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 #:** 3562

**De.2. Measure Title:** Medicare Spending Per Beneficiary – Post Acute Care Measure for Long-Term Care Hospitals

**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 Long-Term Care Hospitals (MSPB-PAC LTCH) 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 LTCH's efficiency relative to that of the national median LTCH. Specifically, the measure assesses Medicare spending by the LTCH and other healthcare providers during an MSPB episode. The measure reports the ratio of the payment-standardized, risk-adjusted MSPB-PAC Amount for each LTCH divided by the episode-weighted median MSPB-PAC Amount across all LTCH facilities. 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 LTCHs. 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., LTCH admissions from October 1, 2015 through September 30, 2017.

Claims-based MSPB-PAC measures were developed in parallel for the LTCH, inpatient rehabilitation facility (IRF), 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 LTCH expertise evaluated LTCH claims and then gave direction on how to adjust for specific patient and case-mix characteristics.

The MSPB-PAC LTCH measure was adopted by the Centers for Medicare & Medicaid Services (CMS) for the LTCH Quality Reporting Program (QRP) and finalized in the FY 2017 LTCH Prospective Payment System (PPS) Final Rule.[1] The measure entered into use on October 1, 2016. Public reporting for the measure began in Fall

2018 through the LTCH Compare website (https://www.medicare.gov/longtermcarehospitalcompare/) using FY 2016-2017 data.

Notes:

[1] Medicare Program; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long Term Care Hospital Prospective Payment System and Policy Changes and Fiscal Year 2017 Rates. Federal Register, Vol. 81, No. 162. https://www.govinfo.gov/content/pkg/FR-2016-08-22/pdf/2016-18476.pdf

**IM.1.1. Developer Rationale:** MSPB-PAC LTCH 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 relationships between facility characteristics and resource use, links between LTCHs' financial incentives and strategic discharge of patients from facilities, and significant opportunities for improvement.[3,4,5,6] The cost and quality link is important, with this resource use measure playing an important role in discerning value of LTCH care.

The MSPB-PAC LTCH measure was adopted by CMS for the LTCH Quality Reporting Program (QRP) and finalized in the FY 2017 LTCH Prospective Payment System (PPS) Final Rule.[7] Public reporting for the measure began in Fall 2018 through the LTCH 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] Medicare Payment Advisory Commission. (2019). Report to the Congress: Medicare and the Health Care Delivery System.

[4] Einav, L., Finkelstein, A., & Mahoney, N. (2018). Provider Incentives and Healthcare Costs: Evidence From Long-Term Care Hospitals. Econometrica, 86(6), 2161-2219.

[5] Eliason, P. J., Grieco, P. L., McDevitt, R. C., & Roberts, J. W. (2018). Strategic patient discharge: The case of long-term care hospitals. American Economic Review, 108(11), 3232-65.

[6] Einav, L., Finkelstein, A., & Mahoney, N. (2018). Long-term care hospitals: A case study in waste (No. w24946). National Bureau of Economic Research.

[7] Medicare Program; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long Term Care Hospital Prospective Payment System and Policy Changes and Fiscal Year 2017 Rates. Federal Register, Vol. 81, No. 162. https://www.govinfo.gov/content/pkg/FR-2016-08-22/pdf/2016-18476.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**

## Criteria 1: Importance to Measure and Report

**1a. High impact or high resource use:**

The measure focus addresses:

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

AND

**1b.** Opportunity for Improvement**:**

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

---

**1a. High Impact or high resource use.**

- The focus of this measure is to assess Medicare spending for Long-Term Care Hospitals (LTCH's) and other healthcare providers during an MSPB-PAC LTCH 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 LTCH divided by the episode-weighted median MSPB-PAC Amount across all LTCHs.
- 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 relationships between facility characteristics and resource use, links between LTCHs' financial incentives and strategic discharge of patients from facilities, and significant opportunities for improvement.
- According to the March 2020 Medicare Payment Advisory Commission (MedPAC) report, the 374 LTCHs that participated in the Medicare program provided about 102,000 LTCH stays to 92,000 Medicare fee-for-service (FFS) beneficiaries, and Medicare FFS spending on LTCH services was $4.2 billion in 2018. On average, FFS beneficiaries accounted for about 60 percent of LTCH stays.

**1b. Opportunity for Improvement:**

- The developer provided publicly reported measure score data for all US providers under Medicare's LTCH Prospective Payment System (PPS) with 20 or more eligible episodes in the reporting period of FY 2016-2017. The data from 422 LTCHs with 20 or more episodes in the reporting period of FY 2016-2017, which include measure scores from 153,864 patient episodes shows the mean of 1.00 with a standard deviation of 0.08. The developer also reported the interquarile range of 0.09 (min: 0.76 and max: 1.50).
- 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:**

- **Required by legislation; per case costs high**
- **Yes**
- **Yes.**
- **post acute care spending varies widely and is significant contributor to Medicare spending. Measure developer cites the importance of cost quality link, but cites no evidence of there being one; however, as part of their validity testing, they see weak association between quality and resource use as measured by this measure.**
- **Yes (required by IMPACT Act)**
- **The developer describes 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.**
- **No concerns**
- **Yes, based on large numbers and high resource use**

**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:**

- **About a 20% difference in 10th to 90th percentile. Reasonable to attack**
- **Yes**
- **Yes, the developer provides publicly-reported data on Medicare's LTCH PPS with ?20 eligible episodes for FY 2016-2017. The mean measure score was 1 (SD=0.08) and IQR of 0.09 (Min = 0.76 and Max = 1.50), with 18% and 23% of the LTCHs having statistically significantly lower and higher than the national average, respectively.**
- **yes, there is variation and opportunity for improvement**
- **The data from 422 LTCHs with 20 or more episodes in the reporting period of FY 2016-2017, which include measure scores from 153,864 patient episodes shows the mean of 1.00 with a standard deviation of 0.08. The developer also reported the interquarile range of 0.09 (min: 0.76 and max: 1.50).**

- **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 LTCH costs.**
- **No concerns**
- **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-5; M-2; L-0; I-0
Methods Panel Individual Validity Ratings: H-2; M-3; L-2; 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 85 – 86.

**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 used two different measures of reliability: 1) Reliability score (signal to noise) to evaluate the extent to which variation in the measure is due to true, underlying differences in provider performance (signal) rather than random variation (noise) and 2) split-sample reliability testing (intraclass correlation or ICC) to examine agreement between two scores for a facility based on randomly-split, independent subsets of Long-Term Care Hospitals (LTCH)

- The performance measure score reliability testing was based on 422 LTCH with 20 or more episodes in the measurement period of 2016-2017.
- The developer reported that the mean reliability score for all agencies was 0.87 with median of 0.90. When examined by facility size, the average reliability score ranged from 0.75 (Q1) to 0.94 (Q4). The ICC for the overall sample was 0.86 with 95% confidence interval of 0.84-0.89. The ICC was lowest for Q1 (0.86) and highest in Q4 (0.90).

**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:**

- **Specifications are clear except for basis for excluding unrelated costs.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.**
- **Why 30-day post-discharge window? How is death handled within episode definition?**
- **No concerns.**
- **Elements are defined Algorithm is complex.**
- **none**
- **No concerns that were not raised by the Scientific Methods Panel.**
- **No concerns**
- **No concerns**

**2a2. Reliability – Testing: Has the developer demonstrated that the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers?Describe any additional concerns you have with the reliability testing results or approach that were not raised by the Scientific Methods Panel.**

**Comments:**

- **S/N ratios sufficiently high, no major concern**
- **Yes**
- **None.**
- **yes, they have demonstrated in a split sample they can reproduce the results**

- **none**
- **No concerns that were not raised by the Scientific Methods Panel.**
- **No concerns**
- **No concerns**

## Validity

**2b1. <u>Specifications align with measure intent</u>:**
The measure specifications are consistent with the measure intent and captures the most inclusive target population.
**2b2. <u>Validity Testing</u>:**
Demonstration that the measure data elements are correct and/or the measure score correctly reflects the cost of care or resources provided.
**2b3. <u>Exclusions:</u>**
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. <u>Risk Adjustment:</u>**
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. <u>Meaningful Differences:</u>**
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. <u>Multiple Data Sources</u>:**
If multiple data sources/methods are specified, there is demonstration that they produce comparable results.
**2c. <u>Disparities</u>:** 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 LTCH. This attribution approach was developed in order to encourage LTCHs to facilitate care coordination 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 LTCH (measure endorsed by NQF (#3480).
    - 3) Correlated this measure with the Percent of Residents or Patients with Pressure Ulcers That Are New or Worsened (Short-Stay) measures (NQF #0678); Catheter-associated Urinary Tract Infection (CAUTI) Outcome Measure (#0138); Central line-associated Bloodstream Infection (CLABSI) Outcome Measure (#0139); and Facility-wide Inpatient Hospital-onset Clostridium difficile Infection (CDI) Outcome Measure (#1717).
- 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 0.94, compared with 1.23 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 1.00, compared to 1.02 for episodes with at least one ER visit (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.
- The developers found a small, significant negative association between MSPB measure scores and the DTC measure scores as hypothesized and a small but statistically significant correlation (Pearson -0.2063; Spearman -0.2916) between MSPB measure scores and DTC measure scores.
- Lastly, the developer found a small, not significant correlation (both Pearson and Spearman) between MSPB PAC-LTCH measure scores and Pressure Ulcers measure scores and infection measure scores.

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

- The developer reports 36.3% of episodes were excluded because of one or more exclusion criteria.

**2b4/2c. Risk adjustment**

- The developer uses a linear regression risk adjustment model with 140 risk factors.
- The developer reported results showing that spending was higher for patients who were dual eligible ($71,554 vs $69,626 for non-duals), non-white ($78,546 for Asians compared to $68,452 for whites).
- 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 422 facilities, the mean score is 1.00 with a standard deviation of 0.08, and a minimum score of 0.76 with a maximum of 1.50. The 10th percentile was 0.90 and the 90th percentile 1.09.

**Questions for the Committee regarding validity:**

- *Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?*

**Guidance from validity 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:**

- **1.No data element testing. Assumption is that claims data is sufficiently accurate. Consistent approach to data element assessment for CMS claims based measures. NEEDS FURTHER DISCUSSION. 2. Also see discussion in 5.2a above about exclusion of "clinically unrelated services. 3. Standardized pricing has strengths and limitations in understanding resources used.**
- **None**
- **No concerns.**
- **Unnecessarily excludes episodes below the 1st percentile and above the 99th, vs. truncating or winsorizing those value. Methodology doesn't separate concurrent events, resulting in overlapping episodes if I understand this correctly. How often does this occur? The measure developer says it is important for continuous accountability. I am not sure I track how this creates continuous accountablity. Overlapping episodes seems problematic, especially when provider #1 isn't controlling the spend once the patient is discharged and then admitted to another facility. Correlating the measure with other measures of use and cost included in the measure (ER, hospitalizations) doesn't seem a good test of validity. I had concerns related to the weak (and not significant) relationship between measures of quality and the MSPB. One would think that these acquired conditions would be related to spending, unless they are relatively rare, in which case they aren't good measures to test validity**
- **Intent for consistency with other PAC MSPB measures yet this one references 31 day period following discharge to community when the IRF measure states 30 days.**
- **No concerns not raised by the Scientific Methods Panel.**
- **No concerns**
- **No concerns**

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

**Comments:**

- **None**
- **Not clear how death is handled if poor care leads to increased death care post-discharge.**
- **None.**

- **there are meaningful differences across LTCHs on this measure. Key concern would be threat from not adjusting for social risk factors.**
- **none**
- **No concerns not raised by the Scientific Methods Panel.**
- **No concerns**
- **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:**

- **Okay**
- **None.**
- **Missing data is reported to be nil.**
- **none**
- **None.**
- **No concerns**
- **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:**

- **No concerns about attribution. Standard issues associated with CMS standardized prices. Standard exclusion of extreme cost cases, and winsorizing of outliers.**
- **None**
- **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?**

- **Don't agree with how they handle outlier episodes (by excluding them). No issues with attribution save the issue of overlapping episodes and counting the costs in both.**
- **none**
- **None.**
- **No concerns**
- **No concerns**

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

**Comments:**

- **None.**
- **The concern from Panel member #3 about the exclusion of 36.3% of episode due to one or more exclusion criteria needs discussion. Agree that this might call into question the overall validity of the measure.**
- **none**
- **none**
- **I am concerned with the inclusion of patients with ventilator services for 96+ hours. What would be the opportunity for cost savings in this specific population. This becomes a larger issue with an influx of long-term ventilator patients as a result of COVID-19. Another threat to validity is the availability of LTCH services. For example, in New York State, there is 1 LTCH. Other LTCH services are provided in hospital or nursing facility spaces. This may impact costs and also change the population of patients who would be admitted to an LTCH.**
- **No concerns**
- **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:**

- **Risk adjustment approach is standard approach for CMS measures. It appears to take into account underlying differences in disease/chronic conditions/age. Because admission to LTCH is usually preceeded by prior admission, there is good coding of condition likely to lead to LTCH admission, and 30 day window post LTCH appears reasonably short, allowing for readmission and other OP services due to original condition for which hospitalized. This is reflected in the high R-square for the risk adjustment, 0.34 for site neutral strata, and 0.47 for Standard strata. Social risk factors analyzed, and as implemented in risk adjustment model, make small-negligible difference in score or adjusted R-square.**
- **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.**
- **While the measure developer offers rationale for not including SES in the final measure, I think 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.**
- **Social risk factors, while tested are excluded as measure developer says they lead to changes (though I'm curious what it looks like for more extreme cases that have a sizeable population with social risk factors). Not adjusting can create peverse incentives for providers to avoid these patients. Hard to understand how the characteristics of benes varies across LTCHs as the developer only shows for entire population of patients, and not by facility (what is mean # of duals and the range across faclities)?**
- **Developer reported results showing spending was higher for patients who were dual eligible ($71,554 vs $69,626 for non duals, or 3% higher which can be a significant amount in the Medicare payment world, non-white (e.g., $78,546 for Asians compared to $68,452 for whites, or 14.7% higher which can have very significant impact on a plan with many non-white members, 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 based on overall R-Square of 0.489. "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.**
- **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 LTCH - LCDS that could be meaningful in the model including: mental status, confusion, depression, nutrition, language, health literacy, transportation, and social isolation.**
- **No concerns**
- **No concerns**

## Combined Scientific Methods Panel Preliminary Analysis of Scientific Acceptability

**Measure Number:** 3562

**Measure Title:** Medicare Spending Per Beneficiary – Post Acute Care Measure for Long-Term Care Hospitals

**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** Measure specification and testing was based on national data. All eligible LTCH patient episodes with LTCH discharges between October 1, 2015, and September 30, 2017, were included in the measure. For patients with multiple LTCH episodes during the measurement period, all eligible episodes were included. A total of 157,004 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 153,864 episodes.

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, 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 LTCH spending. It likely has applicability only to the CMS LTCH compare program for public reporting.

   **Panel Member #3** SEE RELATED COMMENTS REGARDING TERMS, MEASURE SCORE, MEASURE IN CURRENT USE, APPROPRIATE USE OF DATA, TIME PERIOD FOR PERFORMANCE, STANDARDIZING COST OF CARE, AND CLINCIAL DIFFERENCES AMONG PATIENTS ACROSS FACILITIES FROM MY REVIEW OF MEASURE 3561.

   **Panel Member #7** No concerns

   **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 first concern is with the outlier exclusion (Step 6 – Contruction logic). Based on the measure specifications, episodes with residuals below the 1$^{st}$ percentile or above the 99$^{th}$ percentile of the residual distribution are excluded. Four 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. 4) Given the substantial difference between Standard episode and Site Neutral episode, outlier exclusion based on residuals may favor one type of episode. 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.

My second concern is with the exclusion of episodes due to patients' death. While I understand the reason behind that, high death rate (24.5%) and substantial variation of this rate among facilities (table 9 testing form) call in to question how this measure score can be interpreted.

### 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**

   **Panel Member #3** Note: 422 of the 429 LTCH's were included in the analyses because they met the sample size per facility of 20 stays over a 2-year period.

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 analysis (reliability score) and split-sample reliability sampling (Intraclass correlation or ICC) were assessed.

   **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 LTCH 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 with episodes across years evenly distributed. They used the Shrout-Fleixx interclass correlation coefficient (ICC) between the split-half scores to measure reliability. They also calculated ICCs between split-half scores stratified by facility size to assess whether reliability was acceptable across providers of varying sample size.

   **Panel Member #3** Split-half, Shrout-Fleiss ICC approach; and "undefined" Reliability Score approach

   **Panel Member #7** Split sample reliability testing was utilized with four years of data as well as intraclass coefficient. The quartile means and ranges increased from Quartile 1 to Quartile 4.

   **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 the reliability test results with a highe reliability score (0.87) and ICC of 0.86, the performance measure score reliability is good.

   **Panel Member #2** Overall, reliability testing indicated good reliability, regardless of facility size. The average reliability score for all agencies was 0.87 with median of 0.90. When examined by facility size, the average reliability score ranged from 0.75 (Q1) to 0.94 (Q4). The ICC for the overall sample was 0.86 with 95% confidence interval of 0.84-0.89. The ICC was lowest for Q1 (0.86) and highest in Q4 (0.90)

   **Panel Member #3** Generally strong results (0.89 = overall median), with some variation based on size of LTCH (0.78 for smallest LTCHs to 0.95 for largest LTCHs based on episode count).

   **Panel Member #7** Facility reliability score had a mean of 0.87 overall with a range of 0.85 for the 25$^{th}$ percentile and 0.93 for the 75$^{th}$ percentile. Split sample had an intraclass coefficient of 0.86, which was also true for all quartiles except the 4$^{th}$ Quartile, which was 0.90

   **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 6 and 7 above for my rationale.

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

**Panel Member #3** Definition of "Reliability Score" was not presented clearly; variation in reliability based on size of LTCH is problematic if the measure is used for comparison purposes with financial (e.g., bonuses, rewards, changes in reimbursement) nationally.

**Panel Member #7** Reliabilty of the measure score was good across all quartiles, and as an overall group.

**Panel Member #8** NA

**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.** None

**Submission document:** Testing attachment, section 2b2.

**Panel Member #2** NONE

**Panel Member #3** How do seven LTCH's with fewer than 20 episodes of care across two years stay in business and maintain accreditation standards?

This measures exclusions reference "31 days following discharge to community" rather than the 30-day post discharge presented for IRFs. Why the difference?

Substantial elimination of data due to failure of one or more exclusion criteria: Overall, 36.3% of episodes were excluded because of one or more exclusion criteria. This also threatens the overall validity (i.e., generalizability of results across the Medicare population) of the MSPB measure for LTCHs.

**Panel Member #7** No concerns

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

**Submission document:** Testing attachment, section 2b4.

**Panel Member #1** None

**Panel Member #2** NONE

**Panel Member #3** Checked for relationship between score and events such as hospitalization (generally in the expected direction); and for clinical improvement of patients (essentially no relationship between cost of care and four clinical outcomes tested).

These poor results influence whether differences in MSPB (cost of care) between LTCHs can be considered "meaningful differences."

**Panel Member #7** Across 422 facilities, the mean score is 1.00 with a standard deviation of 0.08, and a mimimum score of 0.76 with a maximum of 1.50. The $10^{th}$ percentile was 0.90 and the $90^{th}$ percentile 1.09

**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 #3** Multiple data sources are used, but combining information using LTCH IDs does not seem to be a problem.

**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, given that this would be based in claims analysis, and the exclusion restrictions would ensure that there is rarely any missing data in the analytic dataset that will be used to calculate the performance measure.

**Panel Member #2** NONE

**Panel Member #3** Seems to be very limited missing data, given that the information to compute the measure is derived primarily from claims data with the obvious monetary incentive to submit these data properly.

**Panel Member #7** Missing (problematic) data was present in 0.01% of the episodes and were excluded.

**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 X—
**Panel Member #3** Maybe?
The outline the elements in the prediction model used for risk adjustment were identified for the one of the stratifications (standard payment rate cases), but not for the second (site neutral payment rate case).
Were the same prediction model risk factors used for both stratified models, but allowed to generate different coefficients or where different risk factors used for the two different stratifications?

16d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐ Yes ☒☐ No

16d.3 Is the risk adjustment approach appropriately developed and assessed? ☐☒ Yes ☒☐ No
**Panel Member #3** My answer is based on the lack of information regarding the lack of information about the "site neutral" prediction model.

16d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☐☒ Yes ☒☐ No
**Panel Member #3** The adjusted R-Squared for "standardized" was 0.341, while the same value for the "site neutral" was 0.4697. The authors claim to use a "fully interacting episode type" (whatever that is) of 0.4894—which is higher than either stratification alone. Perhaps there is a problem in the way the information is presented in the document.

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** The MSPB-PAC LTCH risk adjustments were done separately for MSPB-PAC LTCH episode types: LTCH standard and LTCH site neutral (defined in 2b3.1.1) in the testing document. These models are adaptations 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.

The overall adjusted R-squared of the risk model used was 0.489. 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 episode-level predicted costs range from $32,267 to $115,906, indicating that this model has a range of predictions and can predict both high and low costs.

**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 ($71,554 vs $69,626 for non duals, or 3% higher which can be a significant amount in the Medicare payment world, non-white (e.g., $78,546 for Asians compared to $68,452 for whites, or 14.7% higher which can have very significant impact on a plan with many non-white members, 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 based on overall R-Square of 0.489. This is NOT surprising given there were already more than 140 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. The small impact on overall average scores is more likely due to the model specification not capturing the significant differences in costs for patients with SES as described above. Given the empirical findings of 3-15% higher costs, which is significant for Medicare plans, I strongly believe the factors should have been kept in the model. In cases where an LTCH 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 LTCHs for these patients.

**Panel Member #3** The strategy (prediction model and stratification) is conceptually correct**.** However, the lack of clarity regarding the "site neutral" prediction model, and the "fully interacting episode type" does not allow for a convincing argument that the overall results of the risk-adjustment effort was successful.

**Panel Member #7** The overall R-squared was 0.4894, 0.341 for standard episodes and 0.4697 for site neutral episodes. The observed to predicted costs varied between 0.98 and 1.02 for the ten deciles.

**Panel Member #9** Risk-adjustment approach is acceptable although I am not sure about inclusion of MS-LTC-DRGs as risk variable. It may explain the high adjusted R-square.

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

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

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

    **Panel Member #3** The specifications as stated at the outset appear to be in alignment. However, the execution (how these specifications were operationalized) create uncertainty if the resulting measure value matches the measure intent.

18. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):** None

    **Panel Member # 3** See comment #17.

**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 performance measure score validity was sought to be assessed through following correlations: (i) correlation with known indicators of resource or service utilization, specifically hospital admissions and emergency room (ER) visits during the episode period; (ii) correlation between MSPB-PAC LTCH scores and the Discharge to Community (DTC) rates for FY 2016-2017; and (iii) correlation between MSPB-PAC LTCH scores and several assessment-based quality measure scores publicly reported on the LTCH Compare website for FY 2017.

    **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 LTCH 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 LTCH scores and the Discharge to Community (DTC) rates for LTCHs (measure endorsed by NQF (#3480) is publicly reported and also based on Medicare claims. 3. Examined the correlation between MSPB-PAC LTCH 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); Catheter-associated Urinary Tract Infection (CAUTI) Outcome Measure (#0138); Central line-associated Bloodstream Infection (CLABSI) Outcome Measure (#0139); and Facility-wide Inpatient Hospital-onset Clostridium difficile Infection (CDI) Outcome Measure (#1717).

    **Panel Member #3** Checked for relationship between score and events such as hospitalization; and for clinical improvement of patients

    **Panel Member #7** The MSPB-PAC LTCH were examined for correlation with known indicators for resource utilization, specifically hospitalization and emergency room visits**.** Also correlation was examined with discharge to community rates. Finally, they were also examined for correlation with quality measure scores for LTCH's, pressure ulcers, CAUTI, CLABSI, and C-diff.

    **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 a set of quality measures. 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 positive relationship between MSPB-PAC LTCH and hospitalizations and ER visits 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 LTCHs are associated with better discharge to community rates. The low but statistically insignificant correlation between MSPB and Pressure Ulcers and infections measures is consistent with the finding that these measures are not associated with lower rates of successful discharge to community. In summary, these correlations establish the hypothesized relationships between MSPB-PAC LTCH score and other publicly reported measures.

    **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 0.94, compared with 1.23 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 1.00, compared to 1.02 for episodes with at least one ER visit (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 small but statistically significant correlation (Pearson -0.2063; Spearman -0.2916) between MSPB measure scores and DTC measure scores.

    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 LTCHs are associated with better discharge to community rates and fewer unplanned hospitalizations. However, they found very small and not significant correlations between MSPD measure socres and pressure ulcers and infection measure scores, indicating measures of functional improvement are not associated with higher costs.

    **Panel Member #3** Checked for relationship between score and events such as hospitalization (generally in the expected direction); and for clinical improvement of patients (essentially no relationship between cost of care and four clinical outcomes tested).

    These poor results influence whether differences in MSPB (cost of care) between LTCHs can be considered “meaningful differences.”

    **Panel Member #7** The results showed that the correlation was strong for other drivers of cost, but no significant for the quality measures.

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

    **Panel Memver #9** Given the nature of their relationship, the results are as expected for test 1 and 2. The results from test 3 are hard to interpret.

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 are threats to validity and/or relevant threats to validity were not assessed OR if testing methods/results are not adequate)

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

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

**Panel Member #1** Please see my rationale in 21 and 22.

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

**Panel Member #3** The lack of specificity regarding the "site neutral" prediction model, the poor ability to identify meaningful differences, the lack of a relationship between the measure and functional clinical outcomes led to my validity rating as "low."

**Panel Member #8** Low correlation with ER visits concerning.

**Panel Member #9** Mainly due to two concerns I outlied in the section 2.

### FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

☐ **High**

☐ **Moderate**

☐ **Low**

☐ **Insufficient**

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

### ADDITIONAL RECOMMENDATIONS

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

**Panel Member #9** Yes.

My first 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. Four 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. 4) Given the substantial difference between Standard episode and Site Neutral episode, outlier exclusion based on residuals may favor one type of episode. 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.

My second concern is with the exclusion of episodes due to patients' death. While I understand the reason behind that, high death rate (24.5%) and substantial variation of this rate among facilities (table 9 testing form) call in to question how this measure score can be interpreted.

## 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 mentioned that since data are already collected as part of Medicare's payment process, this measure poses no 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:**

- **Measure is routinely compiled from claims and related data, so it is feasible.**
- **Approach appears feasible**
- **No concerns.**
- **yes, construction of this measure is feasible as it relies on secondary data (claims, administrative data, census data)**
- **Think the availability of social risk data that would improve risk adjustment not yet routinely captured**
- **This measure is feasible to collect and calculate.**
- **No concerns**
- **Data generated routinely**

## 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' Long-Term Care Hospital Quality Reporting Program (LTCH QRP). The data are publicly reported on LTCH 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 LTCH QRP includes all LTCHs paid under the LTCH PPS. MSPB-PAC LTCH 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 throughhought 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 LTCH PPS final rule

**Additional Feedback:**

- The developers 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 LTCH 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 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:**

- **Current use is public reporting for accountability, not yet used for payment.**
- **Unclear**
- **No additional comments beyond those noted by the NQF staff.**
- **CMS started publicly reporting the measure in 2018, after providing LTCH confidential reporting in 2017.**
- **yes; Data is publicly available, though unclear whether the measure is easily replicable by others who may want to calculate and compare LTCH spending. It likely has applicability only to the compare program for public reporting.**
- **This measure has been publicly reported since Fall 2018.**
- **No concerns**
- **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.**
- **yes, CMS has given LTCHs the performance results and the LTCHs have been allowed to provide comments in the draft regulations**
- **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 provided LTCHs with sufficient information to reduce costs.**
- **No concerns**
- **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:**

- **This is not clear. 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. There is no indication of areas of high spending or low spending by provider associated with IRF, so not clear if providers have enough information to taret improvement.**
- **Yes, the potential for improvement was demonstrated by developer through adequate incentive for high-cost LTCH compared to the national mean.**
- **the information is shared with LTCH indicating whether the LTCH is above, below or no different than the median nationally. By showing this number, the LTCH would then need to understand its cost drivers (which this measure does not spotlight, so doesn't go the next step to help, unless they break out categories of episodes for LTCHs to see what are some of the drivers). Measure developer asserts that this cost metric combined with quality measures gets at value**
- **Seems to assume it can be used for improvement**
- **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 LTCH costs.**
- **No concerns**
- **Planned use**

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

**Comments:**

- **No obvious harms.**
- **None.**
- **I do worry that absent inclusion of social risk factors that some LTCHs may avoid those with greater social risk factors, especially if CMS ties money to this metric. At this stage, given little data provided, it is hard to know whether this is the case. Would be helpful if the developer could share what the distribution of these factors is across LTCHs (e.g., do some LTCHs have 50 or 100% duals?)**
- **It seems that by not including SES factors in risk adjustment there is a real concern that for LTCHs 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 LTCHs for these patients.**

- **The measure could be more beneficial is the developers explored appropriate resource use and its relationship to health outcomes.**
- **No concerns**
- **No concerns**

## Criterion 5: Related and Competing Measures

- There are no competing measures for #3562 (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:**

- **Part of a complementary set of PAC cost measures. LTC hospitals may be distinctive enough that limited concern of overlap of episodes for these providers.**
- **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 competing measures. Other MSPB measures are related**
- **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.**
- **none.**
- **No concerns**

## 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
  8391 :
  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 22-197 episodes was 0.70, 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 0.4894. 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.

The FAH appreciates that social risk factors were reviewed, and believes 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 LTCH 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 relationships between facility characteristics and resource use, links between LTCHs' financial incentives and strategic discharge of patients from facilities, and significant opportunities for improvement.[3,4,5,6] The cost and quality link is important, with this resource use measure playing an important role in discerning value of LTCH care.

The MSPB-PAC LTCH measure was adopted by CMS for the LTCH Quality Reporting Program (QRP) and finalized in the FY 2017 LTCH Prospective Payment System (PPS) Final Rule.[7] Public reporting for the measure began in Fall 2018 through the LTCH 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] Medicare Payment Advisory Commission. (2019). Report to the Congress: Medicare and the Health Care Delivery System.

[4] Einav, L., Finkelstein, A., & Mahoney, N. (2018). Provider Incentives and Healthcare Costs: Evidence From Long-Term Care Hospitals. Econometrica, 86(6), 2161-2219.

[5] Eliason, P. J., Grieco, P. L., McDevitt, R. C., & Roberts, J. W. (2018). Strategic patient discharge: The case of long-term care hospitals. American Economic Review, 108(11), 3232-65.

[6] Einav, L., Finkelstein, A., & Mahoney, N. (2018). Long-term care hospitals: A case study in waste (No. w24946). National Bureau of Economic Research.

[7] Medicare Program; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long Term Care Hospital Prospective Payment System and Policy Changes and Fiscal Year 2017 Rates. Federal Register, Vol. 81, No. 162. https://www.govinfo.gov/content/pkg/FR-2016-08-22/pdf/2016-18476.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 LTCH measure scores are reported publicly for all US providers paid under Medicare's LTCH Prospective Payment System (PPS) with 20 or more eligible episodes in the reporting period. There were a total of 422 LTCHs with 20 or more episodes in FY 2016-2017. Their scores represent 153,864 patient episodes, after all exclusions were applied. The scores show a good deal of variability – the descriptive statistics are provided below.

MSPB-PAC LTCH score descriptive statistics:

Mean: 1.00

Standard Deviation: 0.08

Min: 0.76

Max: 1.50

Interquartile range: 0.09

Score Percentiles

10: 0.90

20: 0.94

30: 0.96

40: 0.98

50: 0.99

60: 1.01

70: 1.03

80: 1.05

90: 1.09

**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 LTCH 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 LTCHs and other healthcare providers during an MSPB-PAC LTCH episode. An MSPB-PAC LTCH 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 LTCH) and ends 30 days after the discharge from that LTCH. The measure is calculated as the ratio of the payment-standardized, risk-adjusted MSPB-PAC Amount for each LTCH divided by the episode-weighted median MSPB-PAC Amount across all LTCHs. 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 LTCHs.

## 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/LTCH-Quality-Reporting/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 LTCH 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 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 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
- 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 LTCH facilities 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.

- LTCH Compare data: We used this data source to examine the relationship between MSPB and assessment-based quality measures. The LTCH Compare data include publicly reported LTCH quality measures. The data are available at https://data.medicare.gov/data /long-term-care-hospital-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-637237729634498496.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 LTCH measure assesses Medicare spending by LTCH facilities and other healthcare providers during an MSPB-PAC LTCH episode. An MSPB-PAC LTCH 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 LTCH) and ends 30 days after the discharge from that LTCH. The measure is calculated as the ratio of the payment-standardized, risk-adjusted MSPB-PAC Amount for each LTCH divided by the episode-weighted median MSPB-PAC Amount across all LTCH facilities. 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 LTCH facilities.

An MSPB-PAC LTCH measure score of less than 1 indicates that a given LTCH's resource use is less than that of the national median LTCH during a performance period. An MSPB-PAC LTCH measure score of greater than 1 indicates that a given LTCH's resource use is more than that of the national median LTCH 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 LTCH 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 LTCH 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 LTCH episodes is admission to the LTCH, except for readmissions occurring within 7 days to the same provider. The LTCH that triggers the episode is the provider to whom the episode is attributed for the purpose of calculating the MSPB-PAC LTCH measure (attributed provider). We identify an admission to an LTCH based on the inpatient claims with the third character of CMS Certification Number (CCN) equal to "W", or with last four digits of CCN that fall within the range of 2000-2299.[1]

The MSPB-PAC LTCH measure allows different types of claims to trigger different episodes, reflecting differences in payment policy and beneficiaries' underlying health characteristics. In the LTCH setting, the dual payment rate structure, as detailed in the FY 2016 LTCH PPS Final Rule, distinguishes between standard payment rate cases and site neutral payment rate cases.[2] A standard payment rate case is one that is not a psychiatric or rehabilitation Medicare Severity Long-Term Care Diagnosis-Related Group (MS-LTC-DRG), is immediately preceded by an acute care hospital stay, and either (i) the acute care hospital stay included at least 3 days in intensive care unit (ICU) or coronary care unit (CCU) or (ii) the beneficiary received 96+ hours of ventilator services. A site neutral payment rate case is one that does not meet the definition of a standard payment rate case. A standard payment rate case triggers an MSPB-PAC LTCH Standard episode, while a site neutral payment rate case triggers an MSPB-PAC LTCH Site Neutral episode. LTCH Standard and Site Neutral episodes are compared only with LTCH Standard and Site Neutral episodes, respectively, to ensure that the measure is making fair comparisons between clinically similar beneficiaries.

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

Treatment Period. The treatment period of an MSPB-PAC LTCH 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 LTCH occur within 7 or fewer days of one another, the treatment period ends on the day of discharge for the latest LTCH 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 LTCH provider, and that are related to the beneficiary's care plan. Treatment services occurring on the first day of MSPB-PAC LTCH episodes are subject to exclusions related to prior institutional care, including ambulance transport to the attributed LTCH facility and durable medical equipment, prosthetics, orthotics, and supplies (DMEPOS) orders preceding the patient's admission to the LTCH. 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 LTCH 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 LTCH 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 LTCH 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 LTCH 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 LTCH episode may begin during the associated services period of another MSPB-PAC LTCH episode in the 30 days post-treatment. See section S.7.3 for examples of situations in which this occurs and how it is handled in the MSPB-PAC LTCH measure.

Measure Calculation

Certain episodes are excluded from the MSPB-PAC LTCH calculation to ensure that the measure facilitates meaningful comparisons between LTCH providers. These exclusions are distinct from the exclusions for clinically unrelated services discussed above, which exclude a limited set of services from MSPB-PAC LTCH 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).[3] All payment data shown in the MSPB-PAC LTCH 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. The calculation is performed separately for Standard and Site Neutral episodes (described above). 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.

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 LTCH 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 LTCH 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 LTCH Measure. The MSPB-PAC LTCH measure is calculated for individual providers, allowing them to be compared relative to other LTCH providers nationally. The MSPB-PAC LTCH measure is calculated as the ratio of the MSPB-PAC Amount for each LTCH provider divided by the episode-weighted median MSPB-PAC Amount across all LTCH providers. MSPB-PAC LTCH measure calculation is performed separately for LTCH Standard and Site Neutral episodes to ensure that they are compared only to other episodes of the same type. The final MSPB-PAC LTCH measure combines the ratios of the episode types to construct one provider score.

To calculate the MSPB-PAC Amount for each LTCH provider, one calculates the ratio of the standardized spending for LTCH Standard episodes over the expected spending (as predicted in risk adjustment) for LTCH Standard episodes, and the ratio of the standardized spending for LTCH Site Neutral episodes over the expected spending (as predicted in risk adjustment) for LTCH Site Neutral episodes, and then averages these ratios across all episodes for the attributed provider. This quantity is then multiplied by the average episode spending level across all LTCH providers nationally for Standard and Site Neutral episodes.

Mathematically, MSPB-PAC LTCH for individual provider k is:

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

The numerator is the MSPB-PAC LTCH Amount, or the average risk-adjusted episode spending across all episodes for the attributed provider, comparing Standard and Site Neutral episodes only with episodes of the same type. This is then multiplied by the national average episode spending level for all LTCH providers nationally. Mathematically, the MSPB-PAC LTCH 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 LTCH Amounts for all LTCH facilities nationally.

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

Notes:

[1] Resdac – Long-term care hospitals in the Medicare data. https://www.resdac.org/articles/long-term-care-hospitals-medicare-data

[2] Medicare Program; Hospital Inpatient Prospective Payment System for Acute Care Hospitals and the Long-Term Care Hospital Prospective Payment System and Fiscal Year 2016 Rates. Federal Register, Vol. 80, Num. 158. https://www.govinfo.gov/content/pkg/FR-2015-08-17/pdf/2015-19049.pdf

[3] QualityNet, "CMS Price (Payment) Standardization Overview" (April 2019)) https://www.qualitynet.org/inpatient/measures/payment-standardization

**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-637135710443434103.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 LTCH measure methodology does not separate concurrent events. The MSPB-PAC LTCH measure methodology defines an MSPB episode as all claims from the start of admission to the LTCH to 30 days post LTCH discharge. Please refer to section S.8.4., which details the rationale for the construction of the MSPB-PAC LTCH episode, for a discussion of the advantages of this approach.

The definition of MSPB-PAC LTCH episodes allow one episode to overlap with other episodes. One possible scenario occurs where an LTCH provider discharges a beneficiary who is then admitted to another LTCH 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 LTCH stay will be included once as an associated service for the attributed provider of the first MSPB-PAC LTCH episode and once as a treatment service for the attributed provider of the second MSPB-PAC LTCH 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 LTCH episode includes all Medicare Part A and Part B services that fall within the episode window that starts at the LTCH index admission and ends at 30-days post LTCH discharge, except for a limited set of services that are excluded for being clinically unrelated to LTCH 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 LTCH 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 LTCH risk-adjustment model are included to account for beneficiary characteristics prior to the start of an MSPB-PAC LTCH 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 LTCH 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. Different logic is used to handle LTCH Standard episodes with multiple prior claims sharing the same end date. Given that LTCH Standard episodes are defined by the presence of a prior acute IP stay between 0 to 1 days prior to the start of the LTCH Standard episode, only information about this required prior hospitalization is used to assign the episode's clinical case mix category. Given the role of ICU days in

determining eligibility for LTCH Standard payment rates, the clinical case mix category is assigned based on the inpatient stay with the most ICU days. In the event of a tie in the number of ICU days, the clinical case mix category is assigned based on the IP claim with the longer length of stay. Should a tie still persist, the most recent IP claim by discharge date is used. Finally, if the prior criteria do not result in a category assignment, the original hierarchy above is used.

**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 LTCH measure values are included on Medicare claims and enrollment data. The data fields used to calculate the MSPB-PAC LTCH measure (e.g., payment amounts, DRGs, diagnosis and procedure codes, etc.) are included in all Medicare claims because LTCH facilities only receive payments for complete claims. We do have complete data for each beneficiary who has an MSPB-PAC LTCH 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 LTCH 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 LTCH measure assesses the standardized allowed amounts of services performed by LTCH facilities and other healthcare providers during an MSPB-PAC LTCH episode, which includes all Part A and Part B Medicare claims that occur at the LTCH admission through 30 days after discharge from the LTCH 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 LTCH 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 LTCH measure aims to calculate resource use in the period between the start of the treatment period and the end of the associated services period. An MSPB-PAC LTCH episode encompasses all procedures and clinical events that occur between the start of the treatment period and 30 days post LTCH discharge. The clinical topic area includes all LTCH admissions in the United States.

The MSPB-PAC LTCH measure accounts for differences in payment policy and beneficiaries' underlying health characteristics by stratifying by standard and site neutral payment rate admissions. An MSPB-PAC LTCH Standard episode is triggered by a standard payment rate claim, while an MSPB-PAC LTCH Site Neutral episode is triggered by a site neutral payment rate claim. To adjust for beneficiary characteristics that are out of the influence of the attributed LTCH 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. Risk adjustment is performed separately for MSPB-PAC LTCH Standard and Site Neutral cases. 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 LTCH episode. Finally, we account for Medicare Severity-Long-Term Care Diagnosis-Related Groups (MS-LTC-DRGs) identified on the LTCH admission claim.

**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.)*

Grouping methodology:

The grouping methodology includes all Medicare Part A and B services delivered to a beneficiary during the treatment period (from admission to the LTCH through to discharge from the LTCH) and associated services period (from admission to the LTCH through to 30 days after discharge from the LTCH). To simplify the clinical logic and avoid the issue of attributing claims to MSPB 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 LTCH measure.

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 LTCH 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 LTCH care.

Extensive clinical review was performed by clinicians with experience and expertise in LTCH, as well as in collaboration with Medical Officers at CMS. The inpatient, outpatient, Part B physician and supplier, and DMEPOS services least clinically related to the LTCH care were excluded from the measure. For instance, services related to the routine management of preexisting chronic conditions (e.g., dialysis for ESRD, treatment for preexisting cancers, and treatment for organ transplants) were felt to be clinically unrelated to the scope of the type of care that LTCHs provide. Therefore, these types of services were excluded. Services were excluded if there was consensus across clinicians from the measure developer, external clinical experts including TEP members, and CMS medical officers. Please see section S.9.1 for overall clinical consensus regarding the types of exclusions.

Attribution algorithm:

An MSPB-PAC LTCH episode is assigned to the facility of the index admission. A new episode may begin during the associated services period of a previous MSPB-PAC LTCH episode in the 30 days post-discharge from the LTCH. Further details about attribution are provided in section S.13.2.

Risk adjustment:

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. The MSPB-PAC LTCH 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. Diagnosis codes on claims that occur during the 90-day period prior to the start of an MSPB-PAC LTCH episode (90-day "look back") are used to create HCC indicators. For example, the measure accounts for interactions disability status 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 LTCH measure utilizes clinical case-mix categories to create clinically meaningful subgroups that influence the type of services a beneficiary will receive in an LTCH. 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 LTCH 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 LTCH episode. Taking the most recent institutional claim (by end date) in the 60 days prior to the start of an MSPB-PAC LTCH 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)[1]

6) Prior PAC - HHA – beneficiaries who are continuing PAC from a HHA[1]

7) Community – all other beneficiaries[1]

Finally, the MSPB-PAC LTCH Measure includes variables for MS-LTC-DRGs from the LTCH admission. A full list of the MS-LTC-DRGs used in the risk-adjustment model is included in Appendix C of the Measure Specifications document provided in section S.1.

Notes:

[1] This variable is only used in the MSPB-PAC LTCH Site Neutral risk adjustment model. Since LTCH Standard episodes are defined to have a prior IP stay, the MSPB-PAC LTCH Standard risk adjustment model omits variables representing non-IP sources of entry in the clinical case-mix categories.

**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 the MSPB-PAC measures 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. The MSPB-PAC LTCH Measure aims to provide actionable, transparent information to support LTCH providers' efforts to promote care coordination and improve the efficiency of care provided to their patients.

The Pathway for SGR Reform Act of 2013 created a dual-payment rate structure that altered how Medicare pays LTCH providers based on clinical criteria of LTCH patients prior to admission. The MSPB-PAC LTCH measure is stratified to account for these distinct patient populations. A standard payment rate claim triggers an MSPB-PAC LTCH Standard episode, while a site neutral payment rate claim triggers an MSPB-PAC LTCH Site Neutral episode. LTCH Standard and Site Neutral episodes are compared only with LTCH Standard and Site Neutral episodes, respectively, to ensure that the measure is making fair comparisons between clinically similar beneficiaries.

Accounting for patient clinical severity is essential for accurately predicting resource use within the MSPB-PAC LTCH episode. There is ample evidence in both inpatient and post-acute care settings that Medicare episode payments are associated with clinical case-mix or severity. The large variation in clinical severity among post-acute care patients, and with medical conditions and comorbidities, is a major driver of the variation in post-acute care spending.[2] 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 HCCs 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.[3]

Clinical severity groupings, such as MS-DRGs, have been shown to be associated with resource use in the LTCH post-acute setting. For example, the average full PPS payment among LTCHs for the DRG "Respiratory System Diagnosis with Prolonged Mechanical Ventilation" is $78,749, while the payment for the DRG "Aftercare with Complication Conditions or Major Complicating Conditions" is $27,153.[4] The percent of patients in the highest severity of illness category in LTCHs (43%) is more three times that of SNFs and IRFs (approximately 12%) and more than ten times that of HHAs (4%). A similar pattern is repeated in Medicare payments per day by setting.[5]

Patient admission sources, which are similar to clinical case-mix categories defined in section S.8.2, are also associated with resource use. LTCHs with more admissions from acute care hospitals were associated with more resource usage that those with more admissions from the community.[6] The majority of LTCH admissions, 70 percent, come from acute care hospitals, while 14 percent of admissions come from the community. LTCH admissions represent only 1 to 2 percent of Medicare discharges from acute care hospitals.[6,7] Accounting for clinical severity and case-mix is thus important to factor into a measure that estimates the expected cost of the episode to Medicare.

The MSPB-PAC LTCH Measure methodology defines an MSPB-PAC episode as all claims with start dates falling between the treatment period start date (date of admission to LTCH) and the associated services period end date (30-days post LTCH 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."[8] 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; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long-Term Care Hospital Prospective Payment System and Policy Changes and Fiscal Year 2017 Rates. Federal Register, Vol. 81, No. 162. https://www.govinfo.gov/content/pkg/FR-2016-08-22/pdf/2016-18476.pdf

[2] American Hospital Association (AHA). (December, 2015). The role of post-acute care in new care delivery models. Trend Watch. Retrieved from http://www.aha.org/research/reports/tw/15dec-tw-postacute.pdf

[3] 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

[4] Eliason, Paul J., Paul L. E. Grieco, Ryan C. McDevitt, and James W. Roberts. 2018. "Strategic Patient Discharge: The Case of Long-Term Care Hospitals." American Economic Review, 108 (11): 3232-65. doi: 10.1257/aer.20170092.

[5] Einav, Liran, Amy Finkelstein, and Neale Mahoney. "Provider Incentives and Healthcare Costs: Evidence from Long-Term Care Hospitals." Econometrica: Journal Of The Econometric Society 86, no. 6 (2018): 2161-219.

[6] Liu K, Baseggio C, Wissoker D, Maxwell S, Haley J, Long S. Long-term care hospitals under Medicare: facility-level characteristics. Health Care Financ Rev. 2001;23(2):1–18.

[7] Kandilov A, Dalton K. Utilization and payment effects of Medicare referrals to long-term care hospitals: final report. 2011. Final report prepared by Kennell and Associates Inc. and RTI International for Centers for Medicare & Medicaid Services under contract HHSM-500- 2006-00008I

[8] National Quality Forum (NQF). (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 <u>supplemental</u> 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_LTCH_-_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: LTCH admission

Start Date: Date of the LTCH admission

End Date: 30 days after discharge from the LTCH stay

MSPB-PAC LTCH episodes include services that take place during the time period 30 days post-LTCH 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 LTCH admission date and 30 days after LTCH discharge are attributed to the LTCH 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 LTCH episode. Second, the MSPB-PAC LTCH approach incorporates costs due to consequences of care, such as complications, encouraging LTCH 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 LTCH 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 LTCH 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 LTCH 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 LTCH 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 LTCH 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 LTCH episodes because they are clinically unrelated to LTCH care and/or because LTCH 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 LTCH 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 LTCH 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 LTCH 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)*

The MSPB-PAC LTCH measure is stratified by standard and site neutral payment rate admissions. An MSPB-PAC LTCH Standard episode is triggered by a standard payment rate claim, while an MSPB-PAC LTCH Site Neutral episode is triggered by a site neutral payment rate claim. Risk adjustment is then performed separately for MSPB-PAC LTCH Standard and Site Neutral cases. Thus, LTCH Standard and Site Neutral episodes are compared only with LTCH Standard and Site Neutral episodes, respectively, to ensure that the measure is making fair comparisons between clinically similar beneficiaries.

**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 LTCH 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 LTCH measure relies on a detailed price standardization methodology to exclude geographic payment rate differences; in other words, the MSPB-PAC LTCH 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 LTCH measure score of 1 indicates that an LTCH 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 LTCH facilities during a given performance period. An MSPB-PAC LTCH measure score of greater than 1 indicates that an LTCH had higher average risk-adjusted spending levels compared to those of the national median LTCH. For example, a measure score of 1.1 indicates that the LTCH had average risk-adjusted spending levels that are 10 percent higher than the median LTCH. On the other hand, an MSPB-PAC LTCH measure score of less than 1 indicates that an LTCH had lower average risk-adjusted spending levels compared to those of the median LTCH. For example, a measure score of 0.9 indicates that the LTCH had average risk-adjusted spending levels that are 10 percent lower than the median LTCH.

**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 LTCH 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 LTCH model uses a linear regression framework and a 90-day HCC lookback period. The risk adjustment model is estimated on all MSPB-PAC LTCH episodes that meet the exclusion criteria.

The model is estimated separately for Standard and Site Neutral episodes (see section S.7.2 for description of episode types). LTCH episodes are only compared to episodes of the same type (i.e., Standard episodes are only compared to Standard episodes, and Site Neutral episodes to Site Neutral episodes). This ensures that comparisons are fair, meaningful, and reflective of payment policy differences within particular LTCH settings.

Each provider's MSPB-PAC LTCH 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 LTCH 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 LTCH 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
- Medicare Severity-Long-Term Care Diagnosis-Related Groups (MS-LTC-DRGs)

The clinical case-mix category variables used in the MSPB-PAC LTCH 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 LTCH episode. See section S.7.5 for more details on the methodology of assigning clinical case-mix categories to each episode.

Notes:

[1] 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

[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 LTCH measure scores are reported publicly for providers with 20 or more eligible episodes, along with the national average score. The distribution of MSPB-PAC LTCH 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: 17.5%
- Not statistically different from the national average: 60.0%
- Significantly higher than the national average: 22.5%

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

- Minimum: 0.76
- 10th Percentile: 0.90
- 25th Percentile: 0.95
- 75th Percentile: 1.04
- 90th Percentile: 1.09
- Maximum: 1.50

**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 LTCH episode is attributed to the LTCH 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 LTCH episodes allows episodes to overlap with hospital and other MSPB-PAC episodes. MSPB-PAC LTCH episodes may begin within 30 days of discharge from an inpatient hospital discharge as part of a patient's trajectory from an acute to a PAC setting. An MSPB-PAC LTCH 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 LTCH provider discharges a beneficiary who is then admitted to another LTCH within 30 days. The LTCH claim would be included once as an associated service for the attributed provider of the first MSPB-PAC LTCH episode and once as treatment services for the attributed provider of the second MSPB-PAC LTCH 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 LTCH 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, LTCH facilities (and other PAC providers) are required to report data on quality, resource use, and other measures. The MSPB-PAC LTCH measure was created to fulfill the statutory requirement for LTCH facilities to submit measures of resource use for public reporting. As such, LTCH facilities reporting the MSPB-PAC LTCH measure can compare their performance relative to all other Medicare-certified LTCH facilities in the United States.

**S.13.4. Sample size**

Detail the sample size requirements for reporting measure results.

MSPB-PAC LTCH measure scores are publicly reported on LTCH Compare for LTCH facilities with 20 or more eligible episodes. Out of 429 LTCH facilities with FY 2016-2017 episodes, only 7 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 LTCH measure itself is not calculated using benchmarks but rather is a comparison between a given LTCH's MSPB-PAC LTCH Amount and national episode-weighted median MSPB-PAC LTCH 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_LTCH_-_NQF_Testing_Attachment_2020-04-28.docx

**Measure Number** (*if previously endorsed*)**:** Click here to enter NQF number
**Measure Title**: Medicare Spending per Beneficiary Post-Acute Care Measure for Long-Term Care Hospitals
**Date of Submission**: 1/3/2020

**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; LTCH 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 Long-Term Care Hospitals (MSPB-PAC LTHC) 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 long-term care hospitals 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 long-term care hospitals (LTCHs) 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

- **LTCH Compare data**: We used this data source to examine the relationship between MSPB and assessment-based quality measures. The LTCH Compare data include publicly reported LTCH quality measures. The data are available at: https://data.medicare.gov/data /long-term-care-hospital-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 Long-Term Care Hospitals (LTCHs) paid under Medicare's LTCH Prospective Payment System (PPS) and included in the LTCH Quality Reporting Program (QRP) were included, provided they had an eligible episode. A total of 429 LTCHs 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 LTCHs by census region, ownership type, and rurality. Throughout this form, we use "K" to refer to number of providers.

**Table 1. Characteristics of LTCHs included in Specification and Testing of the FY 2016-2017 MSPB-PAC LTCH Measure (K = 429)**

| Characteristic | K (%) |
|---|---|
| **Census Region** | |
| New England | 16 (3.73%) |
| Mid Atlantic | 27 (6.29%) |
| East North Central | 69 (16.08%) |
| West North Central | 27 (6.29%) |
| South Atlantic | 69 (16.08%) |
| East South Central | 37 (8.62%) |
| West South Central | 128 (29.84%) |
| Mountain | 31 (7.23%) |
| Pacific | 25 (5.83%) |
| U.S. Territories | 0 (0.00%) |
| **Ownership type** | |
| Government | 19 (4.43%) |
| Proprietary | 289 (67.37%) |
| Not-For-Profit | 114 (26.57%) |
| Physician Ownership | 3 (0.70%) |
| Hospital District or Authority | 3 (0.70%) |
| Tribal | 1 (0.23%) |

| Characteristic | K (%) |
|---|---|
| **Rurality** | |
| Rural | 22 (5.13%) |
| Urban | 407 (94.87%) |

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

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 LTCH patient episodes with LTCH discharges between October 1, 2015, and September 30, 2017, were included in the measure. For patients with multiple LTCH episodes during the measurement period, all eligible episodes were included. A total of 157,004 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 153,864 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 LTCH Measure (N = 157,004)**

| Characteristic | N (%) | Characteristic | N (%) |
|---|---|---|---|
| **Male** | 79,035 (50.34%) | **Race*** | |
| **Female** | 77,969 (49.66%) | **White** | 112,413 (71.60%) |
| **Male under 65 years** | 26,832 (17.09%) | **Black** | 32,062 (20.42%) |
| **Female under 65 years** | 20,371 (12.97%) | **Other** | 11,529 (7.34%) |

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

*Race information was not available for 1,000 (0.64%) 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 LTCH score calculation removes outliers in the 1st and 99th percentiles of the residual distribution and subsequent testing and analyses were conducted on these 153,864 episodes.

There were a total of 429 LTCHs with FY 2016-2017 episodes. Facility-level performance measure score reliability and validity testing was restricted to the 422 LTCHs 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.

**1.8 What were the social risk factors that were available and analyzed**?

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

We 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 LTCH episodes. Performance measure score reliability testing was restricted to the 422 LTCHs with 20 or more episodes in the FY 2016-2017 measurement period, as only LTCHs 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*.[1] 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 LTCH 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 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)[2] 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 422 LTCHs 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 422 LTCHs from FY 2016-2017, and by sample size quartile. The average reliability score for all agencies was 0.87 and the median was 0.90. When examined by facility size, the average reliability score increased from 0.75 (quartile 1) to 0.94 (quartile 4).

---

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

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

**Table 3. Facility Reliability Score Distribution of the Episode-Level MSPB Risk Adjusted Spending, overall LTCH sample and by sample size quartile, with public reporting exclusions (K = 422)**

| Facility Sample | K | Mean (SD) | 25th Pct* | Median | 75th Pct |
|---|---|---|---|---|---|
| Overall | 422 | 0.87 (0.10) | 0.85 | 0.90 | 0.93 |
| Quartile 1: 22 - 197 episodes | 106 | 0.75 (0.12) | 0.70 | 0.78 | 0.84 |
| Quartile 2: 197 - 288 episodes | 105 | 0.88 (0.04) | 0.87 | 0.89 | 0.91 |
| Quartile 3: 288 - 430 episodes | 106 | 0.91 (0.02) | 0.89 | 0.91 | 0.92 |
| Quartile 4: 430 - 2,521 episodes | 105 | 0.94 (0.03) | 0.93 | 0.95 | 0.96 |

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

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

**Split-sample Reliability Testing Results**. **Table 4** presents ICC(2,1) between the split-sample scores for the overall sample of 422 LTCHs included in this testing, and by sample size quartile. The ICC in the overall sample was 0.86 with a 95% confidence interval (CI) of 0.84 to 0.89. 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 LTCH sample and by sample size quartile, with public reporting exclusions (K = 422)**

| Facility Sample | K | ICC(2,1) (95% CI) |
|---|---|---|
| Overall | 422 | 0.86 (0.84 - 0.89) |
| Quartile 1: 22 - 197 episodes | 106 | 0.86 (0.80 - 0.90) |
| Quartile 2: 198 - 288 episodes | 105 | 0.86 (0.80 - 0.90) |
| Quartile 3: 289 - 429 episodes | 105 | 0.86 (0.80 - 0.90) |
| Quartile 4: 430 - 2,521 episodes | 106 | 0.90 (0.85 - 0.93) |

Analysis of Medicare Claims File for LTCH 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.[3] Reliability among the smallest publicly reported providers (quartile 1) is moderate to good, depending on the analysis used.

[3] 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,"

The reliability score results indicate that the average facility had good reliability. On average, 87 percent of the variation in the risk adjusted MSPB amount was associated with systematic differences between agencies[4], with a range of 75 to 94 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 0.86, with a range of 0.86 to 0.90 (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:

---

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.

[4] 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.

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 LTCH 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 LTCH scores and the Discharge to Community (DTC) rates for FY 2016-2017. The Discharge to Community-Post Acute Care Measure for Long-Term Care Hospitals (DTC-PAC LTCH) is endorsed by NQF (#3480) and is publicly reported as part of the LTCH Quality Reporting Program, and are based on Medicare claims.[5]

   The DTC measure assesses successful discharge to community from an LTCH, with successful discharge to community including no unplanned hospitalizations and no death in the 31 days following discharge. Specifically, this measure reports an LTCH's risk-standardized rate of Medicare FFS patients who are discharged to the community following an LTCH stay, do not have an unplanned admission to an acute care hospital or 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 LTCH scores and several assessment-based quality measure scores publicly reported on the LTCH Compare website for FY 2017. Specifically, we use the following NQF-endorsed measures:

   - Percent of Residents or Patients with Pressure Ulcers That Are New or Worsened (Short-Stay) (#0678)[6]
   - Catheter-associated Urinary Tract Infection (CAUTI) Outcome Measure (#0138)
   - Central line-associated Bloodstream Infection (CLABSI) Outcome Measure (#0139)
   - Facility-wide Inpatient Hospital-onset Clostridium difficile Infection (CDI) Outcome Measure (#1717)

[5] Medicare Program; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long Term Care Hospital Prospective Payment System and Policy Changes and Fiscal Year 2017 Rates. Federal Register, Vol. 81, No. 162. https://www.govinfo.gov/content/pkg/FR-2016-08-22/pdf/2016-18476.pdf

[6] The Pressure Ulcers measure was endorsed by NQF but the endorsement has been removed because the measure has been replaced by the Changes in Skin Integrity Post-Acute Care: Pressure Ulcer/Injury measure.

The Pressure Ulcers measure reports the percent of patients with Stage 2-4 pressure ulcer(s) that are new or worsened since admission. Each of the three infection measures reports standardized infection ratios of healthcare-associated infections. All are calculated using one year of data.

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 these analyses, the relationship depends on whether higher-than-expected rates of pressure ulcers and infections are associated with more adverse outcomes, such as unplanned hospitalizations. To answer this question, we examined the relationship between these measures and the Discharge to Community measure described earlier. We found that Pressure Ulcers and infections measures are not associated with higher rates of adverse outcomes: Pearson correlations between DTC and these measures range between -0.059 and -0.001 (Spearman: -0.093 to -0.024).[7] Accordingly, we hypothesized that the correlations between these measures and MSPB-PAC LTCH should, likewise, be low.

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

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

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 0.94, compared with 1.23 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 1.00, compared to 1.02 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 | 0.94 | 1.20 | 1.32 | 1.41 | 1.50 |
| ER Visits | 1.00 | 1.01 | 1.03 | 1.07 | 1.07 |

Analysis of Medicare Claims File for LTCH 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) | 387 | -0.2063 | <0.001 | -0.2916 | <0.001 |

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

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

[7] See **Appendix Table 2b1.2** for detailed correlation results.

Lastly, we found very small and not statistically significant correlations (both Pearson and Spearman) between MSPB measure scores and Pressure Ulcers and infections measure scores (**Table 7**). These results are consistent with the finding that these assessment-based quality measures are not associated with lower rates of successful discharge to community.

**Table 7. Correlations between MSPB and Functional Improvement Measures**

| Measure Name | K* | Pearson Correlation | p-value | Spearman Correlation | p-value |
|---|---|---|---|---|---|
| Rate of pressure ulcers that are new or worsened | 384 | -0.0927 | 0.0696 | -0.0338 | 0.5087 |
| Catheter-associated urinary tract infections (CAUTI) | 376 | 0.0435 | 0.4003 | 0.0527 | 0.3082 |
| Central line-associated bloodstream infections (CLABSI) | 376 | 0.0074 | 0.8869 | 0.0432 | 0.4030 |
| Clostridium difficile infection (CDI) | 368 | -0.0335 | 0.5218 | 0.0328 | 0.5299 |

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

*Numbers of reflects providers with both MSPB measure scores and scores on the other measure used in the correlation.

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 LTCHs are associated with better discharge to community rates. The low correlation between MSPB and Pressure Ulcers and infections measures is consistent with the finding that these measures are 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, thus, the sum of individual exclusion frequencies exceeds the total number of episodes excluded. Overall, 36.3% of episodes were excluded because of one or more exclusion criteria. 29.9% 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. This high level of insufficient enrollment is due primarily to the relatively high mortality rate among LTCH patients – 60,344 episodes (24.5% of all episodes) were excluded due to patient's death. This exclusion is necessary because the observation window for episodes that end patients' death is likely to be shorter (and include fewer services) than for episodes that do not end in death. Thus, including such patient episodes in a resource use measure would reduce the measure's usability and could introduce perverse incentives for service providers.

9.5% 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, 30.3 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 (median: 29.5%; inter-quartile range: 25.7%-34.0%) and 9.9% 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 (median: 9.6%; inter-quartile range: 7.5%-12.1%).

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

| Exclusion | N* | % |
|---|---|---|
| Any episode that is triggered by a long-term care hospital 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 LTCH 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), or is enrolled in Part C for any part of the lookback period plus episode window | 73,570 | 29.87% |
| 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 | 23,436 | 9.52% |
| Any episode where the claim(s) constituting the attributed LTCH provider'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 | 0.01% |

| | | |
|---|---|---|
| Total Number of Episodes Excluded | 89,280 | 36.25% |

Analysis of Medicare Claims File for LTCH 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 LTCH Measure**

| Exclusion | Mean | 25th Perc. | Median | 75th Perc. |
|---|---|---|---|---|
| Any episode that is triggered by a long-term care hospital 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 LTCH 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 | 30.31% | 25.68% | 29.50% | 34.01% |
| 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 | 9.87% | 7.53% | 9.59% | 12.06% |
| Any episode where the claim(s) constituting the attributed LTCH 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) | 0.01% | 0.00% | 0.00% | 0.00% |

Analysis of Medicare Claims File for LTCH 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.* <u>*Note*</u>*: **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 LTCH Measure**

| Exclusion | Rationale |
|---|---|
| Any episode that is triggered by a long-term care hospital 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 LTCH provider'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 LTCH measure 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 LTCH provider'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 overlap wholly or in part, or are otherwise erroneous or contradictory). | The episode with the most recent processing date is kept to ensure the accuracy of data elements. |

---

## 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** 232 **risk factors**

☒ **Stratification by** 2 **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.

Risk adjustment is performed separately for the MSPB-PAC LTCH episode types listed below:

- LTCH Standard
- LTCH Site Neutral

In calculating expected spending as part of the MSPB-PAC LTCH measure calculation, LTCH episodes are only compared to episodes of the same type (i.e., Standard episodes will only be compared to Standard episodes, and Site Neutral episodes to Site Neutral episodes). This ensures that comparisons are fair, meaningful, and reflective of payment policy differences within particular LTCH settings. The MSPB-PAC LTCH measure allows different types of claims to trigger different episodes, reflecting differences in payment policy and beneficiaries' underlying health characteristics. In the LTCH setting, the dual payment rate structure distinguishes between standard payment rate cases and site neutral payment rate cases. A standard payment rate case is one that is not a psychiatric or rehabilitation Medicare Severity Long-Term Care Diagnosis-Related Group (MS-LTC-DRG), is immediately preceded by an acute care hospital stay, and either (i) the acute care hospital stay included at least 3 days in intensive care unit (ICU) or coronary care unit (CCU) or (ii) the beneficiary received 96+ hours of ventilator services. A site neutral payment rate case is one that does not meet the definition of a standard payment rate case. A standard payment rate case triggers an MSPB-PAC LTCH Standard episode, while a site neutral payment rate case triggers an MSPB-PAC LTCH Site Neutral episode. LTCH Standard and Site Neutral episodes are compared only with LTCH Standard and Site Neutral episodes, respectively, to ensure that the measure is making fair comparisons between clinically similar beneficiaries.

The MSPB-PAC LTCH 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.[8,9] The MSPB-PAC LTCH models use a linear regression framework and a 90-day HCC lookback period. The following beneficiary health status indicators are included as covariates in each MSPB-PAC LTCH risk adjustment model and to the greatest extent possible are consistent across PAC settings (see Appendix C of

[8] 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

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

the Measure Specifications[10] 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
- Original entitlement to Medicare through disability
- ESRD status
- Long-term care institutionalization at start of episode[11]
- 6 clinical case mix categories reflecting recent prior care (described further below)[12]
- Hospice utilization during the episode
- Prior acute ICU utilization day categories
- Prior acute length of stay categories
- Medicare Severity-Long-Term Care Diagnosis-Related Groups (MS-LTC-DRGs)

Clinical case mix categories are also included in the MSPB-PAC LTCH 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 LTCH 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 LTCH 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

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

[11] 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.

[12] There are 7 case mix categories, but one category is removed to prevent collinearity.

(5) **Prior PAC - Institutional**[13] – beneficiaries who are continuing PAC from an institutional PAC setting (i.e., coming from an LTCH, IRF, or SNF)

(6) **Prior PAC - HHA**[14] – beneficiaries who are continuing PAC from a HHA

(7) **Community**[15] – 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. Different logic is used to handle LTCH Standard episodes with multiple prior claims sharing the same end date. Given that LTCH Standard episodes are defined by the presence of a prior acute IP stay between 0 to 1 days prior to the start of the LTCH Standard episode, only information about this required prior hospitalization is used to assign the episode's clinical case mix category. Given the role of ICU days in determining eligibility for LTCH Standard payment rates, the clinical case mix category is assigned based on the inpatient stay with the most ICU days. In the event of a tie in the number of ICU days, the clinical case mix category is assigned based on the IP claim with the longer length of stay. Should a tie still persist, the most recent IP claim by discharge date is used. Finally, if the prior criteria do not result in a category assignment, the original hierarchy above is used.

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

[13] Prior PAC and Community variables are only used in the MSPB-PAC LTCH Site Neutral risk adjustment model. Since LTCH Standard episodes are defined by having a prior IP stay, the MSPB-PAC LTCH Standard risk adjustment model omits variables representing non-IP sources of entry in the clinical case-mix categories.

[14] See previous footnote.

[15] See previous footnote.

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 MSPB-PAC LTCH measure methodology.[16]

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

The MSPB-PAC LTCH 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. 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 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 impacts 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

[16] 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.

[17] Hanchate, Amresh, et al. "Racial and Ethnic Differences in End-of-Life Costs: Why do Minorities Cost More than Whites?" Archives of Internal Medicine. 2009; 169(5):493-504.

[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

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)**

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

The MSPB-PAC LTCH 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 LTCH, we also conducted additional analyses. We also 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 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.**

---

[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

First, we examined the frequency of patients with each social risk factor (SRF). Overall, 45% of LTCH episodes in the FY 2016-2017 period involved patients who were dual-eligible, 28% involved patients who were not White, and 5% involved patients in rural locations (see **Appendix Table 2b3.4b_1**). Standard and Site Neutral episodes had largely similar characteristics.

Second, we compared average observed per-episode spending across patients with each SRF. Overall, we found that spending is lower for patients who are White and patients who live in suburban areas. Spending is highest among patients who reside in areas with average SES in the fourth quartile of the AHRQ SES Index (see **Appendix Table 2b3.4b_2**). Spending is considerably higher in Standard than Site Neutral episodes, consistent with their definitions.

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 in most cases, when added to the model individually as well as when added together with all other SRFs; this is expected given the large sample size.[22] 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 SRFs to the risk adjustment model, we examined the change in individual LTCH's MSPB-PAC LTCH measure scores when computed with and without SRFs. The results are highly correlated: correlation between baseline scores and scores adjusted for all SRFs is 0.997 (Pearson; Spearman rank correlation is 0.996) (**Appendix Figure2b3.4b_1 and Figure2b3.4b_2**). The average absolute change in provider scores is 0.006 (**Appendix Table 2b3.4b_4a**); 91% of LTCH scores change by ±0.01, i.e., less than 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.

### 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 LTCH-level observed and risk-adjusted episode cost.

[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 with Medicaid, the coefficient on the Dual eligibility with Medicaid indicator variable is negative. This is due to collinearity between the social risk factors and the clinical factors already included in the model.

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

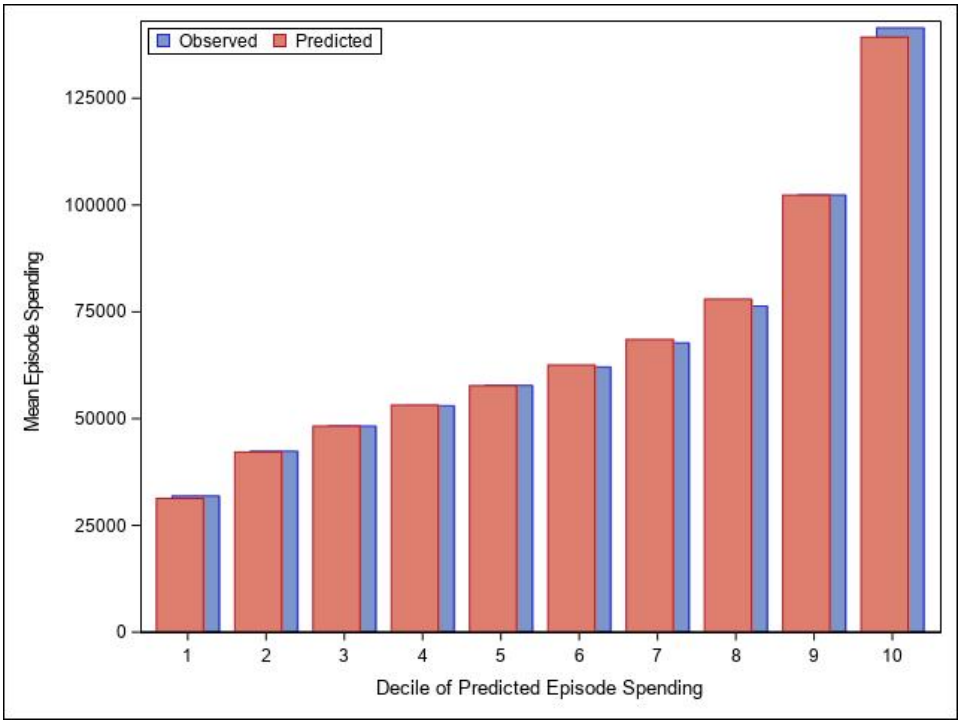
2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*):

The overall adjusted R-squared is 0.4894. The overall adjusted R-squared was computed by fully interacting episode type (Standard vs Site Neutral) with all other coefficients. This shows the combined explanatory power of the strata (episode types) and the covariates. The adjusted R-squared for Standard episodes is 0.341 and 0.4697 for Site Neutral episodes. **Appendix Table 2b3.6_1** also includes regression coefficients, standard errors, and p-values for each of the covariates used in the risk adjustment models.

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

The episode-level predicted costs range from $32,267 to $115,906, 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 2 percentage point or less. The ratio of observed to predicted rates is close to 1 across risk deciles, with the smallest ratio being 0.98 and the largest ratio being 1.02.

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




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

**Table 11. LTCH 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 | 15,382 | 31,878 | 31,330 | -548 | 1.02 |
| 2 | 15,381 | 42,354 | 42,164 | -190 | 1.00 |
| 3 | 15,381 | 48,250 | 48,240 | -10 | 1.00 |
| 4 | 15,381 | 53,020 | 53,145 | 125 | 1.00 |
| 5 | 15,381 | 57,764 | 57,709 | -54 | 1.00 |
| 6 | 15,381 | 62,097 | 62,552 | 455 | 0.99 |
| 7 | 15,381 | 67,737 | 68,523 | 787 | 0.99 |
| 8 | 15,381 | 76,330 | 77,966 | 1,636 | 0.98 |
| 9 | 15,381 | 102,365 | 102,331 | -34 | 1.00 |
| 10 | 15,381 | 141,498 | 139,329 | -2,169 | 1.02 |

Analysis of Medicare Claims File for LTCH 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 0.489. The model controls for over 100 comorbidities (including comorbid interactions), case mix categories, and patient risk factors, and is stratified by payment-related episode types (Standard/Site Neutral). Extensive clinical review was performed by clinicians with experience providing care in LTCH 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 LTCH 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 LTCH 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 LTCH providers. If these services had been included in the LTCH 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. Doing so, however, would have created an inferior measure, as it would lack clinical validity.

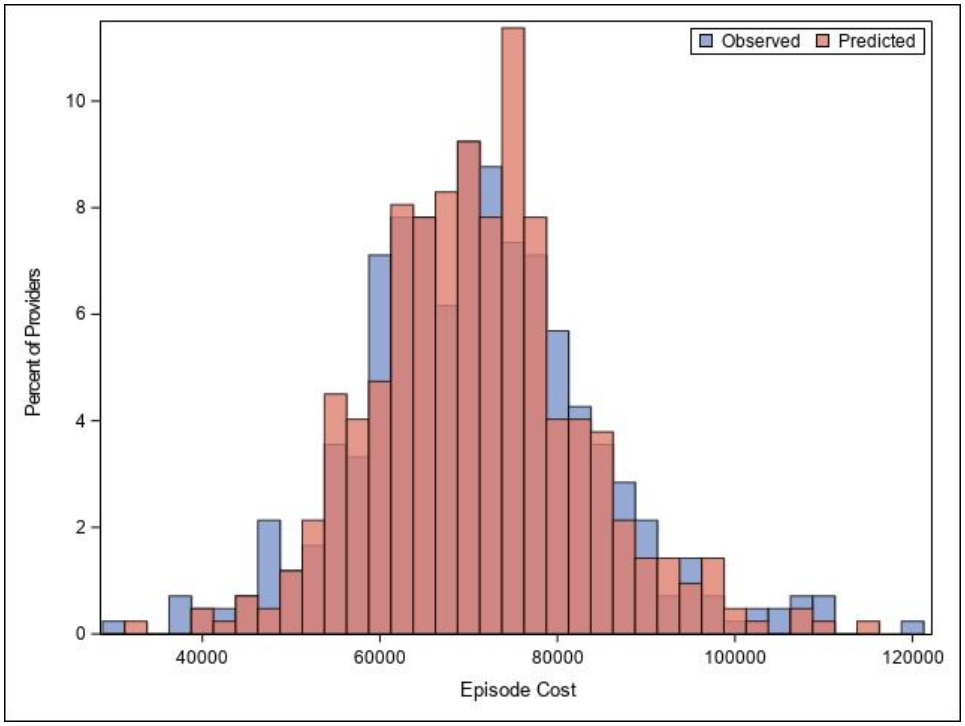
The distribution of facility-level observed and risk-adjusted spending is shown in **Table 12** and **Figure 2.** By taking into account beneficiary characteristics that are outside the provider's control, the model compresses the distribution of provider-level spending and decreases their variability.

**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 | 422 | 70,880 | 13,161 | 55,576 | 62,606 | 70,759 | 78,328 | 86,636 |
| Predicted | 422 | 70,969 | 11,799 | 56,287 | 63,380 | 70,928 | 77,109 | 85,680 |

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

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




Analysis of Medicare Claims File for LTCH 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 LTCH measure scores:

1. We analyzed the distribution of measure scores across all LTCHs 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 $10^{th}$, $25^{th}$, $75^{th}$, and $90^{th}$ percentiles.
2. We used bootstrapping procedures to derive a confidence interval to determine if a LTCH's score is significantly lower than, significantly higher than, or no different than the national average. We bootstrapped samples of LTCHs, 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. LTCHs whose 95% CI was entirely below the national average were considered to be significantly lower than the national average; LTCHs whose CI was entirely above the national average were significantly higher than the national average; and LTCHs 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 LTCH 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 1.97, 1.21, and 1.09, respectively.

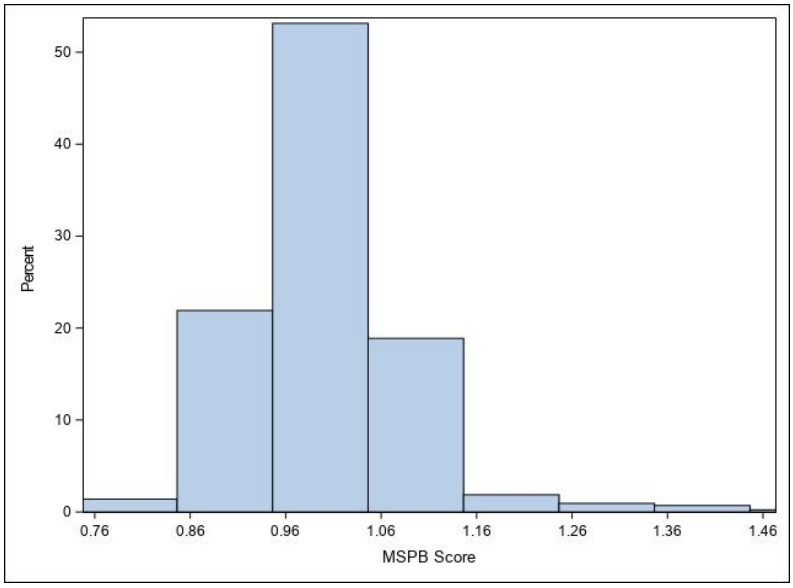
Due to the high level of reliability of the MSPB-PAC LTCH scores, demonstrated in **section 2a2**, small differences in scores can be interpreted as meaningful. This is confirmed by our analysis of statistical significance: 19% of LTCHs had scores that were statistically significantly higher than the national mean, while 20% of LTCHs had scores that were statistically significantly lower (**Table 14**).

**Table 13. Distribution of MSPB-PAC LTCH Scores**

| K | Mean | Standard Deviation | Min | 10th Pct | 25th Pct | 75th Pct | 90th Pct | Max |
|---|---|---|---|---|---|---|---|---|
| 422 | 1.00 | 0.08 | 0.76 | 0.90 | 0.95 | 1.04 | 1.09 | 1.50 |

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

**Figure 3. Distribution of MSPB-PAC LTCH Scores**




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

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

| LTCH Total | Statistically significantly lower than national mean | | Not statistically significantly different from national mean | | Statistically significantly higher than national mean | |
|---|---|---|---|---|---|---|
| K | K | % | K | % | K | % |
| 422 | 74 | 17.5 | 253 | 60.0 | 95 | 22.5 |

Analysis of Medicare Claims File for LTCH 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 LTCH measure is able to identify statistically significant and meaningful differences in performance across LTCHs due to its good reliability and variability. Measure scores range from 0.76 to 1.50, indicating that the model can predict both low and high spending and that there are meaningful differences in facility-level spending. Forty percent of LTCHs 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. 0.01% 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. <u>If this is an eMeasure</u>,** 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 <u>and</u> Planned Use**

| Specific Plan for Use | Current Use (for current use provide URL) |
|---|---|

| | |
|---|---|
| Quality Improvement (Internal to the specific organization) | Public Reporting<br>Long-Term Care Hospital Facilities Quality Reporting Program<br>https://www.medicare.gov/longtermcarehospitalcompare/ |

**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' Long-Term Care Hospital 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 LTCHs. More information about the LTCH QRP can be found at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/LTCH-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 LTCH scores, from the LTCH QRP are publicly reported on the LTCH Compare website at https://www.medicare.gov/longtermcarehospitalcompare/. LTCH quality measure data are also available for download for providers, researchers, and other public at https://data.medicare.gov/data/long-term-care-hospital-compare.
Geographic area and number and percentage of accountable entities and patients included:
The LTCH QRP includes all LTCHs paid under the LTCH PPS. MSPB-PAC LTCH 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 LTCH measure results presented in this submission are based on 429 LTCHs and 157,004 patient episodes; of these, 422 LTCHs 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 LTCH measure were provided to all active LTCH providers under the LTCH 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 LTCH 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 LTCH measure specifications were publicly posted along with the FY 2017 LTCH 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-2016-08-22/pdf/2016-18476.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 LTCH score.

Publicly available data and provider preview reports include the following: reporting period, number of eligible episodes, provider MSPB-PAC LTCH score, and the national average MSPB-PAC LTCH 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 LTCH 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 LTCH 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 LTCH 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 LTCH PPS final rule with public comments and responses can be found at https://www.govinfo.gov/content/pkg/FR-2016-08-22/pdf/2016-18476.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 LTCH 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 LTCH 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 to risk adjust for prolonged ventilator use, a predictor of resource use in LTCHs. However, our model currently includes adjustments for MS-LTC-DRGs that capture ventilator use, making additional adjustments unnecessary. As such, we did not risk adjust for prolonged ventilator use. Additionally, it was suggested to risk adjust for multiple organ failures, also a predictor of resource use. Our testing showed that this condition resulted in a small and not statistically significant impact on predicted spending in LTCHs and inconsistent impacts across other PAC settings. Moreover, our risk adjustment model includes HCCs that capture the most frequent sources of multiple organ failure. For these reasons, we did not implement additional controls for multiple organ failures in our risk adjustment models.

[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/LTCH-Quality-Reporting/Downloads/Copy-of-2016_04_06_mspb_pac_ltch_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:**