NATIONAL QUALITY FORUM

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 2882

Corresponding Measures:

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

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

De.3. Brief Description of Measure: This measure assesses days spent in acute care within 30 days of discharge from an inpatient hospitalization for pneumonia, including aspiration pneumonia or for sepsis (not severe sepsis) with a secondary discharge diagnosis of pneumonia coded in the claim as present on admission (POA) and no secondary diagnosis of severe sepsis coded as POA. This measure is intended to capture the quality-of-care transitions provided to discharge patients hospitalized for an eligible pneumonia condition by collectively measuring a set of adverse acute care outcomes that can occur post-discharge: emergency department (ED) visits, observation stays, and unplanned readmissions at any time during the 30 days post-discharge. In order to aggregate all three events, we measure each in terms of days. The Centers for Medicare & Medicaid Services (CMS) annually reports the measure for patients who are 65 years or older, are enrolled in Medicare fee-for-service (FFS) and are hospitalized in non-federal short-term acute care hospitals.

1b.1. Developer Rationale: The goal of this measure is to improve patient outcomes. Measurement of patient outcomes allows for a broad view of quality of care that encompasses more than what can be captured by individual process-of-care measures. Safely transitioning patients from hospital to home requires a complex series of tasks which would be cumbersome to capture individually as process measures: timely and effective communication between providers, prevention of and response to complications, patient education about post-discharge care and self-management, timely follow-up, and more. Suboptimal transitions contribute to a variety of adverse events post-discharge, including ED evaluation, need for observation, and readmission.

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

S.4. Numerator Statement: The outcome of the measure is a count of the number of days the patient spends in acute care within 30 days of discharge from an eligible index hospitalization with a principal diagnosis of pneumonia, including aspiration pneumonia or a principal diagnosis of sepsis (not severe sepsis) with a

secondary diagnosis of pneumonia (including aspiration pneumonia) coded as POA and no secondary diagnosis of severe sepsis coded as POA. We define days in acute care as days spent in an ED, admitted to an observation unit, or admitted as an unplanned readmission for any cause within 30 days from the date of discharge from the index pneumonia hospitalization.

Additional details are provided in S.5 Numerator Details.

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

The cohort includes admissions for patients discharged from the hospital with a principal diagnosis of pneumonia, including aspiration pneumonia or a principal diagnosis of sepsis (not severe sepsis) with a secondary diagnosis of pneumonia (including aspiration pneumonia) coded as POA and no secondary diagnosis of severe sepsis coded as POA and with continuous 12 months Medicare enrollment prior to admission. CMS publicly reports the measure for those patients 65 years and older who are Medicare FFS or VA beneficiaries admitted to non-federal or VA hospitals, respectively.

Additional details are provided in S.7 Denominator Details.

S.8. Denominator Exclusions: The measure excludes index hospitalizations that meet any of the following exclusion criteria:

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

De.1. Measure Type: Outcome

S.17. Data Source: Claims, Enrollment Data

S.20. Level of Analysis: Facility

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

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? This measure is not formally paired with any measure; however, it is harmonized with a measure of hospital-level, all-cause, 30-day, risk-standardized readmission following pneumonia hospitalization.

Preliminary Analysis: Maintenance Measure

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

Criteria 1: Importance to Measure and Report

1a. Evidence

Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a health outcome measure include providing empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service; if these data not available, data demonstrating wide variation in performance, assuming the data are from a robust number of providers and results are not subject to systematic bias. For measures derived from patient report, evidence also should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.

Summary of prior review in 2016

- This measure calculates excess days in acute care (EDAC) for patients with pneumonia. This measure is intended to capture the quality-of-care transitions provided to discharged patients hospitalized with pneumonia by collectively measuring a set of adverse acute care outcomes that can occur post-discharge: emergency department (ED) visits, observation stays, and unplanned readmissions at any time during the 30 days post-discharge. In order to aggregate all three events, this measure assesses each in terms of days.
- The developer previously cited that "in the context of the Centers for Medicare and Medicaid Services' (CMS') publicly reported readmission measures, the increasing use of ED visits and observation stays has raised concerns that current readmission measures do not capture the full range of unplanned acute care in the post-discharge period (Vashi et al., 2013; Rising et al., 2012; Feng et al., 2012). Observation stays can occur in many different parts of the hospital, including dedicated treatment rooms, the ED, or inpatient units. In particular, there is concern that high use of observation stays could in some cases replace readmissions, and that hospitals with high rates of observation stays in the post-discharge period may therefore have low readmission rates that do not accurately reflect the quality of care (Vashi et al., 2013)."

Changes to evidence from last review

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

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

Updates:

- The developer cites new evidence indicating that pneumonia leads to more than 1 million hospitalizations per year, incurring billions of dollars in healthcare costs. Furthermore, the developer cites that pneumonia is the third leading cause of rehospitalization, accounting for more than 88,800 readmissions at a cost of $1.1 billion in total costs.
- The developer provides a logic model with additional supportive evidence suggesting that hospitals can influence EDAC through a broad range of clinical activities including communication between providers, patient education, prevention of, and response to complications, patient safety, medication reconciliation, better disease management strategies, and coordinated transitions to the outpatient environment.

Question for the Committee:

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

Guidance from the Evidence Algorithm

BOX 1 – Yes à BOX 2 – Yes à PASS

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

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

Maintenance measures – increased emphasis on gap and variation

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

- The developer provides performance data across the three most recent three most recent three-year reporting periods (2016-2019, 2015-2018, and 2014-2017) for all hospitals.
    - The developer reports that the most recent reporting period (2016-2019), for all hospitals, was -65.7 to 178.5 EDAC per 100 admissions; the mean was 4.5 EDAC per 100 admissions and the median was 1.3 EDAC per 100 admissions.
- The developer provides performance data across the three most recent three most recent three-year reporting periods (2016-2019, 2015-2018, and 2014-2017) for hospitals with at least 25 admissions.
    - The developer reports that the most recent reporting period (2016-2019) was -65.7 to 146 EDAC per 100 admissions; the mean was 5.0 EDAC per 100 admissions and the median was 2.9 EDAC per 100 admissions. The 10th percentile was -23.8, the 50th percentile was 2.9, and the 90th percentile was 36.7 EDAC per 100 admissions.

Disparities

- The developer provides performance scores from July 2016 – Jun 2019 for hospitals with at least 25 cases, serving a low proportion of dual eligible patients vs. those serving a high proportion of dual eligible patients and performance scores for hospitals by proportion of patients with Agency for Healthcare Research and Quality (AHRQ) Socioeconomic status (SES) Index scores.
- Low-proportion hospitals are those hospitals whose population of dual-eligible patients or AHRQ SES Index score is small enough to place them at or below the 25th percentile among all hospitals; and high proportion are those hospitals whose population of dual eligible patients or AHRQ SES Index scores is large enough to place them at or above the 75th percentile among all hospitals.
    - Distribution of 30-day Pneumonia EDAC by Proportion of Dual Eligible Patients:
    - Social Risk Proportion(%)//q1:(0-14.4%)//q4:(30.2-97.3%)
    - # of Hospitals//1052//1052
        - Maximum//146.0//125.3
        - 90th percentile//31.1//53.1
        - 75th percentile//16.8//31.2
        - Median//2.9//8.3
        - 25th percentile//-11.7//-11.3
        - 10th percentile//-23.6//-23.9
        - Minimum//-65.7//-52.3

    - Distribution of 30-day Pneumonia EDAC by Proportion of Patients with AHRQ SES Index Scores:
    - Social Risk Proportion within hospital(%)//q1:(0-7.5%)//q4:(33.5-100%)
    - # of Hospitals//1044//1040

- Maximum//87.9//123.4
- 90th percentile//25.6//44.7
- 75th percentile//10.4//22.4
- Median//-3.4//6.0
- 25th percentile//-17.5//-11.9
- 10th percentile//-26.4//-23.1
- Minimum//-49.7//-52.2

Questions for the Committee:

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

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

Committee Pre-evaluation Comments:

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

1a. Evidence to Support Measure Focus: For all measures (structure, process, outcome, patient-reported structure/process), empirical data are required. How does the evidence relate to the specific structure, process, or outcome being measured? Does it apply directly or is it tangential? How does the structure, process, or outcome relate to desired outcomes? For maintenance measures –are you aware of any new studies/information that changes the evidence base for this measure that has not been cited in the submission? For measures derived from a patient report: Measures derived from a patient report must demonstrate that the target population values the measured outcome, process, or structure."

- Readmissions/ED visits/Observation days are negative outcomes to care.
- Responding to the question posed by NQF staff, incorporating data from their logic model, if feasible, could change results.
- Developer cited that there are more than 1 million hospitalizations per year for pneumonia and pneumonia is the third leading cause of readmissions. I am not aware of any new studies or information that changes the evidence for this measure.
- No concerns
- Demonstrated

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

- As measured, from below expected to above expected, there is a wide variation in performance. Data not presented on whether these are stable across years, which would provide additional evidence that differences are persistent rather than random within hospitals.
- A measure specifically built to look at the performance of underserved populations might be revealing.
- Yes. Performance data submitted of the three most recent three-year periods showed variation in performance
- No concerns
- demonstrated

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

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

Reliability

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

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers. For maintenance measures – less emphasis if no new testing data provided.

Validity

2b2. Validity testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For maintenance measures – less emphasis if no new testing data provided.

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

Composite measures only:

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

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

Evaluators: NQF Scientific Methods Panel

Methods Panel Review (Combined)

Methods Panel Evaluation Summary:

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

This measure was reviewed by the Scientific Methods Panel and discussed on the call. A summary of the measure and the Panel discussion is provided below.

Reliability

- The developer conducted reliability testing at the measure score-level.
- The developer estimated measure reliability using intra-class correlation coefficient (ICC). This statistic can be used to assess the correlation and agreement between measurements.
- The developer used a split-sample approach in which hospital performance is measured once using a random subset of patients, and then measured again using a second random subset exclusive of the first, and the agreement of the two resulting performance measures compared across hospitals
- Using the split-sample approach, the developer reported ICC ranges from 0.541 for hospitals with at least two admissions to 0.709 for hospitals with at least 300 admissions. For hospitals with at least 25 admissions, the ICC was 0.576.
- The SMP did not raise any major concerns with reliability and passed the measure on this criterion.

Validity

- The developer conduced face validity and empirical validity testing at the measure score level.
- Face validity was assessed using survey-based information provided by the 16-member technical expert panel.
    - >80% of experts moderately or strongly agreed with the validity of the measure.
- Construct validity was assessed as the relationships between the pneumonia (PN) EDAC measure score and the risk standardized readmission rate (RSRR) group scores, the overall hospital rating scores, and the PN readmission measure.
    - The developer posited a negative relationship between the PN EDAC scores, star-rating readmission score group and star-rating summary scores. They also hypothesized a positive relationship between the pneumonia EDAC scores and the pneumonia readmission measure scores.
        - Correlation with Hospital Star Rating readmission group score: -0.416 (p<.0001),
        - Correlation with Overall Hospital Star Rating summary score: -0.398 (p<.0001),
        - Correlation with PN Readmission Measure: 0.625 (p<.0001).
- Regarding risk adjustment, the developer found a c-statistic of 0.62 and an R2 value of 0.038. Two social risk factors were tested and were found to be statistically significant, but do not appear to meaningfully affect hospital performance estimates and were therefore not included.
- Some SMP members raised concerns with the c-statistic, noting that it was low. Additionally, the SMP raised concern with the choice of variables for testing construct validity, as the approach examined correlated of the EDAC measures with other measures that include the same readmission events, without correction for the overlap.
- The developer responded by conducting additional testing to address the SMP's concerns with respect to overlapping readmission events and found similar results and in the expected direction.
- The SMP acknowledged the developer's response, including the additional validity testing and passed the measure on the validity criterion.

Questions for the Committee regarding reliability:

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

Questions for the Committee regarding validity:

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

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

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

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

- Claims based measure. Readmission, ED and observation status clearly defined.
- Measure specifications appear adequate.
- No concerns. Split sample reliability used. ICC 0.576 for at least 25 admissions, 0.668 for 100+ admissions and 0.709 for 300+ admissions
- No concerns
- no concerns

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

- Reliability appears acceptable at higher volumes.
- It concerns me that not more of the SMP rated the reliability high.
- No concerns. Moderate reliability
- No concerns
- None

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

- Issue of correlation with measures of readmission. Would like to see correlation with other excess day measures, since the same mechanisms are hypothesized to influence outcomes across diseases. Also would like to see correlation of three components of measure. Based on appendix provided, approximately 10% of post-discharge days are ED or observation, so measure is dominated by readmission. Figure 5 in appendix, which has a sample of hospitals ranked by overall score and shows rankings and scores across components shows high correlation of overall score with readmission rates and days, but looks like virtually no correlation with ED or obs days, so inclusion of those seems to just add noise to the measure.
- No.
- No. Face validity of expert panel surveyed showed that 11/12 agreed with validity statement. Empiric validity was done by correlation of this measure with 2 other measures
- No concerns
- None

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

- C-stat on any use moderate. Ability of Poisson regression to estimate days given any days is low. Not clear that this is a more reliable measure than the readmission measure.
- b2. Yes. b3. Social risk factors associate but make no difference in the model outcome. I find that curious. The SMP gave no high ratings. I think the committee should discuss validity.
- SRF analysis for dual eligible status and AHRQ SES index done but no risk adjustment included by developer.
- No concerns
- no concerns

2b4-6. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data)2b4.Identification of Statistically Significant and Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality? 2b5. Comparability of Performance Scores when more than One Set of Specifications: If multiple sets of specifications: Do analyses indicate they produce comparable results? 2b6. Missing Data Analysis and Minimizing Bias/no response: Does missing data constitute a threat to the validity of this measure?

- Missing data does not constitute a threat to validity.
- b4: Primarily through face validity and correlation with other ratings. b6. No threat .5.
- No missing data reported
- No concerns
- appears data is consistent, no concerns

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

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

- ALL data elements are in defined fields in a combination of electronic sources.

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

Questions for the Committee:

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

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

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

- claims based measure. no issues.
- No concerns about the data strategy. Concerns about evaluation of social risk factors for risk adjustment, though these are not available electronically or readily.
- No concerns. No data collection burden to hospitals are providers since electronic sources using administrative claims and enrollment data
- No concerns
- no concerns

Criterion 4: Usability and Use

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

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

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

4a.1. Accountability and Transparency. Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

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

Accountability program details

- Public Reporting - Care Compare
    - https://www.medicare.gov/care-compare/

- Payment Program - Hospital Inpatient Quality Reporting (IQR) Program
    - https://qualitynet.cms.gov/inpatient/iqr

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

Feedback on the measure by those being measured or others

- The developer notes that each hospital generally receives their measure results in April/May of each calendar year through CMS's QualityNet website. The results are then publicly reported on CMS's public reporting websites in the summer of each calendar year.
- The developer adds that since the measure is risk-standardized using data from all hospitals, hospitals cannot independently calculate their score. However, CMS Hospital-Specific Reports with details about every patient from their facility that was included in the measure calculation.
- Lastly, the developer states that Hospital-Specific Reports (HSRs) also provide hospitals with more detailed benchmarks with which to gauge their performance relative to peer hospitals and interpret their results, including comorbidity frequencies for their patients relative to other hospitals in their state and the country.
- Hospitals have access to other resources that are updated in the Spring of each year and are publicly reported. These include HSR guides, tutorial videos, FAQs, annual updates and specification reports, SAS code, measure fact sheets, and other resources.

Additional Feedback:

- Accountable entities and other stakeholders can submit questions via an online portal, in which experts on measure specifications and/or implementation respond to those inquiries.
- Additionally, the developer routinely scans the literature for articles describing research related to this measure.
- The developer has received feedback and inquiries related to the overlap of EDAC and other readmissions measures, interpretation of measure results and performance categories, specifications, performance period, etc.
- Every year, the feedback and literature are considered by technical and clinical experts. The developer states that any issues that warrant additional analytic work due to potential changes in the measure specifications are addressed as a part of annual measure reevaluation.

***Questions for the Committee**:*

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

Preliminary rating for Use: ☒ **Pass** ☐ **No Pass**

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

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

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

Improvement results

- The developer reports that they do not see improvement in the pneumonia EDAC measure across the three performance periods. The posit that a "contributing factor could be the particularly severe 2017-2018 influenza season, (that would have impacted the 2015-2018 and 2016-2019 reporting periods but not the 2014-2017 reporting period), which resulted in an increase in ED visits and inpatient admissions across all age groups, including ages 65 and over."
- Periods//YEAR1619//YEAR1518//YEAR1417
- Number of Hospitals//4082//4126//4175
- Number of Admissions//1326837//1349047//1380472
    - Mean (SD)//4.9(24.7)//4.7(24.1)//4.6(24.8)
    - Range (Min to Max)//-52.3 to 146//-56 to 166.1//-57.8 to 148.9
    - Minimum//-52.3//-56.0//-57.8
    - 10th percentile//-23.8//-23.5//-24.0
    - 20th percentile//-15.8//-15.3//-16.1
    - 30th percentile//-9.4//-9.5//-9.7
    - 40th percentile//-3.2//-3.5//-4.0
    - Median//2.8//2.3//1.6
    - 60th percentile//8.6//8.0//7.8
    - 70th percentile//14.6//14.4//14.3
    - 80th percentile//23.5//23.0//22.6
    - 90th percentile//36.3//35.8//36.2
    - Maximum//146.0//166.1//148.9

4b2. Benefits vs. harms. Benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

Unexpected findings (positive or negative) during implementation

- The developer reports that they did not identify any unintended consequences during measure development or model testing and monitor the measure's use on an ongoing basis for any unintended consequences.

Potential harms

- None reported

Additional Feedback:

- N/A

***Questions for the Committee**:*

- *The Standing Committee should review the measure results over time and consider the rationale provided by the developer for the lack of improvement.*
- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*

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

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

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

- Feedback mechanism described in documentation. Usable by hospital, perhaps (would like some direct commentary from hospitals on how they use the reports). Potential patient use of measure is minimal.

- Yes to above. Some hospitals have used their data to improve care transitions. Feedback from hospitals is sought, though The methodology esp. risk adjustment may not be useful at the hospital level.

- This measure is publicly reported on Care Compare, CMS. Payment program: CMS Hospital inpatient quality reporting Program. Hospitals receive measure results in April/May each year through CMS QualityNet website and the results are reported on CMS websites in the summer. Developer responded to SMP members concerns about validity. They retested or provided supporting evidence for their methodology

- No concerns
- measurement looks at transition and impact on preventing readmission

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

- No obvious harms.

- The measure has yet to show substantive impact. An emphasis should be placed on the logic model components.

- unintended consequences reported.

- No improvement seen in the pneumonia EDAC across three reporting periods. The developer contributes the severe infuenza season that impacted the 2015-2018 and 2016-2019 reporting periods. No unintended negative consequences noted.

- No concerns

- no concerns

Criterion 5: Related and Competing Measures

Related

- 0229: Hospital 30-Day, All-Cause, Risk-Standardized Mortality Rate (RSMR) Following Heart Failure (HF) Hospitalization
- 0230: Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute myocardial infarction (AMI) hospitalization
- 0330: Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization
- 0505: Hospital 30-day all-cause risk-standardized readmission rate (RSRR) following acute myocardial infarction (AMI) hospitalization.
- 0506: Hospital 30-day, All-Cause, Risk-Standardized Readmission Rate (RSRR) Following Pneumonia Hospitalization
- 1551: Hospital-level 30-day risk-standardized readmission rate (RSRR) following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA)
- 1789: Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)
- 1891: Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization
- 2515: Hospital 30-day, all-cause, unplanned, risk-standardized readmission rate (RSRR) following coronary artery bypass graft (CABG) surgery
- 2880: Excess days in acute care (EDAC) after hospitalization for heart failure (HF)
- 2881: Excess days in acute care (EDAC) after hospitalization for acute myocardial infarction (AMI)

Harmonization

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

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

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

- Correlation with PNE readmission measure is about 0.6. Not clear which measure should be preferred. That would require more analysis than provided.
- Related measures are for other disease-specific entities.
- Developer stated they completely harmonized the cohort definition and risk-adjustment strategy with those of the existing CMS 30-day Pneumonia readmission measure. Developer noted differences in these two measures
- No concerns
- none known

Combined Methods Panel Scientific Acceptability Evaluation

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 2882

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

RELIABILITY: SPECIFICATIONS

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

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

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

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

   Panel Member 3: I am confused about the impact of excluding MI admissions within 30 days of a priori pneumonia admission.

   Panel Member 4: No concerns.

   Panel Member 6: No major concerns.

   Panel Member 7: none

   Panel Member 9: None

RELIABILITY: TESTING

Type of measure:

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

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

Data Source:

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

Level of Analysis:

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

Measure is:

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

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

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

   ☐ **Yes** ☐ **No**
6. **Assess the method(s) used for reliability testing**

   Submission document: Testing attachment, section 2a2.2

   Panel Member 1: Split-sample ICC estimation

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

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

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

   Panel Member 5: split-sample reliability testing: 0.56 for >= 25 admissions and 0.69 for >= 200 admissions

   Panel Member 6: No major concerns.

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

   Panel Member 8: Random half/half split sample test re-test ICC

   Panel Member 9: Methods were appropriate.
7. **Assess the results of reliability testing**

   Submission document: Testing attachment, section 2a2.3

   Panel Member 1: ICC = 0.71 with at least 300 discharges. However, only one-third of hospitals had at least 300 discharges. On the other hand, these hospitals include >75% of all discharges.

Panel Member 3: Different from measures #2880 and #2881, ICCs for hospitals with ≥50 admissions the majority of hospitals were >.60, which would be in the range of adequate under current guidance.

Panel Member 4: The test result is modest to high for 4 of the 6 scenarios of hospital admission counts. The 4 largest hospital counts (where the smallest count is. >=100 admits) the split sample ICC was 0.628 and 0.709. For the hospital counts where the range is >=25, the split sample ICC was modest at 0.576. My summary here removes the smallest hospital count (>=2) as CMS' typical minimum threshold is 25 cases.

Panel Member 5: split-sample reliability testing: 0.56 for >= 25 admissions and 0.69 for >= 200 admissions

Panel Member 6: No major concerns.

Panel Member 7: ICC=0.576 at current volume threshold for reporting, although it would rise to 0.668 at a plausible increase in the threshold from 25 to 100.

Panel Member 8: For hospitals with>25 admissions (89.6%) ICC=0.576, for >50 0.628, >300 (34.9% of hospitals) 0.709. Reasonable

Panel Member 9: Reliability is moderate - ICC values range from .57 to .70 across a reasonable range of sample sizes.

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

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

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

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

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

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

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

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

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

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

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

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

    Panel Member 1: Reasonable reliability in large hospitals

    Panel Member 3: The ICC's are in the range of what would be considered as moderate/adequate under current guidance.

    Panel Member 4: The test result is modest to high for 4 of the 6 scenarios of hospital admission counts. The 4 largest hospital counts (where the smallest count is. >=100 admits) the split sample ICC was 0.628 and 0.709. For the hospital counts where the range is >=25, the split sample ICC was modest at 0.576. My summary here removes the smallest hospital count (>=2) as CMS' typical minimum threshold is 25 cases.

    Panel Member 6: No major concerns.

    Panel Member 8: ICC appropriate test and results acceptable

Panel Member 9: Scores in .57 - .7 range.

VALIDITY: TESTING

12. **Validity testing level: ☒ Measure score ☐ Data element ☐ Both**
13. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?**
    NOTE that data element validation from the literature is acceptable.

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

    ☒ **Yes**

    ☐ **No**

    ☒ **Not applicable** (data element testing was not performed)
14. **Method of establishing validity of the measure score:**

    ☒ **Face validity**

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

    ☐ **N/A (score-level testing not conducted)**
15. **Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?**

    Submission document: Testing attachment, section 2b1.

    ☒ **Yes**

    ☒ **No**

    ☐ **Not applicable** (score-level testing was not performed)
16. **Assess the method(s) for establishing validity**

    Submission document: Testing attachment, section 2b2.2

    Panel Member 1: Face validity assessed by expert committee Empiric validity assessed by correlation of measure with star ratings and readmission measure.

    Panel Member 3: Face validity was assessed using survey-based information provided by the 16 member TEP assembled to address this and other measures. Construct validity was assessed as the relationship between the pneumonia EDAC measure score and the risk standardized readmission group scores, the overall hospital rating scores and the pneumonia readmission measure. The developer posited a negative relationship between the pneumonia EDAC scores, star-rating readmission score group and star-rating summary scores. They also hypothesized a positive relationship between the pneumonia EDAC scores and the pneumonia RSRR scores.

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

    Panel Member 5: "correlation analysis:

    - Hospital Star Rating readmission group score: -0.416 (p<.0001),
    - Overall Hospital Star Rating summary score: -0.398 (p<.0001)
    - PN Readmission Measure: 0.625 (p<.0001), "

    Panel Member 6: No major concerns.

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

Panel Member 8: Face validity (11/12 of Expert Panel) Empirical validity: correlation with three other metrics, two related specifically to readmissions and one related to overall hospital quality of which readmission is part.

Panel Member 9: Developer correlated several measures that all include readmission events for the same patient population. Some correction must be made for the guaranteed correlation that this approach creates. Correlations without correction are not interpretable.

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

Submission document: Testing attachment, section 2b2.3

Panel Member 1: >80% of experts moderately or strongly agreed with validity of measure Directionally appropriate correlation of measure with both star ratings and readmission risk

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

Panel Member 4: Regarding measure score testing, the correlation with the EDAC measure and: - CMS readmits group rating was moderate at -0.416 [p18] - CMS overall star rating was moderate at -0.398 [p19] -CMS PN readmits rating was moderate to strong at 0.625 [p20]

Panel Member 5 "correlation analysis:

- Hospital Star Rating readmission group score: -0.416 (p<.0001),
- Overall Hospital Star Rating summary score: -0.398 (p<.0001)
- PN Readmission Measure: 0.625 (p<.0001),"

Panel Member 6: No major concerns.

Panel Member 7: Correlations are adequate - for example, r=0.625 between EDAC and readmissions - but the test is very easy to pass given that readmissions drive EDAC. Process-outcome correlations or pre-post analyses of intervention effects are strongly preferred.

Panel Member 8: Circular logic--measuring correlation of ER visits and readmission with readmissions is not a test of the validity of the measure as a quality metric in the care of patients with pneumonia--correlation with mortality, need for VV ECMO or other clinically validated measure would be more meaningful

Panel Member 9: Per note above - not interpretable as presented.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

Submission document: Testing attachment, section 2b2.

Panel Member 1: Medicare Advantage

Panel Member 4: No concerns with the exclusions.

Panel Member 6: No major concerns.

Panel Member 7: none

Panel Member 8: exclusions seem appropriate although not clear that planned procedures would necessarily be excluded due to the broad nature of what are considered acute admission diagnoses

Panel Member 9: None

19. **Risk Adjustment**

Submission Document: Testing attachment, section 2b3

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

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

☐ Yes ☐ No ☒ Not applicable

19c. **Social risk adjustment:**

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

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

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

19d.**Risk adjustment summary:**

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

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

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

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

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

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

Panel Member 1: Two-part model (any days, number of days | >0 days), with adjustment for age, sex, and comorbidity

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

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

Panel Member 4: The risk adjustment methods are appropriate for the given measure. No issues with the measures, but issues with the results, which follow: The c-statistics noted in response to 2b3.6 are 0.616 & 0.62. While there are no hardlines as to a high, moderate & low c-statistic, from everything I've seen at, or below, 0.6 is generally unacceptable. In turn, marginally over 0.6 is between questionable to marginally acceptable. The R squared result of 0.038 is low. The response in 2b3.10 essentially confirms this in the statement: "0.038 indicates that patients' clinical risk factors can explain 3.8% of the variation in the numbers of excess days in acute care" Having said the above, the risk decile plot (fig. 5) show acceptable performance in each decile with the exception of the 1st and 10th decile.

Panel Member 5: cannot evaluate model validation because validation was not performed in a validation data set

Panel Member 6: No major concerns.

Panel Member 7: The models perform poorly, with c=0.62 from the first stage logistic model and R2=0.038 from the second stage Poisson model. Two tested social risk factors are statistically significant in the Poisson model, but do not appear to meaningfully affect hospital performance estimates.

Panel Member 8: C-statistic fair, only 0.62 with no change if SRS included. Minimal median effect (0.6/100 admissions) for adding SRS. Problem is the median is not where the effect would be felt--never checked net reclassification index or how it would impact outliers.

Panel Member 9: I actually didn't read this section closely - it's essentially like the two other EDAC meausures, and the approach to validity is a fatal flaw. Without that being fixed, risk adjustment doesn't matter in the context of validity.

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

    Submission document: Testing attachment, section 2b4.

    Panel Member 1: Low discrimination

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

    Panel Member 4: No concerns as there is a fair degree of variation expressed when testing at the 95% confidence interval. Specifically, of hospitals with 25 or more cases, 14% had "fewer days" and 25% had "more days".

    Panel Member 5: none

    Panel Member 6: No major concerns.

    Panel Member 8: Adequate distribution

    Panel Member 9: As in other measures, there was no effort made to identify and analyze "meaningful". The measure can be used to detect statistical outliers, but it's not clear whether the measure will detect "meaningful" differences.

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

    Submission document: Testing attachment, section 2b5.

    Panel Member 3: N/A

    Panel Member 4: No concerns.

    Panel Member 6: No major concerns.

    Panel Member 8: Basically, based on claims; other data sources used to identify patients and eligibility--no concerns.

    Panel Member 9: N/A

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

    Submission document: Testing attachment, section 2b6.

    Panel Member 3: No missing data reported.

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

    Panel Member 6: No major concerns.

    Panel Member 7: none

    Panel Member 8: No concerns

    Panel Member 9: None

For cost/resource use measures ONLY:

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

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

24. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**
25. **OVERALL RATING OF VALIDITY considering 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 4: Regarding measure score testing, the correlation with the EDAC measure and: - CMS readmits group rating was moderate at -0.416 [p18] - CMS overall star rating was moderate at -0.398 [p19] -CMS PN readmits rating was moderate to strong at 0.625 [p20]

    Panel Member 5: cannot evaluate model validation because validation was not performed in a validation data set

    Panel Member 6: No major concerns.

    Panel Member 7: Inappropriate choice of variables for testing construct validity; weak risk-adjustment models. Also, the impact of using Medicare claims data vs. VA data is not adequately explored.

    Panel Member 8: Empirical testing not appropriate/meaningful and risk adjustment only fair and without SRS despite known impact of SRS on outcome of interest

    Panel Member 9: Correlational analyses used to establish empirical validity must be corrected for inclusion of the same readmission events on both sides of the correlation calculation.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

    ☐ **High**

    ☐ **Moderate**

    ☐ **Low**

    ☐ **Insufficient**

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

ADDITIONAL RECOMMENDATIONS

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

    Panel Member 6: No major concerns.

    Panel Member 7: Inappropriate choice of variables for testing construct validity, generalizability of the same methods from Medicare claims data to VA data.

Panel Member 8: Bias of measure developers evident in use of the term "extra" when they have no way to determine if the additional days were "extra" (i.e., not needed if adequate care had been delivered in index admission)