level measures needed to compute the Agency for Healthcare Research and Quality (AHRQ) Socioeconomic Status (SES) index score for social risk factor testing. This information is available on the US Census website: http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t

- **Rural-Urban Continuum Codes 2013**: We used this data source to construct rural-urban identifiers to test the impact of social risk factors on measure performance. These codes include county FIPS indicators which are then merged onto the episode files. Additional information on this data source can be found at: https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/

- **Provider of Services Current Files (POS File)**: We used this data source to describe the characteristics of Inpatient Rehabilitation Facilities (IRFs) included in specification and testing, such as census region, ownership type, and rurality, reported in Table 1. The POS file contains data on characteristics of hospitals and other types of healthcare facilities, including the name and address of the facility and the type of Medicare services the facility provides, among other information. The data are collected through the CMS Regional Offices. General information about the POS Files is available at: https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/index.html

- **IRF Compare data**: We used this data source to examine the relationship between MSPB and assessment-based quality measures. The IRF Compare data include publicly reported IRF quality measures. The data are available at https://data.medicare.gov/data/inpatient-rehabilitation-facility-compare

- **Common Medicare Environment (CME) database**: We extracted patient-level dual eligibility information from the CME database for social risk factor testing. CMS has designated the CME database as the single, enterprise-wide authoritative source for Medicare beneficiary enrollment and demographic data. The CME database integrates and standardizes different types of beneficiary data from CMS legacy systems. The CME database receives information from the Enrollment Database (EDB) and also contains additional information not available in the EDB. Description of the CME is available at: https://www.ccwdata.org/documents/10280/19002256/medicare-enrollment-impact-of-conversion-from-edb-to-cme.pdf

What are the dates of the data used in testing?

Fiscal years 2016 and 2017

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: Click here to describe	☐ other: Click here to describe

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*)

This measure is based on national data. All IRFs paid under Medicare's IRF Prospective Payment System (PPS) and included in the IRF Quality Reporting Program (QRP) were included, provided they had an eligible episode. A total of 1,161 IRFs with eligible episodes in FY 2016-2017 were included in measure specification, testing, and analysis, with differences noted in **section 1.7**. **Table 1** summarizes the frequency of IRFs by census region, ownership type, and rurality. Throughout this form, we use "K" to refer to number of providers.

Table 1. Characteristics of IRFs included in Specification and Testing of the FY 2016-2017 MSPB-PAC IRF Measure (K = 1,161*)

Characteristic	K (%)
Census Region	
New England	34 (2.93%)
Mid Atlantic	149 (12.83%)
East North Central	199 (17.14%)
West North Central	96 (8.27%)
South Atlantic	163 (14.04%)
East South Central	77 (6.63%)
West South Central	231 (19.90%)
Mountain	88 (7.58%)
Pacific	104 (8.96%)
U.S. Territories	6 (0.52%)
Ownership type	
Government	121 (10.42%)
Proprietary	382 (32.90%)
Not-For-Profit	635 (54.69%)
Physician Ownership	6 (0.52%)
Tribal	1 (0.09%)
Rurality	
Rural	144 (12.40%)
Urban	1,003 (86.39%)

Analysis of Medicare Claims File for IRF FY 2016-2017 and 2016-2017 POS.

*Provider census region and rurality information was not available for 14 IRFs. Provider ownership type information was not available for 16 IRFs.

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*)

Measure specification and testing was based on national data. All eligible IRF patient episodes with IRF discharges between October 1, 2015, and September 30, 2017, were included in the measure. For patients with multiple IRF episodes during the measurement period, all eligible episodes were included. Readmissions of the same patient to the same provider within 7 or fewer days after discharge do not trigger a new episode and instead were included in the treatment period of the original episode to reflect the likelihood that these closely adjacent stays are related. A total of 630,737 patient episodes were included after sample exclusions were applied. Note that the MSPB score calculation removes outliers at the 1st and 99th percentile of the residual distribution and subsequent testing and analyses were conducted on these 618,123 episodes. **Table 2** presents demographic characteristics of the patient episodes. Throughout this form, we use "N" to refer to number of patient episodes.

Table 2. **Demographic Characteristics of Episodes Included in Specification and Testing of the MSPB-PAC IRF Measure (N = 630,737)**

Characteristic	N (%)	Characteristic	N (%)
Male	277,802 (44.04%)	**Race***	
Female	352,935 (55.96%)	**White**	532,063 (84.36%)
Male under 65 years	37,648 (5.85%)	**Black**	65,840 (10.44%)
Female under 65 years	36,880 (5.97%)	**Other**	28,592 (4.53%)

Analysis of Medicare Claims File for IRF FY 2016-2017 and Medicare Enrollment database.

*Race information was not available for 4,242 (0.67%) episodes

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.

MSPB-PAC IRF score calculation removes outliers in the 1st and 99th percentiles of the residual distribution and subsequent testing and analyses were conducted on these 618,123 episodes.

There were a total of 1,161 IRFs with FY 2016-2017 episodes. Facility-level performance measure score reliability and validity testing was restricted to the 1,147 IRFs with 20 or more episodes in FY 2016-2017. We applied this restriction because this measure is only to be publicly reported for providers with a minimum of 20 stays in the two-year measure calculation period.

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 analyzed the impact of the following beneficiary-level and community-level social risk factors:

- Medicare/Medicaid dual eligibility,
- Race/ethnicity,

- Urbanicity, based on beneficiary ZIP code, and
- Socioeconomic status (SES), based on beneficiary ZIP code and using data from the 2017 American Community Survey (5-year file).

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)

We examined the measure score's ability to capture between-facility differences versus random error using a signal-to-noise reliability score, defined below. We also examined measure score repeatability by assessing agreement between a facility's MSPB scores based on randomly-split independent subsets of IRF episodes. Performance measure score reliability testing was restricted to the 1,147 Inpatient Rehabilitation Facilities (IRFs) with 20 or more episodes in the FY 2016-2017 measurement period, as only IRFs with a minimum of 20 episodes in the two-year measurement period will be publicly reported.

Reliability Score: Measure reliability scores reflect the extent to which variation in the measure is due to true, underlying provider performance rather than random variation (i.e., statistical noise) due to the sample of cases observed. In the case of MSPB-PAC, the reliability score captures how much of the variance in measure scores is due to differences in episode payments between agencies rather than differences in episode payments within a facility's set of episodes. This score is calculated for each facility, using the following formula:

Reliability Score: $R_k = V_b/(V_b + (V_{w_k}/n_k))$,

where R_k is the reliability for facility *k*, V_b is the between-facility variance, V_{w_k} is the within-facility variance for facility *k*, and n_k is the number of MSPB episodes for facility *k*.[2] We report the facility-level distribution of the reliability score.

Split-sample Reliability Testing: This test examined agreement between two performance measure scores for a facility based on randomly-split, independent subsets of IRF episodes. Good agreement indicates that the performance score is more the result of facility characteristics, like efficiency of care, rather than statistical noise due to random variation. We used four years of data (FY 2014-2017) to achieve numbers of

---

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

episodes per facility in the split-half samples that are comparable to the numbers used for the actual measure scores. The sample was stratified by fiscal year, thus ensuring that episodes within each fiscal year were evenly distributed across the split-halves. We calculated performance measure scores for each split-half sample using the same measure specification. We then calculated Shrout-Fleiss intraclass correlation coefficients ICC(2,1)[3] between the split-half scores to measure reliability.

We also calculated ICCs between split-half scores stratified by facility size to assess whether reliability was acceptable across providers of varying sample size. To do this, we first split our sample of 1,147 IRFs into quartiles based on facility size. We then calculated ICCs within each quartile using the split-half performance measure scores derived above. Lower ICC scores indicate less correlation between the two estimates, a score of 1 would mean the estimates are exactly the same.

The Reliability Score and the ICC capture related, but distinct, concepts. ICC(2,1) will tend to differ from the Reliability Score metric for two reasons: the denominator of ICC(2,1) (i) includes statistical variation arising from true differences in a provider's performance across performance periods; and (ii) imposes a common variance for the residuals across providers, ignoring differences in precision arising from differences in case sizes. Reason (i) makes ICC(2,1) a less relevant metric in this context, since program goals actually require accurately distinguishing systematic performance changes from one period to another, rather than treating them as statistical noise. To avoid this issue, one could alternatively calculate ICC(2,1) using split-half samples from a single performance period. However, this approach also underestimates reliability of the measure for use in the program; in this case, under-estimation occurs because case sizes are artificially cut in half from true case sizes, mechanically reducing precision from the intended application of the measures. For these reasons, we view the Reliability Score as the preferred and more relevant metric of reliability. We still present both reliability metrics for completeness.

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*)

Reliability Score Results. **Table 3** presents the mean, 25th, 50th, and 75th percentile values of the reliability scores among 1,147 IRFs from FY 2016-2017, and by sample size quartile. The average reliability score for all agencies was 0.86 and the median was 0.89. When examined by facility size, the average reliability score increased from 0.70 (quartile 1) to 0.96 (quartile 4).

Table 3. Facility Reliability Score Distribution of the Episode-Level MSPB Risk Adjusted Spending, overall IRF sample and by sample size quartile, with public reporting exclusions (k = 1,147)

Facility Sample	K	Mean (SD)	25th Pct*	Median	75th Pct
Overall	1,147	0.86 (0.12)	0.81	0.89	0.94
Quartile 1: 20-190 episodes	288	0.70 (0.12)	0.64	0.73	0.78
Quartile 2: 191-334 episodes	286	0.85 (0.04)	0.83	0.85	0.87
Quartile 3: 335-686 episodes	287	0.91 (0.02)	0.90	0.91	0.93
Quartile 4: 687-5,941 episodes	286	0.96 (0.01)	0.96	0.97	0.97

* Pct = percentile. Analysis of Medicare Claims File for IRF FY 2016-2017.

Note: Facility size can vary based on the sample in which the regression was estimated due to outlier exclusions.

[3] Shrout, Patrick E., and Joseph L. Fleiss. "Intraclass correlations: uses in assessing rater reliability." *Psychological bulletin* 86, no. 2 (1979): 420.

Split-sample Reliability Testing Results. **Table 4** presents ICC(2,1) between the split-sample scores for the overall sample of 1,147 IRFs included in this testing, and by sample size quartile. The ICC in the overall sample was 0.87 with a 95% confidence interval (CI) of 0.85 to 0.88. The ICC was lowest in the first sample size quartile and increased progressively with increasing quartile.

Table 4. Split-sample reliability: Intraclass correlation coefficients between split-sample performance measure scores for the overall IRF sample and by sample size quartile, with public reporting exclusions (N = 1,147)

Facility Sample	K	ICC(2,1) (95% CI)
Overall	1,147	0.87 (0.85-0.88)
Quartile 1: 20-190 episodes	287	0.81 (0.76-0.84)
Quartile 2: 191-334 episodes	287	0.87 (0.84-0.90)
Quartile 3: 335-686 episodes	286	0.93 (0.91-0.94)
Quartile 4: 687-5,992 episodes	287	0.95 (0.94-0.96)

Analysis of Medicare Claims File for IRF FY 2014-2017.

Note: Facility size can vary based on the sample in which the regression was estimated due to outlier exclusions.

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?*)

Overall, reliability testing results indicated good performance measure score reliability.[4] Reliability among the smallest publicly reported providers (quartile 1), is moderate to good, depending on the analysis used.

[4] Thresholds for sufficient measure reliability (including the ICC and other reliability methods) vary across sources (see, for example, Portney and Watkins, 2000, for a discussion). Nunnally (1978) is often cited to justify a threshold of 0.7 for "sufficient" reliability. Other authors provide other thresholds. For example, Landis and Koch (1977) classify Kappa statistics in the 0.41-0.60 range as "moderate," 0.61-0.80 range as "substantial," and 0.81-1.00 range as "almost perfect." Koo and Li (2016), on the other hand, classify ICC values in the 0.5-0.75 range as "moderate," 0.75-0.9 range as "good," and above 0.9 as "excellent." The Department of Education provides the following thresholds: "*Reliability of an outcome measure may be established by meeting the following minimum standards: (a) internal consistency (such as Cronbach's alpha) of 0.50 or higher; (b) temporal stability/test-retest reliability of 0.40 or higher; or (c) inter-rater reliability (such as percentage agreement, correlation, or kappa) of 0.50 or higher.*" (What Works Clearinghouse (WWC) Standards Handbook v4, p.78)

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, *15*(2), 155-163.

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

Nunnally, J. C. (1978). Psychometric theory. New York, NY: McGraw-Hill.

U.S. Department of Education (2018), What Works Clearinghouse (WWC) Standards Handbook version 4.0.

The reliability score results indicated that the average facility had good reliability. On average, 86 percent of the variation in the risk adjusted MSPB amount was associated with systematic differences between agencies[5], with a range of 70 to 96 percent (on average) among the smallest and largest facility quartiles, respectively.

The split-half reliability analysis provides further evidence of reliability and repeatability of the performance measure. Reliability (ICC) was good overall, at 0.87, with a range of 0.81 to 0.95 (on average) among the smallest and largest facility quartiles, respectively.

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 <u>performance measure score</u> as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*) **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)*

To empirically test Performance Measure Score validity, we used the following methods, we:

1. Evaluated the empirical validity of the MSPB-PAC measure by examining correlation with known indicators of resource or service utilization, specifically hospital admissions and emergency room (ER) visits during the episode period. For this analysis, we compared the ratio of observed over expected spending for MSPB-PAC IRF episodes with and without hospital admissions occurring in the episode period. We also compared the observed over expected spending for episodes with and without ER visits. This analysis sought to confirm the expectation that variation in service utilization is captured by the MSPB-PAC cost measure.
2. Examined the correlation between MSPB-PAC IRF scores and the Discharge to Community (DTC) rates for FY 2016-2017. The Discharge to Community-Post Acute Care Measure for Inpatient Rehabilitation Facilities (DTC-PAC IRF) is endorsed by NQF (#3479), is publicly reported as part of the IRF Quality Reporting Program, and is based on Medicare claims.[6]

---

[5] Thompson, M. P., Kaplan, C. M., Cao, Y., Bazzoli, G. J., & Waters, T. M. (2016). Reliability of 30-Day Readmission Measures Used in the Hospital Readmission Reduction Program. *Health services research*, *51*(6), 2095-2114.

[6] Medicare Program; Inpatient Rehabilitation Facility Prospective Payment System for Federal Fiscal Year 2017 Federal Register, Vol. 81, No. 151. https://www.gpo.gov/fdsys/pkg/FR-2016-08-05/pdf/2016-18196.pdf

The DTC measure assesses successful discharge to community from an IRF, with successful discharge to community including no unplanned hospitalizations and no death in the 31 days following discharge. Specifically, this measure reports an IRF's risk-standardized rate of Medicare fee-for-service (FFS) patients who are discharged to the community following an IRF stay, and do not have an unplanned admission to an acute care hospital or long-term care hospital (LTCH) in the 31 days following discharge to community, and who remain alive during the 31 days following discharge to community. DTC is calculated using two consecutive years of data.

We hypothesized that there would be a negative association between the MSPB measure and DTC measure, indicating that providers with lower MSPB scores (more efficient providers) would have higher rates of successful discharge to the community. Providers whose patients have adverse events, such as re-hospitalization, at a rate higher than would be expected based on patient characteristics, should have lower DTC scores and higher (less cost efficient) MSPB scores. The reverse should be true for providers whose patients have fewer than expected adverse events.

3. We examined the correlation between MSPB-PAC IRF scores and provider's scores on the Percent of Residents or Patients with Pressure Ulcers That Are New or Worsened (Short-Stay) measures (NQF #0678). Pressure ulcers are high-cost adverse events, and their prevalence is an important measure of quality of care.[7] The Pressure Ulcers measure was endorsed by NQF (the endorsement has been removed because the measure has been replaced by the Changes in Skin Integrity Post-Acute Care: Pressure Ulcer/Injury measure).[8] The measure reports the percent of Medicare patients with Stage 2-4 pressure ulcer(s) that are new or worsened since admission, is assessment-based, and is publicly reported on the IRF Compare website.

   The relationship between resource use and outcome measures depends on many factors, including the exact construction of each measure, payment policies, and the real-world relationship between service provision and both immediate and longer-term patient outcomes. For this analysis, the relationship depends in part on whether the prevalence of pressure ulcers is associated with other adverse outcomes, such as unplanned hospitalizations. To answer this question, we examined the relationship between pressure ulcers and the claims-based measure which captures adverse outcomes, as described earlier: DTC. We found that pressure ulcers are not associated with lower rates of discharge to community: Pearson correlation between DTC and Pressure Ulcers measures is -0.038 (Spearman: 0.021).[9] Accordingly, we hypothesized that the correlations between the Pressure Ulcers measure and MSPB-PAC IRF should, likewise, be low.

2b1.3. What were the statistical results from validity testing?

(*e.g., correlation; t-test*)

[7] Medicare Program; Inpatient Rehabilitation Facility Prospective Payment System for Federal Fiscal Year 2018. Federal Register, Vol. 82, No. 148. https://www.govinfo.gov/content/pkg/FR-2017-08-03/pdf/2017-16291.pdf

[8] https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/IRF-Quality-Reporting/IRF-Quality-Reporting-Program-Measures-Information-

[9] Pearson correlation p-value=0.220, Spearman correlation p-value=0.503.

We found a positive relationship between MSPB and known indicators of resource or service utilization. The mean observed to expected cost ratio for episodes without a hospital admission is 0.91, compared with 1.39 for episodes with at least one hospital admission during the episode period (p-value<0.0001). The mean observed to expected cost ratio for episodes without an ER visit is 0.98, compared to 1.09 for episodes with at least one ER visits (p-value<0.0001). We also observe a positive relationship between the mean observed to expected cost ratio and the number of hospitalizations/ER visits (**Table 5**).

Table 5. Mean Cost Ratio, by Number of Hospitalizations/ER Visits

Number of High-Cost Event	0	1	2	3	4
Hospitalizations	0.91	1.35	1.59	1.73	1.81
ER Visits	0.98	1.08	1.14	1.15	1.17

Analysis of Medicare Claims File for IRF FY 2016-2017.

We also found a small, significant negative association between MSPB measure scores and the DTC measure scores (**Table 6**). Both Pearson and Spearman rank correlations revealed similar relationships.

Table 6. Correlations between MSPB and Discharge to Community (DTC) Measures

Measure Name	K*	Pearson Correlation	p-value	Spearman Correlation	p-value
Discharge to Community (DTC)	1,067	-0.193	<0.0001	-0.165	<0.0001

Analysis of Medicare Claims File for IRF FY 2016-2017.

*Number of reflects providers with both MSPB and DTC measure scores.

Lastly, we found very small, statistically significant correlation (both Pearson and Spearman) between MSPB measure scores and Pressure Ulcers measure scores (**Table 7**). These results are consistent with the finding that pressure ulcers measure scores are not associated with lower rates of successful discharge to community; they may indicate that prevalence of pressure ulcers is associated with somewhat higher spending in the 30 days after discharge from IRF, on average.

Table 7. Correlations between MSPB and Pressure Ulcers Measures

Measure Name	K*	Pearson Correlation	p-value	Spearman Correlation	p-value
Rate of pressure ulcers that are new or worsened	1,061	0.1207	0.0001	0.1198	0.0001

Analysis of Medicare Claims File for IRF FY 2016-2017 and IRF Compare data FY 2017.

*Number of reflects providers with both MSPB and Pressure Ulcers measure scores.

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 positive relationship between MSPB and known indicators of resource/service utilization confirms that the MSPB measure is sensitive to both the occurrence and the intensity of high cost events. The small, significant negative correlation between MSPB and DTC measures confirms that, on average, more efficient

IRFs are associated with better discharge to community rates and fewer unplanned hospitalizations. The low correlation between MSPB and Pressure Ulcers measures is consistent with the finding that the prevalence of pressure ulcers is not associated with lower rates of successful discharge to community.

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*)

Measure exclusion criteria and rationale for exclusions are presented in section 2b2.3. We examined the episode-level frequency of each exclusion and the facility-level distribution of exclusions. The exclusions were required to ensure availability of complete and valid data for measure specification (e.g., excluding episodes in which a patient is not enrolled in Medicare FFS or where Medicare was not the primary payer for the entirety of a 90-day lookback period plus episode window).

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 8 displays the overall number and percentage of episodes excluded based on each criterion. Because the exclusions are not applied sequentially, one episode could be excluded for multiple reasons and the sum of individual exclusion frequencies may exceed the total number of episodes excluded. Overall, 18.48% of episodes were excluded because of one or more exclusion criteria. 10.01% of episodes were excluded due to a patient not being enrolled in Medicare FFS for the entirety of the 90-day lookback period plus the episode window. 10.21% of episodes were excluded because a patient had a primary payer other than Medicare for any part of the 90-day lookback period plus the episode window.

Table 9 shows the distribution of exclusions at the facility level. On average, 10.52% of a facility's episodes were excluded due to a patient not being enrolled in Medicare FFS for the entirety of the 90 day lookback period plus the episode window and 10.18% of a facility's episodes were excluded because a patient had a primary payer other than Medicare for any part of the 90-day lookback period plus the episode window.

Table 8. Episode Frequencies of Exclusion Criteria for the MSPB-PAC IRF Measure

Exclusion	N*	%
Any episode that is triggered by an IRF claim outside the 50 states, D.C., Puerto Rico, or U.S. territories	0	0.00%
Any episode where the claim(s) constituting the attributed IRF provider's treatment have a standard allowed amount of zero or where the standard allowed amount cannot be calculated	0	0.00%
Any episode in which a beneficiary is not enrolled in Medicare FFS for the entirety of the 90-day lookback period (i.e., a 90-day period prior to the episode trigger) plus episode window (including where a beneficiary dies),	77,442	10.01%

or is enrolled in Part C for any part of the lookback period plus episode window		
Any episode in which a beneficiary has a primary payer other than Medicare for any part of the 90-day lookback period plus episode window	78,981	10.21%
Any episode where the claim(s) constituting the attributed IRF's treatment include at least one related condition code indicating that it is not a prospective payment system bill	0	0.00%
Any episode with problematic claims data (e.g., anomalous records for stays that overlap wholly or in part, or are otherwise erroneous or contradictory)	18	0.00%
Total Number of Episodes Excluded	143,001	18.48%

Analysis of Medicare Claims File for IRF, FY 2016-2017

*Exclusions are not mutually exclusive; one episode could be excluded for multiple reasons. The sum of individual exclusion frequencies may exceed the total number of episodes excluded.

Table 9. **Facility-Level Distribution of Exclusion Criteria for the MSPB-PAC IRF Measure**

Exclusion	Mean	25th Perc.	Median	75th Perc.
Any episode that is triggered by an IRF claim outside the 50 states, D.C., Puerto Rico, or U.S. territories	0.00%	0.00%	0.00%	0.00%
Any episode where the claim(s) constituting the attributed IRF provider's treatment have a standard allowed amount of zero or where the standard allowed amount cannot be calculated	0.00%	0.00%	0.00%	0.00%
Any episode in which a beneficiary is not enrolled in Medicare FFS for the entirety of the 90-day lookback period (i.e., a 90-day period prior to the episode trigger) plus episode window (including where a beneficiary dies), or is enrolled in Part C for any part of the lookback period plus episode window	10.52%	7.58%	9.29%	11.46%
Any episode in which a beneficiary has a primary payer other than Medicare for any part of the 90-day lookback period plus episode window	10.18%	7.95%	9.77%	12.13%
Any episode where the claim(s) constituting the attributed IRF provider's treatment include at least one related condition code indicating that it is not a prospective payment system bill	0.00%	0.00%	0.00%	0.00%
Any episode with problematic claims data (e.g., anomalous records for stays that overlap wholly or in part, or are otherwise erroneous or contradictory)	0.00%	0.00%	0.00%	0.00%

Analysis of Medicare Claims File for IRF, FY 2016-2017

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*)

Exclusions for the MSPB-PAC measure are listed in **Table 10**, along with the rationale for each exclusion. Enrollment in Medicare FFS with Medicare as the primary payer is necessary to calculate the Medicare spending measures. Without these exclusion criteria, it would not be possible to calculate the full cost to the Medicare program. The risk adjustment methodology relies on Medicare FFS claims during the 90-day look back period and, therefore, the model will not be accurate without full Medicare claims data for the period of analysis.

Table 10. **Exclusion Criteria and Rationale for the MSPB-PAC IRF Measure**

Exclusion	Rationale
Any episode that is triggered by an IRF claim outside the 50 states, D.C., Puerto Rico, or U.S. territories.	This exclusion ensures that complete claims data are available for each provider.
Any episode where the claim(s) constituting the attributed IRF's treatment have a standard allowed amount of zero or where the standard allowed amount cannot be calculated.	Episodes where the claim(s) constituting the attributed PAC provider's treatment are zero or have unknown allowed payment do not reflect the cost to Medicare. Including these episodes in the calculation of MSPB-PAC measures could potentially misrepresent a providers' resource use.
Any episode in which a beneficiary is not enrolled in Medicare FFS for the entirety of the 90-day lookback period (i.e., a 90-day period prior to the episode trigger) plus episode window (including where a beneficiary dies), or is enrolled in Part C for any part of the lookback period plus episode window.	Episodes meeting this criteria do not have complete claims information that is needed for risk adjustment and the measure calculation, as there may be other claims (e.g., for services provided under Medicare Advantage [Part C]) that are not observable in the Medicare Part A and B claims data. Including these episodes in the MSPB-PAC measures could potentially misrepresent a provider's resource use. This exclusion also allows us to faithfully construct Hierarchical Condition Categories (HCCs) for each episode by scanning the lookback period prior to its start without missing claims.
Any episode in which a beneficiary has a primary payer other than Medicare for any part of the 90-day lookback period plus episode window.	Where a patient has a primary payer other than Medicare, complete claims data may not be observable. These episodes are removed to ensure that the measures are accurately calculated using complete data.
Any episode where the claim(s) constituting the attributed IRF's treatment include at least one related condition code indicating that it is not a prospective payment system bill.	Claims that are not a prospective payment system bill may not report sufficient information to allow for payment standardization.
Any episode with problematic claims data (e.g., anomalous records for stays that	The episode with the most recent processing date is kept to ensure the accuracy of data elements.

Exclusion	Rationale
overlap wholly or in part, or are otherwise erroneous or contradictory).	

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** 146 **risk factors**

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

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

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.

The MSPB-PAC IRF risk adjustment models are adapted from the model used in the NQF-endorsed hospital MSPB measure (#2158), which is itself an adaptation of the standard CMS-HCC risk adjustment model.[10,11] The MSPB-PAC IRF model uses a linear regression framework and a 90-day HCC lookback period. The following beneficiary health status indicators are included as covariates in the MSPB-PAC IRF risk adjustment model and to the greatest extent possible are consistent across PAC settings (see Appendix C of the Measure Specifications[12] for a comprehensive list of independent variables used in the risk adjustment models):

- 70 HCCs
- 11 HCC interactions
- 11 brackets for age at the start of the episode

[10] QualityNet, "Measure Methodology Reports: Medicare Spending Per Beneficiary (MSPB) Measure," (2015). http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1228772057350

[11] CMS, "Medicare Risk Adjustment Information" (2016) https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Risk-Adjustors.html

[12] https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/Downloads/2016_07_20_mspb_pac_ltch_irf_snf_measure_specs.pdf

- Original entitlement to Medicare through disability
- ESRD status
- Long-term care institutionalization at start of episode[13]
- 6 clinical case mix categories reflecting recent prior care (described further below)[14]
- Hospice utilization during the episode
- Prior acute ICU utilization day categories
- Prior acute length of stay categories
- Rehabilitation Impairment Categories (RICs)

Clinical case mix categories are also included in the MSPB-PAC IRF risk adjustment models to account for differences in the intensity and type of care received by beneficiaries prior to the start of an MSPB-PAC IRF episode. A beneficiary is assigned to a clinical case mix category using the following methodology. Taking the most recent institutional claim (by end date) in the 60 days prior to the start of an MSPB-PAC IRF episode, the episode is assigned to one of the following mutually exclusive and exhaustive clinical case mix categories:

(1) **Prior Acute Surgical IP – Orthopedic** – beneficiaries who have most recently undergone orthopedic surgery in an acute inpatient hospital

(2) **Prior Acute Surgical IP – Non-Orthopedic** – beneficiaries who have most recently undergone a non-orthopedic surgery in an acute inpatient hospital

(3) **Prior Acute Medical IP with ICU** – beneficiaries who have most recently stayed in an acute inpatient hospital for non-surgical reasons and had a stay in the ICU

(4) **Prior Acute Medical IP without ICU** – beneficiaries who have most recently stayed in an acute inpatient hospital for non-surgical reasons but did not have a stay in the ICU

(5) **Prior PAC - Institutional** – beneficiaries who are continuing PAC from an institutional PAC setting (i.e., coming from an LTCH, IRF, or SNF)

(6) **Prior PAC - HHA** – beneficiaries who are continuing PAC from an HHA

(7) **Community** – all other beneficiaries

In the event that there are multiple prior claims with the same end date in the 60 days prior to the start of a PAC episode, additional logic is employed to determine the episodes' clinical case mix category. For conflicts occurring between two IP claims, the clinical case mix category corresponding to the claim with the longest length of stay (LOS) is assigned. For all other types of conflicts including those where the LOS is the same between two IP claims, the clinical case mix category is assigned using a hierarchy in the order of the categories listed above.

---

[13] Identifies beneficiaries who have been institutionalized for at least 90 days in a given year. The indicator is based on 90-day assessments from the Minimum Data Set (MDS) and is calculated based on CMS' definition of institutionalized individuals.

[14] There are 7 case mix categories, as described below, but one category is removed to prevent collinearity.

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.

Not applicable

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?

Clinical Factors

The CMS-HCC model was selected based on previous studies evaluating its appropriateness for use in risk-adjusting Medicare claims data. This model was developed specifically for use in the Medicare population, meaning that it accounts for conditions found in the Medicare population and is calibrated on Medicare Fee-for-Service (FFS) beneficiaries. In addition, the CMS-HCC model is annually updated for changes in coding practices and is exhaustive on these code sets. Because the CMS-HCC model has already been extensively tested, we focus on adapting the CMS-HCC model to the IRF measure methodology.[15]

Extensive clinical review was performed by clinicians with experience providing care in IRF settings, in collaboration with Medical Officers at CMS. The hospitalizations and outpatient services felt least clinically related to the IRF care were excluded from resource use calculation. Services were only added to the exclusions list if there was consensus across IRF and CMS clinicians.

The MSPB-PAC IRF measure accounts for comorbid interactions by incorporating a number of health status interactions, such as disability and selected HCC groups (e.g., Cystic Fibrosis, Severe Hematological Disorders, Opportunistic Infections, among others). Given the fact that beneficiaries often have more than one comorbidity, the model includes commonly observed paired condition interactions, (e.g., chronic obstructive pulmonary disease [COPD] and congestive heart failure [CHF]) and commonly observed triple-interactions (e.g., diabetes mellitus, congestive heart failure, and renal failure). The beneficiary health status indicators included in the model are described in **Section 2b3.1.1**.

Social Risk Factors

A number of studies have shown that socioeconomic status is associated with the amount of resources used during the period in which patients are hospitalized, as well as during post-acute care. For example, individuals on Medicaid or who are uninsured, individuals with lower income, and individuals living in metropolitan areas are more likely to use SNF versus IRF care.[16] End-of-life care for Medicare beneficiaries who are Black or Hispanic is also substantially different than the end-of-life hospital services received by Medicare beneficiaries who are White. Much of the variation in end-of-life care is due to differences in utilization levels among hospitalized patients. Beneficiaries who are Black and beneficiaries who are Hispanic are significantly more likely to be admitted to the ICU than beneficiaries who are White, and minorities also

[15] Pope, Gregory C., John Kautter, Melvin J. Ingber, Sara Freeman, Rishi Sekar, and Cordon Newhart. "Evaluation of the CMS-HCC Risk-Adjustment Model: Final Report." RTI International: March 2011.

[16] Kaiser Family Foundation. "Medicare Chartbook" Fourth Edition, 2010. http://www.kff.org/medicare/upload/8103.pdf

receive significantly more intensive procedures, such as resuscitation and cardiac conversion, mechanical ventilation, and gastrostomy for artificial nutrition.[17]

According to a 2014 National Quality Forum report, the mechanisms underlying differences in resource use by socioeconomic status and race are complex and may be impacted by factors such as financial resources, community resources, historical and current discrimination, and reduced access to preventive services. Provider assumptions or implicit biases may impact quality of care for beneficiaries of different races.[18] These factors may result in inefficient care, increased disease severity, or greater morbidity, leading to higher Medicare spending for beneficiaries depending on socioeconomic status or race.

Given the conceptual and empirical relationship between income, race, and resource use, we analyzed the impact of the following beneficiary-level and county-level social risk factors: dual eligibility, race/ethnicity, urbanicity based on beneficiary residence, and socioeconomic status (SES). We used the CMS Enrollment Database (EDB) and Common Medicare Environment (CME) to determine dual eligibility, race, and beneficiary ZIP code. Urbanicity was defined by cross-walking beneficiary residence ZIP codes to Federal Information Processing Standard Publication (FIPS) codes,[19] then cross-walking FIPS codes to Rural-Urban Continuum Codes (RUCC_2013).[20] Socioeconomic status was determined using the Agency of Healthcare Research and Quality (AHRQ) SES Index[21], calculated based on beneficiary residence ZIP Code Tabulation Area (ZCTA). ZCTA was found by cross-walking the beneficiary residence ZIP code with ZCTA. We used data from the 2017 American Community Survey (5-year file) to calculate the AHRQ SES Index, with higher values indicating higher SES.

Using these data, we conducted a number of analyses for each social risk factor:

- Calculated the frequency of patients with each social risk factor;
- Calculated average Medicare spending for patients with each social risk factor;
- Assessed the difference in the measure scores estimated with and without adjustment for the social risk factors

The outcomes of these analyses are discussed in **Section 2b3.4b**.

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)**

---

[17] Freburger, Janet K., George M. Holmes, Li-Jung E. Ku, Malcolm P. Cutchin, Kendra Heatwole-Shank, and Lloyd J. Edwards. "Disparities in post–acute rehabilitation care for joint replacement." *Arthritis care & research* 63, no. 7 (2011): 1020-1030.

[18] National Quality Forum. "Risk Adjustment for Socioeconomic Status or Other Sociodemographic Factors: Technical Report." National Quality Forum: August 2014.

[19] https://www.huduser.gov/portal/datasets/usps_crosswalk.html

[20] https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/

[21] Bonito, A. J., Bann, C., Eicheldinger, C., & Carpenter, L. (2008). Creation of new race-ethnicity codes and socioeconomic status (SES) indicators for Medicare beneficiaries. *AHRQ Publication*, (08-0029). Available at: https://archive.ahrq.gov/research/findings/final-reports/medicareindicators/medicareindicators.pdf

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

The MSPB-PAC IRF measure broadly replicates the CMS-HCC model. The literature has extensively tested the use of the HCC model as applied to Medicare claims data. It was also adopted for the NQF-endorsed MSPB-Hospital measure (#2158). Although the variables in the HCC model were chosen to predict annual cost, CMS also uses this risk adjustment model in a number of other settings (e.g., ACOs and physician QRUR programs). More information on selection of factors included in the CMS-HCC model can be found at Pope et al. (2011). During the development phase of MSPB-PAC IRF, we also conducted additional analyses. We tested stratifying by clinical case mix categories (e.g., Prior Acute Surgical IP – Orthopedic). We found that the improvement in model fit was small, while the approach also created small sample sizes in some strata. Further, we tested a longer look-back period (180 days) for identifying patients' comorbidities. This approach resulted in lower model fit and increased the number of excluded episodes. Lastly, we used bootstrapping techniques to investigate the significance of individual risk factors. We found that the risk factors that were not consistently significant across replications were clinically analogous to the risk factors that were consistently significant. As a results, we did not remove such risk factors from the model.

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.**

First, we examined the frequency of patients with each social risk factor (SRF). Overall, 19% of IRF episodes in the FY 2016-2017 period involved patients who were dual-eligible, 16% involved patients who were not White, and 5% involved patients in rural locations (see **Appendix Table 2b3.4b_1**).

Second, we compared average observed per-episode spending across patients with each SRF. Overall, we found that spending is higher for patients who are dual-eligible patients, patients who are Black, and patients who live in urban areas. Spending is highest among patients who reside in areas with average SES in the first quartile of the AHRQ SES Index (see **Appendix Table 2b3.4b_2**).

Third, we examined risk-adjustment models with all or some of the SRFs added (see **Appendix Tables 2b3.4b_3a-e**). We found that each SRF is statistically significant, when added to the model individually as well as when added together with all other SRFs;[22] this is expected given the large sample size. However, adding SRFs, individually or together, does not substantially improve overall model fit: adjusted R-squared values increase by less than 0.001 in all cases.

Fourth, to further examine the impact of adding social risk factors to the risk adjustment model, we examined the change in individual provider's MSPB-PAC IRF measure scores when computed with and without social risk factors. The results are highly correlated: correlation between baseline scores and scores adjusted for all SRFs is 0.997 (Pearson and Spearman). (See **Appendix Figure2b3.4b_1 and Figure2b3.4b_2**). The average absolute change in provider scores is 0.004 (**Appendix Table 2b3.4b_4a**); 91% of IRF scores change by ±0.01, i.e., about one tenth of a standard deviation of the measure scores (**Appendix Tables 2b3.4b_4b**).

Given the minimal impact of including SRFs in the risk-adjustment model on measure scores, we do not recommend adjusting the scores for these social risk factors.

[22] Some coefficients do not have the same sign as the absolute difference in spending between the reference group and the controlled group. For example, while spending is higher among patients who are Hispanic than among those who are categorized as White, the coefficient on the Hispanic indicator variable is negative. This is due to collinearity between the social risk factors and the clinical factors already included in the model.

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*)

To test the adequacy of this model, we conducted several analyses and tests:

1. **Model Discrimination**: We examined the model's fit (adjusted R-squared).
2. **Risk-decile testing and plots:** We calculated the distribution of episode spending by decile to examine the model's ability to predict both very low and high cost episodes. Specifically, we created a "risk score" for each episode calculated as the predicted cost values from each episode divided by the national average predicted cost value. After arranging episodes into deciles based on the risk score, we calculated the difference and ratio between predicted and observed cost for each decile.
3. **Score Distribution**: We examined the distribution of IRF-level observed and risk-adjusted episode cost.

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

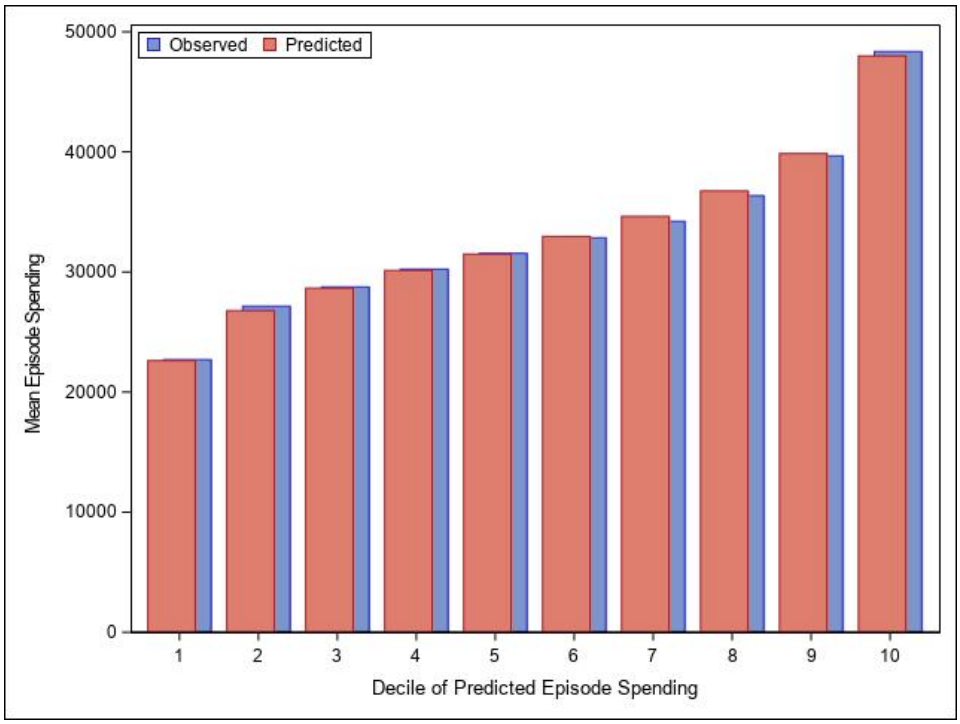
2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*):

The adjusted R-squared is 0.1595. Regression coefficients, standard errors, and p-values for each of the covariates used in the risk adjustment model are shown in **Appendix Table 2b3.6_1**.

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

The episode-level predicted costs range from $26,376 to $41,876, indicating that this model has a range of predictions and can predict both high and low costs. The results of our analysis comparing the observed and predicted costs by risk decile are displayed in **Figure 1** and **Table 11.** In each risk decile, the observed and predicted costs are close, with a difference of 1 percentage points or less. The ratio of observed to predicted rates is close to 1 across risk deciles, with the smallest ratio being 0.99 and the largest ratio being 1.01.

Figure 1. IRF Model Diagnostics: Comparison of Observed and Predicted Spending by Predicted Spending Deciles



Analysis of Medicare Claims File for IRF FY 2016-2017.

Table 11. IRF Model Diagnostics: Comparison of Observed and Predicted Spending by Predicted Spending Deciles

Deciles of predicted episode cost	Number of episodes	Observed episode cost	Predicted episode cost	Predicted minus observed cost	Observed/ predicted costs
1	61,800	22,702	22,616	-85.61	1.00
2	61,799	27,152	26,783	-368.48	1.01
3	61,799	28,757	28,652	-104.68	1.00
4	61,801	30,242	30,131	-111.18	1.00
5	61,798	31,553	31,490	-63.53	1.00
6	61,799	32,851	32,961	110.31	1.00
7	61,800	34,219	34,629	410.17	0.99
8	61,799	36,357	36,744	386.35	0.99
9	61,799	39,667	39,860	193.02	1.00
10	61,799	48,355	47,989	-366.21	1.01

Analysis of Medicare Claims File for IRF FY 2016-2017.

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

The results of our analysis comparing the observed and predicted spending by deciles of spending risk are displayed in **section 2b3.7** above.

2b3.9. Results of Risk Stratification Analysis:

Not applicable

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)?

(i.*e., what do the results mean and what are the norms for the test conducted*)

The model discrimination and calibration results demonstrate good predictive ability across the full range of episodes, from low to high spending risk. There was no evidence of excessive under- or over-estimation at the extremes of episode risk. The overall adjusted R-squared is 0.1595. The model controls for over 100 comorbidities (including comorbid interactions), case-mix categories, and patient risk factors. Extensive clinical review was performed by clinicians with experience providing care in IRF settings, in collaboration with Medical Officers at CMS, to identify and review relevant risk factors. Furthermore, certain features of the model improve its policy and practical usability while potentially reducing its fit statistics (adjusted R-squared). Most importantly, unrelated services, such as planned hospital admissions and routine management of certain preexisting chronic conditions (see **section S.9.1 of the Intent to Submit form**), were purposefully and carefully excluded to improve the ability to interpret and compare MSPB-PAC IRF scores across providers. The R-squared cannot be evaluated alone and must be considered in combination with the costs excluded from the measure to ensure clinical validity. Since unrelated services may be well predicted by patient risk factors, excluding them can reduce the explained portion of the cost variance and the model's adjusted R-squared. For example, MSPB-PAC IRF excluded services such as routine dialysis for end-stage renal disease (ESRD), as they were not felt to be prescribed by or within the scope of the IRF providers. If these services had been included in the IRF measure, doing so would have increased the R-squared because the ESRD indicator variable in the risk adjustment model would explain much of the variation due to dialysis. This, however, would have created an inferior measure, as it would lack clinical validity.

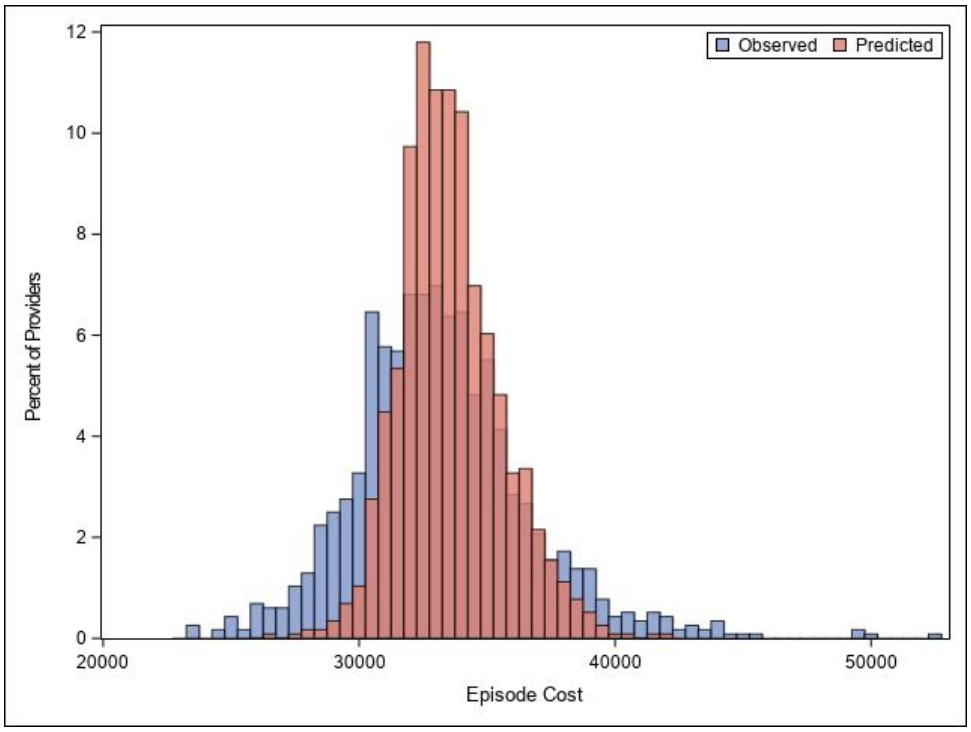
The distribution of facility-level observed and risk-adjusted spending is shown in **Table 12** and **Figure 2.** By taking into account beneficiary characteristics that are outside the provider's control, the model compresses the distribution of provider-level spending and decreases their variability. The degree of compression demonstrates that there exists a significant amount of variation in IRF spending that is not explained by the observed beneficiary risk factors.

Table 12. Distribution of Provider-Level Observed and Risk-Adjusted Episode Spending

Group	K	Mean	SD	10th Pct	25th Pct	50th Pct	75th Pct	90th Pct
Observed	1,161	33,185.0	3,454.9	29,256.2	31,022.0	32,936.3	34,931.9	37,389.5
Predicted	1,161	33,562.4	1,959.6	31,305.5	32,253.9	33,345.3	34,687.3	36,272.9

Analysis of Medicare Claims File for IRF FY 2016-2017.

Figure 2. Distribution of Provider-Level Observed and Risk-Adjusted Episode Spending



Analysis of Medicare Claims File for IRF FY 2016-2017.

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*)

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)*

We used the following two methods to identify statistically significant and meaningful differences in MSPB-PAC IRF measure scores:

1. We analyzed the distribution of measure scores across all IRFs that meet the public reporting criteria. Specifically, we looked at the standard deviation of the scores and multiple points of their distribution: min, max, and 10th, 25th, 75th, and 90th percentiles.
2. We used bootstrapping procedures to derive a confidence interval to determine if an IRF's score is significantly lower than, significantly higher than, or no different than the national average. We

bootstrapped samples of IRFs, with replacement, with 100 replications, and re-estimated the facility average episode ratio (observed spending over expect spending) within each replication. The calculation of the expected episode spending within each bootstrap includes the addition of normally distributed noise proportional to the standard error of the facility-specific average episode spending. We use the 2.5 and 97.5 percentile of the average episode ratio from the bootstrapped distribution of each facility to calculate the full width of the 95 percent confidence interval (CI). We then compared each facility's 95% CI to the national episode-level average ratio to determine if the provider's performance was significantly different from the national mean. IRFs whose 95% CI was entirely below the national average were considered to be significantly lower than the national average; IRFs whose CI was entirely above the national average were significantly higher than the national average; and IRFs whose CI overlapped the national average were no different than the national average.

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*)

As can be seen in **Table 13** and **Figure 3**, MSPB-PAC IRF scores are distributed fairly symmetrically and have a good deal of variability. The standard deviation is 0.08, and the max/min, 90/10, and 75/25 ratios are 1.98, 1.21, and 1.10, respectively.

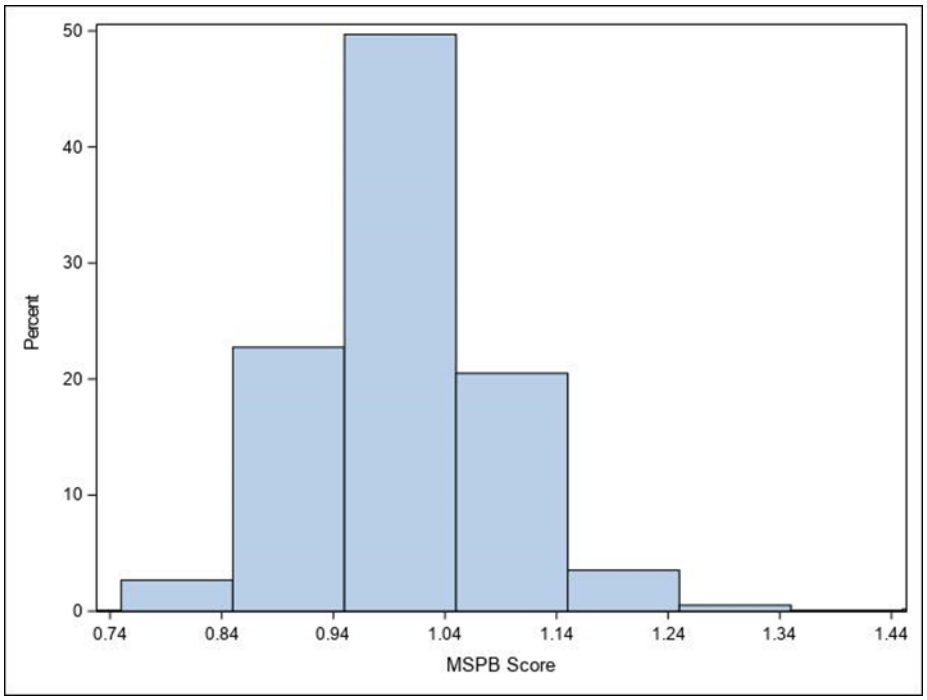
Due to the high level of reliability of the MSPB-PAC IRF scores, demonstrated in **section 2a2**, small differences in scores can be interpreted as meaningful. This is confirmed by our analysis of statistical significance: 46% of IRFs had scores that were statistically significantly higher than the national mean, while 30% of IRFs had scores that were statistically significantly lower (**Table 14**).

Table 13. Distribution of MSPB-PAC IRF Scores

K	Mean	Standard Deviation	Min	10th Pct	25th Pct	75th Pct	90th Pct	Max
1,147	1.00	0.08	0.74	0.91	0.95	1.05	1.10	1.47

Analysis of Medicare Claims File for IRF FY 2016-2017.

Figure 3. Distribution of MSPB-PAC IRF Scores



Analysis of Medicare Claims File for IRF FY 2016-2017.

Table 14. Proportion of IRFs with Scores Statistically Significantly Different From the National Average

IRF Total	Statistically significantly lower than national mean		Not statistically significantly different from national mean		Statistically significantly higher than national mean	
K	K	%	K	%	K	%
1,147	341	29.7	280	24.4	526	45.9

Analysis of Medicare Claims File for IRF FY 2016-2017.

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 MSPB-PAC IRF measure is able to identify statistically significant and meaningful differences in performance across IRFs due to its good reliability and variability. Measure scores range from 0.74 to 1.47, indicating that the model can predict both low and high spending and that there are meaningful differences in facility-level spending. Seventy six percent of IRFs have scores that are statistically significantly different from the national average, supporting the conclusion that even small difference between facility scores can be treated as meaningful.

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*)

Not applicable

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*)

Not applicable

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*)

Not applicable

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased

due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

This measure is calculated using Medicare FFS claims data; because submission and completion of claims is tied to provider reimbursement, missing data are rare. Our measure excludes episodes that are missing key measure specification data, under the exclusion criterion of claims with data that are problematic.

0.002% of episodes were excluded from our measure due to problematic claims data (e.g., anomalous records for stays that overlap wholly or in part, or are otherwise erroneous or contradictory). Thus, missing data are rare and do not have an impact on the measure. Consequently, we do not perform any formal missing data analyses.

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*)

As described in **section 2b6.1** above, missing data are rare and do not have an impact on the measure. Consequently, we do not perform any formal missing data analyses.

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased

due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i*.e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Not applicable

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 a combination of electronic sources

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.

This measure uses Medicare Enrollment data and Medicare FFS claims from the home health, inpatient, outpatient, and physician office settings claims data, which are routinely collected for payment purposes. These data are electronically available from the Centers for Medicare & Medicaid Services (CMS) at no cost beyond that of data processing and can be used to specify, publicly report, and track the measure in a timely fashion. Since data are already collected as part of Medicare's payment process, this measure poses no additional data collection burden on providers, and because claims are used for payment, data are complete and subject to audit. In addition, this measure uses data from the Minimum Data Set (MDS). The MDS is necessary to construct one of the risk adjustment variables, indicating beneficiaries who have been institutionalized for at least 90 days in a given year. The submission of MDS is part of the federally mandated process for clinical assessment of all residents in Medicare and Medicaid certified nursing homes and does not pose additional burden on providers.

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)?

None

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)
Quality Improvement (Internal to the specific organization)	Public Reporting Inpatient Rehabilitation Facilities Quality Reporting Program https://www.medicare.gov/inpatientrehabilitationfacilitycompare/

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

Name of program and sponsor:

This measure is publicly reported as part of the Center of Medicare & Medicaid Services' Inpatient Rehabilitation Facilities Quality Reporting Program.
Purpose:
Section 1895(b)(3)(B)(v)(II) of the Social Security Act (SSA) requires the Secretary to establish quality reporting requirements for IRFs. More information about the IRF QRP can be found at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/IRF-Quality-Reporting. In addition to tracking quality of care, quality measure data are intended to help consumers make informed decisions when selecting healthcare providers. Most quality measure data, including MSPB-PAC IRF scores, from the IRF QRP are publicly reported on the IRF Compare website at https://www.medicare.gov/inpatientrehabilitationfacilitycompare/. IRF quality measure data are also available for download for providers, researchers, and other public at https://data.medicare.gov/data/inpatient-rehabilitation-facility-compare.
Geographic area and number and percentage of accountable entities and patients included:
The IRF QRP includes all IRFs paid under the IRF PPS. MSPB-PAC IRF scores are publicly reported for active providers with 20 or more eligible episodes in the reporting period; thus, the number of providers included in the measures can vary by reporting period. The MSPB-PAC IRF measure results presented in this submission are based on 1,161 IRFs and 630,737 patient episodes; of these, 1,147 IRFs had 20 or more eligible episodes.

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?*)

Not applicable – measure is 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.*)

Not applicable – measure is 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.

Confidential feedback reports on the MSPB-PAC IRF measure were provided to all active IRF providers under the IRF QRP starting in Fall 2017. These on-demand, user requested, reports are available via the internet Quality Improvement and Evaluation System (iQIES) application. Public reporting of the MSPB-PAC IRF measure began in Fall 2018. Providers have a 30-day preview period to check their provider preview reports and submit suppression requests if there is evidence of errors in their data. CMS maintains an active provider helpdesk to which providers can submit any questions about the measure, including questions about performance data and interpretation. Individual responses are provided to each question. In addition, CMS conducts open door forums during which stakeholders can ask general questions about the measure. Along with the publicly-reported data, CMS includes consumer-friendly language to help consumers interpret measure data. Finally, MSPB-PAC IRF measure specifications were publicly posted along with the FY 2017 IRF PPS final rule at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/Downloads/2016_07_20_mspb_pac_ltch_irf_snf_measure_specs.pdf and https://www.govinfo.gov/content/pkg/FR-2016-08-05/pdf/2016-18196.pdf, respectively. The measure specifications are detailed and precise, allowing stakeholders to replicate measure calculations if they would like.

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.

See U.2.1.1.

Confidential feedback reports include the following data, for the provider and for the national average: reporting period, number of eligible episodes, spending during treatment period, spending during associated services period, total spending during episode, average risk-adjusted spending, national median risk-adjusted spending, and the MSPB-PAC IRF score.

Publicly available data and provider preview reports include the following: reporting period, number of eligible episodes, provider MSPB-PAC IRF score, and the national average MSPB-PAC IRF score.

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.

In addition to the processes described above, we solicited public comments on the MSPB-PAC IRF measure via a 24-day public comment period during January – February 2016. We posted the call for public comment on a CMS website and reached out via CMS listserv and notified TEP members. We received 45 comments during this period.

We also sought feedback on the measure through the pre-rulemaking process. We received four public comments after the release of the Measures Under Consideration (MUC) List on December 1, 2015. The MAP PAC/LTC Workgroup met on December 14-15 to consider this measure, and provided the preliminary decision of "encourage continued development" for the MSPB-PAC IRF measure. Following the release of the MAP PAC Workgroup's preliminary recommendation, the report was open for a public comment period. Eight public comments on this measure were received in this time. The MAP Coordinating Committee considered these comments alongside the Workgroup recommendation and finalized the recommendation of "encourage continued development," releasing their final recommendations in February 2016. Members of the public could comment during both MAP meetings.

The measure was subject to public comment during the FY 2017 IRF QRP rulemaking process. Stakeholders could comment on the specifications that were posted with the rule.

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

Comments were received from a range of stakeholders, including providers and provider associations, at each of these public comment opportunities. The comments covered a range of topics, including episode construction, exclusions, score calculation, risk adjustment, and reporting. Several commenters expressed support for the approach taken on these topics. Several commenters commented on issues such as: usefulness of setting-specific MSPB-PAC measures, usefulness of a resource use measure as a measure of quality, the adequacy of the risk adjustment model, and the process of sharing measure scores with providers. All comments were addressed, either by revising the measure or by providing the rationale why revisions are not necessary or appropriate.

The public comment summary report can be found at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/2016_03_24_mspb_pac_public_comment_summary_report.pdf.

The supplementary materials for public comment summary report can be found at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/2016_03_24_mspb_pac_public_comment_summary_report_supplementary_materials.pdf

The MAP recommendations and summaries of public comments can be found at http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=81593

The FY 2017 IRF PPS final rule with public comments and responses can be found at https://www.govinfo.gov/content/pkg/FR-2016-08-05/pdf/2016-18196.pdf.

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

Comments were received from a range of stakeholders, including researchers, government agencies, information system vendors, advocacy groups, and individuals at each of these public comment opportunities.

See U.2.2.2 for details and links to public comment summaries.

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

CMS and Acumen, LLC., reviewed and considered all public comments during development and implementation, before finalizing the measure in the FY 2017 IRF PPS final rule. Details of these considerations were provided in the public comment summary report (see link in U.2.2.2). For example, in response to public comments about the inclusion of hospice services, we added a risk adjustor for when a hospice claim begins within the beneficiary's episode window. This ensures that the IRF continues to have incentives for the efficient delivery of services, but also accounts for the higher cost of episodes with hospice. We also considered public comments about risk adjusting for prior hospital stays and added risk adjustors for length of prior inpatient and ICU stays.

Additionally, in response to public comments requesting more detail about the clinically unrelated excluded services, we provided detailed descriptions of the systematic process we used during development to identify clinically unrelated services.[1] This systematic process included organizing all claims into meaningful service categories, populating all services representing significant costs into a web tool used by clinicians with expertise in PAC care to determine service exclusions, and having multiple rounds of reviews to refine the list of exclusions.

We also considered other feedback that we did not implement. For example, commenters suggested extending the lookback period from 90 to 180 days. We had discussed this earlier with our TEP and investigated the impact of moving to the longer lookback period – we found that it would remove a large number of episodes due to beneficiary continuous enrollment requirements without a consistent increase in predictive power. As such, we did not extend the lookback period.

[1] The process for determining clinically unrelated services is described in Appendix D of the Measure Specifications, available at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/Downloads/2016_07_20_mspb_pac_ltch_irf_snf_measure_specs.pdf. The complete list of excluded services is available at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/IRF-Quality-Reporting/Downloads/Copy-of-2016_04_06_mspb_pac_irf_service_exclusions.xlsx.

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**

Not applicable

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.

Not applicable

U.4.1. Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.

No unexpected findings have been noted during implementation of this measure. Monitoring of patient characteristics and provider scores over time did not indicate unintended impacts on patients, to date. We are aware of the need to continuously monitor for unintended impacts on patients, such as cost-cutting at the expense of quality of care or avoiding complex patients. Our monitoring plans include monitoring trends in process and patient outcome measures, as well as trends in patient case-mix.

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

Not applicable

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)

2158 : Medicare Spending Per Beneficiary (MSPB) - Hospital

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

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.)

Not applicable. There are currently no measures that address both the same measure focus AND the same target population.

MSPB-PAC measures are harmonized across PAC settings as well as with MSPB-Hospital. MSPB-PAC measures were developed in parallel for all PAC settings to meet the mandate of the IMPACT Act. To align with the goals of standardized assessment across PAC settings, these measures were conceptualized uniformly across the four settings in terms of the construction logic, the approach to risk adjustment, and measure calculation. The measures mirror the general construction of MSPB-Hospital. Aligning the MSPB-Hospital and MSPB-PAC

measures in this way creates continuous accountability and aligns incentives to improve care planning and coordination across inpatient and PAC settings.

Contact Information

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

Co.2 Point of Contact: Ronique, Evans, ronique.evans1@cms.hhs.gov, 410-786-3966-

Co.3 Measure Developer if different from Measure Steward: Acumen, LLC

Co.4 Point of Contact: Mikhail, Pyatigorsky, mspb-pac-measures-support@acumenllc.com, 650-558-8882-

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.

A technical expert panel (TEP) was convened in Fall 2015. The TEP consisted of clinicians, researchers, and health care administrators with relevant expertise in PAC settings. TEP members provided input on measure conceptualization, definitions, specifications, exclusion criteria, unintended consequences, and other considerations related to development and implementation. The TEP summary report can be found at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/Technical-Expert-Panel-on-Medicare-Spending-Per-Beneficiary_Jan-2016.pdf.

TEP members' names and organizations:

1. Alma Allen; Inova VNA Home Health, Visiting Nurse Associations of America
2. Brian Bell; Spartanburg Regional Healthcare System
3. Dexanne Clohan; Foundation for Physical Medicine and Rehabilitation Evidence- Based Practice Committee of the American Academy of Physical Medicine and Rehabilitation
4. Jean de Leon; University of Texas Southwestern Medical Center
5. Scott Guevin; Penn State Hershey Rehabilitation Hospital, AMRPA
6. Kurt Hope; Mayo Clinic, Academy of Physical Medicine and Rehabilitation
7. Steven Lichtman; Helen Hayes Hospital
8. Craig Miller; Michigan Health & Rehabilitation Services, American Physical Therapy Association
9. Mary Ousley; American Health Care Association
10. Mary Shaughnessy; Partners Continuing Care, Spaulding Rehabilitation Network and Partners Health Care at Home
11. Christopher Vaz; American Hospital Association
12. Joanne Wisely; Genesis Rehab Services, AHCA, NASL

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2018

Ad.3 Month and Year of most recent revision: 09, 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? 09, 2020

Ad.6 Copyright statement:

Ad.7 Disclaimers:

Ad.8 Additional Information/Comments: