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: Click 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 #: 3457

Corresponding Measures:

De.2. Measure Title: Minimizing Institutional Length of Stay

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

De.3. Brief Description of Measure: The proportion of admissions to an institutional facility (e.g., nursing facility, intermediate care facility for individuals with intellectual disabilities [ICF/IID]) for managed long-term services and support (MLTSS) plan enrollees that result in successful discharge to the community (community residence for 60 or more days) within 100 days of admission. This measure is reported as an observed rate and a risk-adjusted rate.

1b.1. Developer Rationale: This measure evaluates how well MLTSS plans can minimize length of stay in institutions for MLTSS enrollees (i.e., keep stays to less than 100 days). This is important to MLTSS enrollees who are newly admitted to institutional facilities for temporary skilled nursing needs, but could live independently in a community setting with appropriate support. The 100-day cut point is generally considered a cross-over point because after that time (1) residence in the institutional facility becomes considered semi-permanent, (2) dual eligible enrollees lose Medicare coverage for their institutional stay, and (3) they often lose any community-based housing they were previously using further limiting the probability they will ever return to the community.

Most individuals who need long-term services and supports (LTSS) want to live at home or in community settings (Guo et al. 2015; Keenan 2010). Although state Medicaid agencies, which pay for the majority of LTSS, have made significant progress in rebalancing their LTSS systems to provide more HCBS, many Medicaid beneficiaries still reside in institutions, which is both costly and can be associated with adverse outcomes such as hospital readmissions. A key goal of MLTSS programs is to minimize lengths of stay in institutions. MLTSS plans achieve this goal by conducting person-centered assessment and care planning, providing timely access to high quality HCBS, and coordinating LTSS and medical care across providers and settings (Felix et al. 2011; Greiner et al. 2013; Sands et al. 2012). While it may not be possible to avoid the use of institutional care entirely, this measure assesses the effectiveness of MLTSS plans in minimizing length of stay in institutions when the same services can be provided in the home or community settings. Keeping individuals in home and community settings and minimizing their stays in institutions helps to increase their quality of life and care experience.

Finally, there is some evidence from the general population (not limited to Medicaid or LTSS users) suggesting the duration of institutional stays among LTSS users may vary by state. Among all nursing home admissions of

Medicare beneficiaries in 2012 (not limited to those dually eligible for Medicare and Medicaid or LTSS users), the median percentage of stays that lasted 100 days or more was 18.7 in 2012; the highest share was in Louisiana (35 percent), and the lowest was in Arizona (8.9 percent) (Irvin et al. 2016). However, these state-level estimates might mask variation by programs across states, such as differences in 1915(c) HCBS waiver programs. It is also important to note that while this variation has been identified across states, it may be less relevant in an MLTSS environment where managed care plans are all operating within a single state, and subject to the same LTSS requirements.

Citations

Felix, Holly C., Glen P. Mays, M. Kathryn Stewart, Naomi Cottoms, and Mary Olson. "Medicaid Savings Resulted When Community Health Workers Matched Those With Needs to Home and Community Care." Health Affairs, vol. 30, no. 7, 2011, pp. 1366-1374.

Greiner, Melissa A., Laura G. Qualls, Isao Iwata, Heidi K. White, Sheila L. Molony, M. Terry Sullivan, Bonnie Burke, Kevin A. Schulman, and Soko Setoguchi. "Predicting Nursing Home Placement Among Home- and Community-Based Services Program Participants." The American Journal of Managed Care, vol. 20, no. 12, 2014, pp. e535-e536.

Guo, Jing, R. Tamara Konetzka, Elizabeth Magett, and Willian Dale. "Quantifying Long-Term Care Preferences." Medical Decision Making, vol. 35, no. 1, 2015, pp. 106-113.

Irvin, Carol, Noelle Denny-Brown, Eric Morris, and Claire Postman. Table 5. Indicators of performance of state long-term services and supports systems, "Pathways to Independence: Transitioning Adults Under Age 65 from Nursing Home to Community Living (), table 5, Indicators of performance of state long-term services and supports systems. Cambridge, MA: Mathematica Policy Research, 2016, pp. 30-31. Available at https://www.medicaid.gov/medicaid/ltss/downloads/mfpfieldreport19.pdf.

Keenan, Teresa A. "Home and Community Preferences of the 45+ Population." Washington, DC: AARP Public Policy Institute, November 2010. Available at http://assets.aarp.org/rgcenter/general/home-community-services-10.pdf.

Sands, Laura P., Huiping Xu, Joseph Thomas, III, Sudeshna Paul, Bruce A. Craig, Marc Rosenman, Caroline C. Doebbeling, and Michael Weiner. "Volume of Home- and Community-Based Services and Time to Nursing-Home Placement." Medicare & Medicaid Research Review, vol. 2, no. 3, 2012, pp. E1-E21.

S.4. Numerator Statement: The count of discharges from an institutional facility to the community that occurred within 100 days or less from admission and resulted in successful discharge to the community (community residence for 60 or more days).

S.6. Denominator Statement: New admissions to an institutional setting for MLTSS enrollees age 18 and older.

S.8. Denominator Exclusions: None.

De.1. Measure Type: Outcome

S.17. Data Source: Claims, Enrollment Data

S.20. Level of Analysis: Health Plan

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

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? Not applicable.

Criteria 1: Importance to Measure and Report

1a. Evidence

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.

Evidence Summary

- This is a measure of how well MLTSS plans can minimize MLTSS enrollee length of stay in institutions.
- The developer notes that improvement on this outcome will require MLTSS plans to develop discharge plans in collaboration with nursing facility staff, and coordinate appropriate home and community based services to ensure a successful transition.

Question for the Committee:

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

Guidance from the Evidence Algorithm

Box 1: The measure assesses a healthcare outcome →Box 2: The developer has provided empirical data that there is a relationship between the measured outcome and at least one healthcare outcome → Pass

The highest possible rating is 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 notes significant variation across HPPLs in performance with regard to performance. The risk adjusted median performance is 36.49 with a range of 0.0 to 65.88.

Disparities

- The developer did not test differences in performance by race/ethnicity in the testing data. Age, gender, and dual eligible status are included in the statistical risk model.
- The developer noted limited evidence on potential disparities in the quality of LTSS care and hypothesized this could be caused by the lack of valid and reliable performace measures in this area.
- The developer identified two studies that suggest there may be disparities in care in this area. One study of dually eligible Medicaid HCBS recipients age 65 and older found that blacks and Hispanics were less likely than whites to have a nursing home stay; they remained in the community longer and were more physically and cognitively impaired upon admission (Cai and Temkin-Greener 2015). Still another study of Medicaid enrollees found that those newly admitted to nursing homes were more likely to be older white women (Schmitz et al. 2014), which is consistent with the findings for Medicaid HCBS users. The developer found that taken together, these two studies suggest there may be

racial/ethnic disparities in the use of nursing homes and/or cultural differences that lead to variance in the settings where people receive long-term care.

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.

- There was not much evidence specifically for MLTSS individuals regarding the link between MLTSS programs and successful reduction of long term cases.
- Yes-empirical data for health outcome measure

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?

- There was minimal evidence for identifying gaps in care for MLTSS patients; gaps were identified for HCBS recipients.
- There is variability in utilization of institutional LTC in these populations, though whether that indicates there is a quality gap remains a question.
- Seems like performance varies across states

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:

- Larry Glance
- Karen Joynt Maddox
- Marybeth Farquhar
- Eugene Nuccio
- Christie Teigland
- Steve Horner**Evaluation of Reliability and Validity (and composite construction, if applicable)**:

Summary of Methods Panel Review:

- Subgroup members found the measure to be reliable and valid in their preliminary analyses, thus this measure was not discussed during the Methods Panel measure evaluation call.

Standing Committee Action Item(s):

- The Standing Committee can discuss the reliability and validity testing, or agree to accept the ratings of the Scientific Methods Panel. It is important to note that the appropriateness of inclusion or exclusion of social risk factors was not within scope for the Scientific Methods Panel ratings.

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?
    - Please note the inclusion or exclusion of social risk factors was not within the scope of the Scientific Methods Panel ratings.

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

Scientific Acceptability

Measure Number: 3457
Measure Title: Minimizing institutional length of stay

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

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.*

   - Some Methods Panel members expressed concerns about the specifications being precise, unambiguous, and complete.
     - Please note that to be eligible for a moderate rating per NQF's reliability algorithm submitted specifications must be specifications precise, unambiguous, and complete so that they can be consistently implemented.
   - The developer provided several clarifying comments on the measure specifications to address the Methods Panel members concerns.
     - The developed noted that CMS has established a technical assistance mailbox facilitate implementation, and has other implementation support planned for 2019.
     - The developer clarified that this measure identifies new admissions, therefore they remove transfers from another institution (defined as a Medicaid or Medicare certified nursing facility or institutional care facility for individuals with intellectual disabilities) in order to correctly identify the first stay in a sequence of stays. Similarly, in a sequence where institutional stay #1 is followed by a hospital stay and then instutional stay #2, only the first stay is counted in the measure. Instiutional stays that end in death are not counted in the measure, given these enrollees do not have an opportunity for successful discharge.
   - A summary of the Methods Panel members feedback is provided below:

- **PANEL MEMBER 1:** Measure specifications could be presented more clearly.
- **PANEL MEMBER 2:** Yes
- **PANEL MEMBER 3:** Description of the measure is relatively clear, however, the steps to calculate the measure algorithm includes removal of admissions such as "transfer from another institution", "admission from hospital that originated from another institution", "death", etc.

2. **Briefly summarize any concerns about the measure specifications.**
   - **Methods Panel noted the following concerns about the measure specifications.**
     - **PANEL MEMBER 3:** Developer should include the exclusions as they are listed in S.14 of the MIF for measure calculation.
     - **PANEL MEMBER 4:** Narrative and rationale for this Minimize LOS measure is identical to Admissions to Institution measure. There should be something that differentiates these two measures; please identify.
       - The Brief Description contains details (e.g., definition of successful discharge as 60+ days) that do not appear in the specific statements for Numerator and Denominator. The specific statements seem incomplete in the MIF document.
   - Note: the developer has since revised the MIF (item S.4) to address these concerns.

RELIABILITY: TESTING

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

3. **Reliability testing level** ☒ **Measure score** ☐ **Data element** ☐ **Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure**
   ☒ **Yes** ☐ **No**
5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical <u>VALIDITY</u> testing** of <u>patient-level data</u> conducted?
   ☐ **Yes** ☐ **No**
6. **Assess the method(s) used for reliability testing**
   Submission document: Testing attachment, section 2a2.2
   - Reliability was assess at the measure scocre level. The developer used a signal-to-noise analysis using the Morris method.
   - The developer provided clarification on the source of their testing data. There are no existing, nationally standardized datasets for Medicaid beneficiaries enrolled in managed long-term services and supports (MLTSS) plans, which is the target population for this measure. Therefore, the developer worked directly with health plans to obtain the enrollment and claims data needed to support measure testing. These data represented 4 parent health plan organizations, and 14 different health plan product lines (HPPLs) from 10 states, located in geographically diverse regions of the country. Health plans are anticipated to calculate this measure utilizing their own data, similar to reporting for the Healthcare Effectiveness Data and Information Set (HEDIS) meaures.The measure also specifies that only enrollees with both LTSS and medical benefits are eligible. This ensures that health plans should have access to information on both LTSS enrollment and services provided as well as institutional admissions.
   - A summary of the Methods Panel members feedback is provided below. This feedback is intended to inform Standing Committee discussion.
     - **PANEL MEMBER 1:** Tested measure reliability using SNR.
     - **PANEL MEMBER 2:** Signal to noise
     - **PANEL MEMBER 3:** Used signal-to-noise ratio, which is appropriate.

- **PANEL MEMBER 4**: Signal to Noise (ratio of standard deviations) for measure across 13 plans; value seemed adequate. Note: source data across all states is likely to be inconsistent and vary in quality.

7. **Assess the results of reliability testing**
   Submission document: Testing attachment, section 2a2.3
   - Results of Signal-To-Noise Reliability testing [a]

HPPL	Mean SNR
HPPL-01	0.9511
HPPL-02	0.8454
HPPL-03	0.6326
HPPL-04	0.9907
HPPL-05	0.9799
HPPL-06	0.9237
HPPL-07	0.9861
HPPL-08	0.8592
HPPL-09	0.9309
HPPL-11	0.9809
HPPL-12	N/A
HPPL-13	0.9862
HPPL-14	0.8736
HPPL-17	0.9870
Average	**0.9175**

Source: Mathematica analysis of data from four MLTSS plans and fourteen health plan product lines.

[a] SNR values excluding plans with fewer than 11 enrollees in the numerator.

- Average and plan-level SNRs reported: Plan SNRs ranged 0.63 to 0.99
- A summary of the Methods Panel members feedback is provided below. This feedback is intended to inform Standing Committee discussion.
  - **PANEL MEMBER 1**: Mean was 0.86, consistent with almost perfect reliability. using Landis scale.
  - **PANEL MEMBER 2:** Average reliability score is high
  - **PANEL MEMBER 3:** SNR = 0.92 indicating high
  - **PANEL MEMBER 4:** Value of Signal-to-Noise was 0.92; acceptable.

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):

- Ultimately, the Scientific Methods Panel gave this measure an overall rating of reliability of moderate. Individual members' ratings ranged from low to high. A rationale for members' ratings is provided in item 11 below.

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:** Mean reliability score was 0.86, consistent with almost perfect reliability using Landis scale.
    - **PANEL MEMBER 2:** Reliability reasonable for all subsets and excellent for some
    - **PANEL MEMBER 3:** Results of the testing for reliability. Needs clarification of exclusions for numerator/denominator.
    - **PANEL MEMBER 4:** National Medicare data are inconsistent—as the authors note in their response. Hence, while the sample data used appears to be of reasonable quality, generalization of this level of quality across the 50 states is probably not warranted.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

12. **Please describe any concerns you have with measure exclusions.**
    Submission document: Testing attachment, section 2b2.
    - The Scientifc Methods Panel members raised several concerns with the measure exclusions.
    - The developer clarified that this measure identifies new admissions, therefore they remove transfers from another institution (defined as a Medicaid or Medicare certified nursing facility or institutional care facility for individuals with intellectual disabilities) in order to correctly identify the first stay in a sequence of stays. Similarly, in a sequence where institutional stay #1 is followed by a hospital stay and then instutional stay #2, only the first stay is counted in the measure. Instiutional stays that end in death are not counted in the measure, given these enrollees do not have an opportunity for successful discharge.
    - A summary of Methods Panel members concerns are listed below:
        - **PANEL MEMBER 3:** Developer indicated "no exclusions" but there seem to be a number of them when calculating the measure.
        - **PANEL MEMBER 4:** There were no exclusions to this measure identified by the authors. However, in S.14 Calculation Algorithm/Measure Logic lists "exclusions" in steps 2 -4 "Remove admissions…". Numerator and Denominator statements are incomplete.
13. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.**
    Submission document: Testing attachment, section 2b4.
    - Methods Panel members raised a number of question about the ability to identify meaningful differences in performance. One Panel member noted the Standing Committee consider the developer's decision to dichotomize at 100 days.
        - The developer clarified that The 100-day cut point is generally considered a cross-over point in long-term care because after that time: (1) residence in the institutional facility becomes considered semi-permanent, (2) dual eligible enrollees lose Medicare coverage for their institutional stay, and (3) they often lose any community-based housing they were previously using further limiting the probability they will ever return to the community.
    - The developer clarified that they evaluate differences in performance in Section 2b4 by comparing the value of the observed to expected rate for each plan, and evaluating the results for statistically significant differences.
    - A summary of the Methods Panel members feedback is provided below:
        - **PANEL MEMBER 1:** None. There appears to be a reasonable spread of risk adjusted rates across units of analysis.

- **PANEL MEMBER 2:** This is a question for the content group rather than the methods group, but dichotomizing at 100 days seems like a strange choice – so much information is lost. That said, if it's appropriate from a content standpoint it is simple to model from a methodological standpoint.
- **PANEL MEMBER 3:** No concerns.
- **PANEL MEMBER 4:** Meaningful difference criteria among plans are not specified.

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.
- One Methods Panel member noted a concern about the quality of the data source. The developer clarified they e worked directly with health plans to obtain the Medicaid enrollment and claims data needed to support measure testing, which reflects how the measures will be calculated for implementation. These data represented 4 parent health plan organizations, and 14 different health plan product lines (HPPLs) from 10 states, located in geographically diverse regions of the country. National Medicare data was not used.

15. **Please describe any concerns you have regarding missing data.**
Submission document: Testing attachment, section 2b6.
- One Methods Panel member noted concerns about the quality of the source Medicare data. Please see item 14 above for the developer's clarification.

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
- The developer stated that no social risk factors included; however, NQF consider dual eligible status to be a social risk factor. The developer noted that this measure was not intended for use in the more heterogeneous non-dually eligible Medicare population.

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
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
16d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration)
☒ Yes ☐ No
16d.5.Appropriate risk-adjustment strategy included in the measure? ☒ Yes ☐ No
PANEL MEMBER 1: Adjusting for social risk may not be important since this measure focuses on a population with high social risk

16e**. Assess the risk-adjustment approach**
- Risk adjustment with 18 factors
- Risk Model: A logistic regression was used to model the log-odds of an admission being successfully discharged within 100 days of admission.

- Note: The panel members suggest that Standing Committee weigh in on the 100-day threshold; they were unsure whether this threshold was appropriate.
- Risk Factors: Age, gender, dual-eligibility for Medicaid and Medicare, hospital utilization in the period prior to admission to the institutional facility, enrollment in the MLTSS plan for 6 months, and comorbid conditions (Alzheimer's disease and related disorders, asthma, intellectual disabilities, mental health conditions, stroke).

- Developers noted a conceptual rationale for marital status and functional status, but cited lack of data as reason for not analyzing these variables.
- The developer stated that no social risk factors included; however, NQF consider dual eligible status to be a social risk factor. The developer noted that this measure was not intended for use in the more heterogeneous non-dually eligible Medicare population.
  - Note: Developer cites NQF's SES trial results as rationale for not including SES. NQF does not agree with this interpretation.
- Scientific Methods Panel members noted concern of a lack of compelling analysis to support the exclusion of dual status in the risk adjustment model. However, the data provided by the developer states that dual eligible status is a variable in the risk model. As stated above, the developer notes that they do not consider this an adjustment for social risk.
- A summary of Methods Panel Feedback is provided below. This summary is intended to inform Standing Committee discussion.
  - **PANEL MEMBER 1:** Used stepwise selection for covariate selection. Automated methods suffer frm lack of input from content experts.
    - Final model includes Age, gender, dual-eligibility for Medicaid and Medicare, hospital utilization in the period prior to admission to the institutional facility, enrollment in the MLTSS plan for 6 months, and comorbid conditions (Alzheimer's disease and related disorders, asthma, intellectual disabilities, mental health conditions, stroke).
    - Logistic regression model
    - OE ratio
    - Discrimination is modest with C statistic of 0.63. Did not provide value in validation data.
    - Although HL statistic was not significant (indicative of good calibration), the HL statistic is likely to be not significant in such a small sample size used for testing (n=3,384).
    - Unclear why, given the very large size of the sample, the sample in validation set for the HL stat was so small.
  - **PANEL MEMBER 2:** Appropriate discrimination and calibration.
  - **PANEL MEMBER 3:** Appropriate risk adjustment for this population. Developer reported on several other risk factors suggested by TEP members and did a fairly through analysis on each one. Some of the suggested factors were included in the model and others were ruled out.

VALIDITY: TESTING

17. **Validity testing level:** ☒ **Measure score** ☐ **Data element** ☐ **Both**
18. **Method of establishing validity of the measure score:**
    - ☐ **Face validity**
    - ☒ **Empirical validity testing of the measure score**
    - ☐ **N/A (score-level testing not conducted)**
19. **Assess the method(s) for establishing validity**

Submission document: Testing attachment, section 2b2.2

- Validity was assessed at the measure score level.
- Measure score
    - The developer conducted score-level validity testing (construct validity) by comparing measure results with results from two other measures (#3456: Admission to an institution from the community and #3458: Successful transition after long-term institutional stay).
- The measure developer sought input from a TEP in the development process but evaluated convergent validity to assess validity of the measure score.
- A summary of the Methods Panel members feedback is provided below:
    - **PANEL MEMBER 1:** TEP was involved in the development of this measure, but did not conduct formal assessment of face validity of measure.
        - Convergent validity was evaluated by comparing measure to 2 other measures currently undergoing endorsement consideration. Since these measures are not currently endorsed, the value of these measures for evaluating convergent validity is unclear
        - Empirical validity is assessed using predictive validity of risk adjustment model. Model calibration cannot be effectively evaluated using this small data set. It would have been helpful to provide a calibration plot.
    - **PANEL MEMBER 2:** Convergent validity
    - **PANEL MEMBER 3:** Developer used Convergent validity and face validity through input of four technical expert panels. They also included a public comment period.
    - **PANEL MEMBER 4:** A variety of empirical, consensus (expert panel), and literature validity tests were performed.

20. **Assess the results(s) for establishing validity**
    Submission document: Testing attachment, section 2b2.3

- Results of the analyses were generally as hypothesized:
    - There was a moderate, negative correlation between a measure of utilization of long-stay institutional care and performance on this measure of minimizing institutional length of stay.
    - There was an even stronger positive correlation between a measure of successful transition to the community after long-stay institutional stay and this measure of minimizing institutional length of stay (correlation= 0.89, p-value= 0.0005).
- Results of Correlation Analyses:

Measure	**Strata**	**Correlation Coefficient**	**p-value**
Admission to an institution from the community: Long Stay (101+ days)**	Age 18-64	-0.12321	0.6747
	Age 65-74	-0.54125*	0.0456
	Age 75-84	-0.47965	0.0826
	Age 85+	-0.56292*	0.0361
Successful transition after long-term (101+ days) institutional stay	Risk-adjusted for age, gender and co-morbid conditions	0.89091**	0.0005

Source: Mathematica analysis of data from four MLTSS plans and fourteen health plan product lines.

*Correlation was significant at p<.05

**Correlations between the proposed measure and Long-stay admissions was hypothesized to be negative because the measure of long-stay admissions is a "lower is better" measure.

- A summary of the Methods Panel members feedback is provided below:
    - **PANEL MEMBER 1:** TEP was involved in the development of this measure, but did not conduct formal assessment of face validity of measure.
        - Convergent validity was evaluated by comparing measure to 2 other measures currently undergoing endorsement consideration. Since these measures are not currently endorsed, the value of these measures for evaluating convergent validity is unclear
        - Empirical validity is assessed using predictive validity of risk adjustment model. Model calibration cannot be effectively evaluated using this small data set. It would have been helpful to provide a calibration plot.
    - **PANEL MEMBER 2:** Reasonable correlation with other measures
    - **PANEL MEMBER 3:** Results showed moderate to good associations for hypothesized relationships suggesting that measure meets the test of validity.
    - **PANEL MEMBER 4:** The results from these tests seemed to support the focus of the measure.

21. **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)

22. **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)

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

    - Ultimately, the Scientific Methods Panel gave this measure a preliminary overall rating of validity of moderate. Individual members' ratings ranged from low to high. A rationale for members' ratings is provided under item 24 below.

24. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**
    - **PANEL MEMBER 1:** The predictive validity of the risk adjustment model has not been demonstrated
    - **PANEL MEMBER 2:** No concerns
    - **PANEL MEMBER 3:** Results of the testing.
    - **PANEL MEMBER 4:** The measure validity would be more accurately as good or very good, assuming that the risk adjustment results are positive.

ADDITIONAL RECOMMENDATIONS

25. **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.**

    - No additional concerns were identified

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?

- Reliability was high using the SNR methodology; however, variability in state data may result in differences in reliability if widely disseminated.
- The measure description includes language re: ICF/IID beneficiaries, but then ?intellectual disabilities are used as a risk adjuster and 100 day cut off is used because MCR no longer pays for post-acute/SNF care after that time. If this is about ICF LTC (rather than all long term care) utilization, is the 100 day cut off relevant as stated? Do agree that 100+ days generally means someone is headed for permanent institutional LTC. Also agree that this is already a high risk population but perhaps some of the variability may be related to geography as well and "area-level socioeconomic" indicators?
- None

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

- Per SNR seems very reliable
- No

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

- Testing based on comparisons to other measures
- No

2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data): 2b4. Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality? 2b5. Comparability of performance scores: If multiple sets of specifications: Do analyses indicate they produce comparable results? 2b6. Missing data/no response: Does missing data constitute a threat to the validity of this measure?

- The 100 day cut point is based on financial and rules regarding length of stay. It wasn't clear why further stratification wasn't examined for stays less than 100 days to determine the applicability of the metric for lower LOS.
- 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?

- Modest C statistic (0.63); construct validity was determined via convergent validity by comparing metrics, with a moderate level of association. No TEP input for this metric. This appears to be the primary issue with this metric; with an arbitrary 100-day selection, results may be somewhat removed from a measure of effective programs that attempt to reduce LOS.
- Still wonder about SES factors, especial area related access to robust LTSS services. In some communities, even if MMCO will pay, there aren't any service providers. Does this measure quality or access or both?
- Appears some exclusions should be considered

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.

- The measure is calculated using claims data.
- All data elements are in defined fields in electronic claims.
- The American Hospital Association holds a copyright to the Uniform Bill Codes ("UB") contained in the measure specifications. The UB Codes in this specification are included with the permission of the AHA. The UB Codes contained in this specification may be used by health plans and other health care delivery organizations for the purpose of calculating and reporting the measure results or using the measure results for their internal quality improvement purposes. All other uses of the UB Codes require a license from the AHA.

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?
- Is the data collection strategy ready to be put into operational use?

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 data. Should be feasible
- Data already collected

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 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.

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

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

OR

Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

This measure is intended for use by states and health plans to monitor and improve the quality of care provided for the Medicaid MLTSS enrollee population. The intended use of the measure is to allow states to compare and evaluate the quality of LTSS care being provided in MLTSS plans with which they contract using a common, standardized and validated measure. This specific measure focuses on one outcome of high quality MLTSS care – minimizing institutional stays and ensuring successful return to the community. In May 2017, the National MLTSS Health Plan Association recommended a set of model LTSS performance measures and network adequacy standards in an effort to assist states in complying with the 2016 final rule on managed care in Medicaid and Children's Health Insurance (CHIP). This measure is included in the set of recommended measures that assesses person-centered planning and coordination.

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

Not applicable. This measure has not been implemented yet. Unlike Medicare measures, there is no formal process by which draft results for Medicaid measures are shared with measured entities, such as a Dry Run used in the Hospital Inpatient Quality Reporting (IQR) and Outpatient Quality Reporting (OQR) programs. Feedback on the measure will be available after the measure has been implemented by states in their MLTSS programs.

Additional Feedback:

n/a

***Questions for the Committee**:*

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

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

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

4b. Usability 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.

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

Improvement results

This measure is not yet implemented, thus longitudinal data is not available.

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

This measure has not been implemented yet. There were no unexpected findings identified during testing of this measure.

Potential harms

This measure has not been implemented yet. There were no unexpected findings identified during testing of this measure.

Additional Feedback:

n/a

Questions for the Committee:

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

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

RATIONALE:

This measure is a new measure, thus usability results are unavailable. Although the Use and Usability criterion is not met, the measure may be suitable for endorsement based on an assessment of the strength of the measure in relation to the other three evaluation criteria and the strength of the competing and related measures to drive improvement.

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

4a. 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?

- The metric is not currently in use.
- Not currently being used, sounds like they reached out to stakeholders and there is agreement that MLTSS providers need some quality measures.
- Observational

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.

- Unknown
- Observational but could allow more capacity

Criterion 5: Related and Competing Measures

Related or competing measures

- NQF did not identify competing measures.

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

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

- None
- Resolved

Public and Member Comments

NQF received no public or member comments on this measure as of January 25, 2019.

Brief Measure Information

NQF #: 3457

Corresponding Measures:

De.2. Measure Title: Minimizing Institutional Length of Stay

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

De.3. Brief Description of Measure: The proportion of admissions to an institutional facility (e.g., nursing facility, intermediate care facility for individuals with intellectual disabilities [ICF/IID]) for managed long-term services and support (MLTSS) plan enrollees that result in successful discharge to the community (community residence for 60 or more days) within 100 days of admission. This measure is reported as an observed rate and a risk-adjusted rate.

1b.1. Developer Rationale: This measure evaluates how well MLTSS plans can minimize length of stay in institutions for MLTSS enrollees (i.e., keep stays to less than 100 days). This is important to MLTSS enrollees who are newly admitted to institutional facilities for temporary skilled nursing needs, but could live independently in a community setting with appropriate support. The 100-day cut point is generally considered a cross-over point because after that time (1) residence in the institutional facility becomes considered semi-permanent, (2) dual eligible enrollees lose Medicare coverage for their institutional stay, and (3) they often lose any community-based housing they were previously using further limiting the probability they will ever return to the community.

Most individuals who need long-term services and supports (LTSS) want to live at home or in community settings (Guo et al. 2015; Keenan 2010). Although state Medicaid agencies, which pay for the majority of LTSS, have made significant progress in rebalancing their LTSS systems to provide more HCBS, many Medicaid beneficiaries still reside in institutions, which is both costly and can be associated with adverse outcomes such as hospital readmissions. A key goal of MLTSS programs is to minimize lengths of stay in institutions. MLTSS plans achieve this goal by conducting person-centered assessment and care planning, providing timely access to high quality HCBS, and coordinating LTSS and medical care across providers and settings (Felix et al. 2011; Greiner et al. 2013; Sands et al. 2012). While it may not be possible to avoid the use of institutional care entirely, this measure assesses the effectiveness of MLTSS plans in minimizing length of stay in institutions when the same services can be provided in the home or community settings. Keeping individuals in home and community settings and minimizing their stays in institutions helps to increase their quality of life and care experience.

Finally, there is some evidence from the general population (not limited to Medicaid or LTSS users) suggesting the duration of institutional stays among LTSS users may vary by state. Among all nursing home admissions of Medicare beneficiaries in 2012 (not limited to those dually eligible for Medicare and Medicaid or LTSS users), the median percentage of stays that lasted 100 days or more was 18.7 in 2012; the highest share was in Louisiana (35 percent), and the lowest was in Arizona (8.9 percent) (Irvin et al. 2016). However, these state-level estimates might mask variation by programs across states, such as differences in 1915(c) HCBS waiver programs. It is also important to note that while this variation has been identified across states, it may be less relevant in an MLTSS environment where managed care plans are all operating within a single state, and subject to the same LTSS requirements.

Citations

Felix, Holly C., Glen P. Mays, M. Kathryn Stewart, Naomi Cottoms, and Mary Olson. “Medicaid Savings Resulted When Community Health Workers Matched Those With Needs to Home and Community Care.” Health Affairs, vol. 30, no. 7, 2011, pp. 1366-1374.

Greiner, Melissa A., Laura G. Qualls, Isao Iwata, Heidi K. White, Sheila L. Molony, M. Terry Sullivan, Bonnie Burke, Kevin A. Schulman, and Soko Setoguchi. "Predicting Nursing Home Placement Among Home- and Community-Based Services Program Participants." The American Journal of Managed Care, vol. 20, no. 12, 2014, pp. e535-e536.

Guo, Jing, R. Tamara Konetzka, Elizabeth Magett, and Willian Dale. "Quantifying Long-Term Care Preferences." Medical Decision Making, vol. 35, no. 1, 2015, pp. 106-113.

Irvin, Carol, Noelle Denny-Brown, Eric Morris, and Claire Postman. Table 5. Indicators of performance of state long-term services and supports systems, "Pathways to Independence: Transitioning Adults Under Age 65 from Nursing Home to Community Living (), table 5, Indicators of performance of state long-term services and supports systems. Cambridge, MA: Mathematica Policy Research, 2016, pp. 30-31. Available at https://www.medicaid.gov/medicaid/ltss/downloads/mfpfieldreport19.pdf.

Keenan, Teresa A. "Home and Community Preferences of the 45+ Population." Washington, DC: AARP Public Policy Institute, November 2010. Available at http://assets.aarp.org/rgcenter/general/home-community-services-10.pdf.

Sands, Laura P., Huiping Xu, Joseph Thomas, III, Sudeshna Paul, Bruce A. Craig, Marc Rosenman, Caroline C. Doebbeling, and Michael Weiner. "Volume of Home- and Community-Based Services and Time to Nursing-Home Placement." Medicare & Medicaid Research Review, vol. 2, no. 3, 2012, pp. E1-E21.

S.4. Numerator Statement: The count of discharges from an institutional facility to the community that occurred within 100 days or less from admission and resulted in successful discharge to the community (community residence for 60 or more days).

S.6. Denominator Statement: New admissions to an institutional setting for MLTSS enrollees age 18 and older.

S.8. Denominator Exclusions: None.

De.1. Measure Type: Outcome

S.17. Data Source: Claims, Enrollment Data

S.20. Level of Analysis: Health Plan

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

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? Not applicable.

1. Evidence and Performance Gap – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

2._MinInstit_EvidenceAttachment_9.19.18.docx

1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission?

Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** N/A

Measure Title: Minimizing institutional length of stay

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

Date of Submission: 11/1/2018

Instructions

- *Complete 1a.1 and 1a.2 for all measures. If instrument-based measure, complete 1a.3.*
- *Complete EITHER 1a.2, 1a.3 or 1a.4 as applicable for the type of measure and evidence.*
- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *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.
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

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

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Outcome: 3 Empirical data demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service. If not available, wide variation in performance can be used as evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence 4 that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: 5 a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence 4 that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence 4 that the measured structure leads to a desired health outcome.
- Efficiency: 6 evidence not required for the resource use component.
- For measures derived from patient reports, evidence should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.
- Process measures incorporating Appropriate Use Criteria: See NQF's guidance for evidence for measures, in general; guidance for measures specifically based on clinical practice guidelines apply as well.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines and/or modified GRADE.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to

the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.
6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

☒ Outcome: The proportion of admissions to an institution for managed long-term services and supports (MLTSS) plan enrollees that result in successful discharge to the community (community residence for 60 or more days) within 100 days of admission. This measure is reported as an observed rate and a risk-adjusted rate.

☐Patient-reported outcome (PRO):

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (*A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)*

☐ Intermediate clinical outcome (*e.g., lab value*):

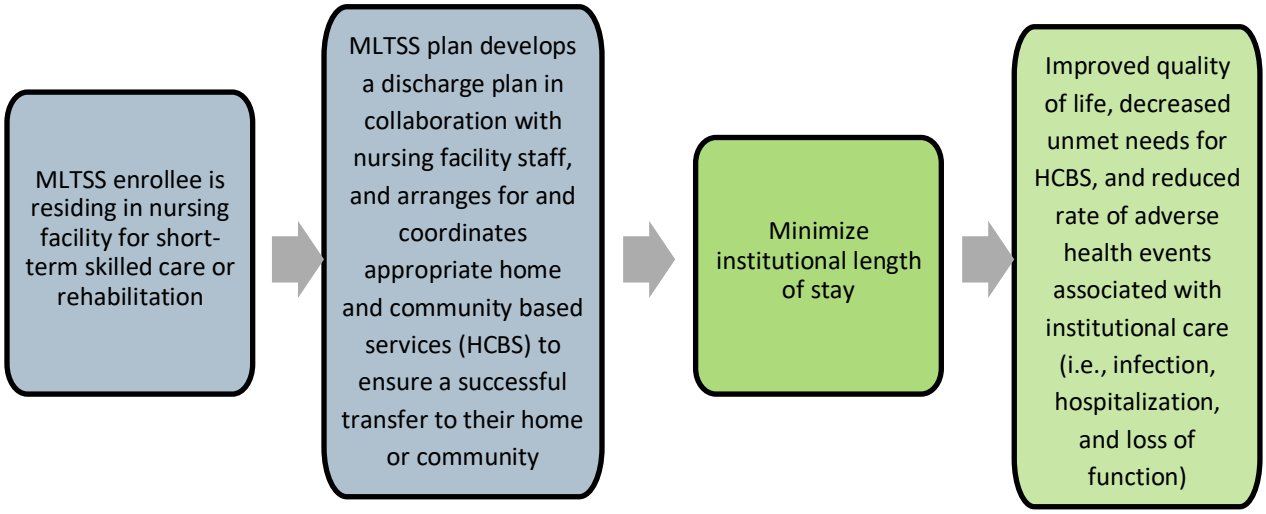
☐ Process:

☐ Appropriate use measure:

☐ Structure:

☐ Composite:

1a.2 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.



Most individuals who need long-term services and supports (LTSS) want to live at home or in community settings (Guo et al. 2015; Keenan 2010). Although state Medicaid agencies, which pay for the majority of LTSS, have made significant progress in rebalancing their LTSS systems to provide more HCBS, many Medicaid beneficiaries still reside in institutions, which is both costly and can be associated with adverse outcomes such as hospital readmissions. A key goal of MLTSS programs is to reduce the use of institutional care by decreasing unnecessary admissions, minimizing lengths of stay, and helping long-term residents return to the community if that is where they wish to live and receive services and long-stays. MLTSS plans achieve this goal by conducting person-centered assessment and care planning, providing timely access to high quality HCBS, and coordinating LTSS and medical care across providers and settings (Felix et al. 2011; Greiner et al. 2013; Sands et al. 2012). While it may not be possible to avoid the use of institutional care entirely, this measure assesses the effectiveness of MLTSS plans in minimizing avoidable admissions and reducing long stays in institutions when the same services can be provided in the home or community settings. Keeping individuals in home and community settings and minimizing their stays in institutions helps to increase their quality of life and care experience.

Citations

Felix, Holly C., Glen P. Mays, M. Kathryn Stewart, Naomi Cottoms, and Mary Olson. "Medicaid Savings Resulted When Community Health Workers Matched Those With Needs to Home and Community Care." *Health Affairs*, vol. 30, no. 7, 2011, pp. 1366-1374.

Greiner, Melissa A., Laura G. Qualls, Isao Iwata, Heidi K. White, Sheila L. Molony, M. Terry Sullivan, Bonnie Burke, Kevin A. Schulman, and Soko Setoguchi. "Predicting Nursing Home Placement Among Home- and Community-Based Services Program Participants." *The American Journal of Managed Care,* vol. 20, no. 12, 2014, pp. e535-e536.

Guo, Jing, R. Tamara Konetzka, Elizabeth Magett, and Willian Dale. "Quantifying Long-Term Care Preferences." *Medical Decision Making*, vol. 35, no. 1, 2015, pp. 106-113.

Keenan, Teresa A. "Home and Community Preferences of the 45+ Population." Washington, DC: AARP Public Policy Institute, November 2010. Available at http://assets.aarp.org/rgcenter/general/home-community-services-10.pdf.

Sands, Laura P., Huiping Xu, Joseph Thomas, III, Sudeshna Paul, Bruce A. Craig, Marc Rosenman, Caroline C. Doebbeling, and Michael Weiner. "Volume of Home- and Community-Based Services and Time to Nursing-Home Placement." *Medicare & Medicaid Research Review*, vol. 2, no. 3, 2012, pp. E1-E21.

1a.3 Value and Meaningfulness: IF this measure is derived from patient report, provide evidence that the target population values the measured ***outcome, process, or structure*** and finds it meaningful. (Describe how and from whom their input was obtained.)

Not applicable.

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ** (1a.2 has been completed)**

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES - Provide empirical data demonstrating the relationship between the outcome (or PRO) to at least one healthcare structure, process, intervention, or service.

Background on Managed Long-Term Services and Supports (LTSS) Plans: LTSS includes a wide range of care offered in both the institutional (e.g., nursing facility) and community (e.g., home and community based services) setting for adults who needs some assistance with daily tasks due to disability, aging or chronic illness. LTSS is primarily non-medical in nature – many of the services provided are intended to maximize individual's independence living in their preferred environment. The majority of older adults and people with disabilities prefer to receive LTSS in home or community settings, rather than institutions (Guo et al. 2015; Keenan 2010), and states have an obligation under the Supreme Court *Olmstead* decision to provide LTSS in the most integrated setting appropriate to the needs of qualified beneficiaries (Olmstead v. L.C. 1195.Ct. 2176, 1999).

Medicaid is the largest payer for LTSS in the U.S. and most states have traditionally provided LTSS under fee-for-service (FFS) arrangements with providers, facilities and community-based organizations. In recent years, however, state Medicaid agencies have shifted LTSS from FFS to managed care delivery models, in which states contract with managed care plans to deliver LTSS. States may contract with managed LTSS (MLTSS) plans that only provide LTSS, or comprehensive managed care plans that provide both LTSS and medical care. Almost half of all states (24) provide LTSS through either type of MLTSS arrangement in 2017, 50 percent more than the 16 states that did so in 2012 (Lewis et al. 2018). States have adopted MLTSS to achieve several goals, including improved participant outcomes and quality of care, increased access to HCBS, and improved care coordination. However, if not well-designed, MLTSS could disrupt longstanding relationships (e.g. if enrollees' providers are not part of the managed care plan's network) and create barriers to obtaining needed care (e.g., through gatekeeping or coverage restrictions). Consequently, it is important to systematically measure the quality of care delivered to people in MLTSS plans and their effectiveness in helping individuals with disability to live in the community.

There are currently no NQF endorsed measures of MLTSS quality, and states with MLTSS programs generally do not use reliable, validated LTSS quality measures. People who receive LTSS typically have chronic conditions and their functional ability is likely to decline over time due to the nature of their disability or age. Thus, outcomes such as improvements in health status and function are not applicable to MLTSS enrollees; instead, outcomes such as improvement in quality of life, community integration, and avoidance or delay of institutionalization are more relevant and important (MACPAC 2018).

The intended use of the measure is to allow states to compare and evaluate the quality of LTSS care being provided in MLTSS plans with which they contract using a common, standardized and validated measure. NQF confirmed the importance of developing measures of the degree to which people who need LTSS are served in home and community settings, citing current state-specific measures that are similar to this measure, but vary by state and have not been rigorously tested (NQF 2016, p. 34-35). This specific measure focuses on one critical outcome of high quality MLTSS care – minimizing the time spent in institutional settings.

Evidence to support relationship between LTSS services and successful transition from the institutional setting to community setting (outcome): This measure evaluates how well MLTSS plans can minimize length of stay in institutions for MLTSS enrollees (i.e., keep stays to less than 100 days). This is important to MLTSS enrollees who are newly admitted to institutional facilities for temporary or post-acute care skilled nursing needs, and can return to their home or a community setting with appropriate support. The 100-day marker is generally considered a cross-over point because permanent residence in the institutional facility becomes more likely because individuals often lose community-based supports and housing after that time (Arling et al. 2010). In addition, MLTSS enrollees who are dually eligible for Medicare and Medicaid lose Medicare coverage for post-acute care in a nursing facility after 100 days, so MLTSS plans (as Medicaid payers for MLTSS enrollees) become wholly responsible for paying for LTSS and have substantial ability to determine where enrollees receive such services and supports.

With adequate discharge planning and high-quality community-based LTSS, individuals can successfully transition to community settings from an institution. Evidence from the national evaluation of the Money Follows the Person (MFP) Demonstration program shows that individuals can successfully transition to community settings from institutions, even after long-stays in an institution. MFP participants – those who resided in an institution for 90 days or more and moved to a home or community setting with HCBS support -- had fewer unmet needs for personal assistance, and they faced fewer barriers to integrating into the community (Irvin et al. 2015). Compared to a matched comparison group of people who transitioned to the community without the help of MFP services and transition coordinators, MFP participants had lower odds of reinstitutionalization of 30 days or more, and lower odds of any institutional use within 6 months of transition (Irvin et al. 2017). In addition, older MFP participants (age 65 and older) were significantly less likely to be hospitalized than those who transition to the community without MFP assistance, and after discharge, were more likely to use home health. States with the highest numbers of MFP transitions, and the lowest rates of reinstitutionalization, took a systematic approach to the transition process: identifying those who wanted to

return to the community, comprehensive assessment of their needs and risks before transition, providing specialized services and supports (e.g., home modifications) that increase the likelihood of successful transitions, and robust monitoring to ensure services are delivered in a timely way (Denny-Brown et al. 2015).

Among individuals who had been receiving HCBS prior to their admission, it is often easier to re-establish some of those services after a short interruption than having to create an entirely new care plan with new providers many months or years later. In addition, housing options typically diminish the longer an individual remains in an institution, making it more difficult for institutional residents to be discharged (Arling et al. 2010) , so it is critical to focus on discharge planning early in the institutional stay when the person may be able to return to the same home or apartment. When individuals need financial assistance to afford an apartment that is accessible, a rigorous study found that housing vouchers that subsidize rental costs combined with coordination of HCBS increased community transition rates (Hoffman et al. 2017).

Evidence suggests that states with less investment in HCBS and higher rates of nursing home use have higher proportions of "low-care" residents (i.e., individuals who do not require skilled nursing services and could reasonably be living in community settings with support) and lower rates of discharge to the community (Mor et al. 2007; Arling et al. 2011). Low-care nursing home residents are particularly good candidates for discharge to the community, and the prevalence of low-care residents in institutions indicates an opportunity for improvement (Ross et al. 2012).

Citations

Arling, Greg, Robert L. Kane, Valerie Cooke, and Teresa Lewis. "Targeting Residents for Transitions from Nursing Home to Community." *Health Services Research*, vol. 45, no. 3, June 2010, pp. 691-711.

Arling, Greg, Kathleen A. Abrahamson, Valerie Cooke, Robert L. Kane, and Teresa Lewis. "Facility and Market Factors Affecting Transitions from Nursing Home to Community." *Medical Care*, vol. 49, no. 9, September 2011, pp. 790–796.

Denny-Brown, N., B. Hagen, C. Bradnan, and S. Williams. "The Right Supports at the Right Time: How Money Follows the Person Programs Are Supporting Diverse Populations in the Community." National Evaluation of the MFP Program, Report from the Field, Number 16, April 2015.

Hoffman, D., M. Kehn, and D. Lipson. "The Missing Link: Examining the Impact of Housing Vouchers and Community-Based Services and Supports on Transitions From Nursing Facilities to the Community." *Journal of Disability Policy Studies*, 27(4) 243–251, 2017.

Irvin, Carol V., Alex Bohl, Kate Stewart, Susan R. Williams, Allison Steiner, Noelle Denny-Brown, Andrea Wysocki, Rebecca Coughlin, Jason Smoot, and Victoria Peebles. "Money Follows the Person 2015 Annual Evaluation Report." Cambridge, MA: Mathematica Policy Research, May 11, 2017.

Irvin, C. V., N. Denny-Brown, A. Bohl, J. Schurrer, A. Wysocki, R. Coughlin, and S. R. Williams. "Money Follows the Person 2014 Annual Evaluation Report." Mathematica Policy Research, Cambridge, MA. December 18, 2015.

Lewis, E., S. Eiken, A. Amos, and P. Saucier. 2018. The growth of managed long-term services and supports programs: 2017 update. Cambridge, MA: Truven Health Analytics, IBM Watson Health. https://www.medicaid.gov/medicaid/managed-care/downloads/ltss/mltssp-inventory-update-2017.pdf

Medicaid and CHIP Payment and Access Commission (MACPAC). "Managed Long-Term Services and Supports: Status of State Adoption and Areas of Program Evolution." Report to Congress on Medicaid and CHIP, June 2018. Chapter 3. https://www.macpac.gov/publication/managed-long-term-services-and-supports-status-of-state-adoption-and-areas-of-program-evolution/

Mor, Vincent, Jacqueline Zinn, Pedro Gozalo, Zhanlian Feng, Orna Intrator, and David C. Grabowski. "Prospects for Transferring Nursing Home Residents to the Community." *Health Affairs*, vol. 26, no. 6, 2007, pp. 1762–1771.

National Quality Forum. "Quality in Home and Community-Based Services to Support Community Living: Addressing Gaps in Performance Measurement." Final Report, September 2016. http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=83433.

Ross, Jessica, Sam Simon, Carol Irvin, and Dean Miller. "Institutional Level of Care Among Money Follows the Person Participants." Mathematica Policy Research, Cambridge, MA. The National Evaluation of the Money Follows the Person Demonstration Grant Program, Reports from the Field, Number 10, October 2012.

1a.3. SYSTEMATIC REVIEW(SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES, INCLUDING THOSE THAT ARE INSTRUMENT-BASED) If the evidence is not based on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

What is the source of the systematic review of the body of evidence that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (IOM)

☐Clinical Practice Guideline recommendation (with evidence review)

☐US Preventive Services Task Force Recommendation

☐Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*)

☐Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	

Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

_____

1a.4 OTHER SOURCE OF EVIDENCE

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

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

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

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

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

1b.1. Briefly explain the rationale for this measure *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

<u>If a COMPOSITE</u> (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.

This measure evaluates how well MLTSS plans can minimize length of stay in institutions for MLTSS enrollees (i.e., keep stays to less than 100 days). This is important to MLTSS enrollees who are newly admitted to institutional facilities for temporary skilled nursing needs, but could live independently in a community setting with appropriate support. The 100-day cut point is generally considered a cross-over point because after that time (1) residence in the institutional facility becomes considered semi-permanent, (2) dual eligible enrollees lose Medicare coverage for their institutional stay, and (3) they often lose any community-based housing they were previously using further limiting the probability they will ever return to the community.

Most individuals who need long-term services and supports (LTSS) want to live at home or in community settings (Guo et al. 2015; Keenan 2010). Although state Medicaid agencies, which pay for the majority of LTSS, have made significant progress in rebalancing their LTSS systems to provide more HCBS, many Medicaid beneficiaries still reside in institutions, which is both costly and can be associated with adverse outcomes such as hospital readmissions. A key goal of MLTSS programs is to minimize lengths of stay in institutions. MLTSS plans achieve this goal by conducting person-centered assessment and care planning, providing timely access to high quality HCBS, and coordinating LTSS and medical care across providers and settings (Felix et al. 2011; Greiner et al. 2013; Sands et al. 2012). While it may not be possible to avoid the use of institutional care entirely, this measure assesses the effectiveness of MLTSS plans in minimizing length of stay in institutions when the same services can be provided in the home or community settings. Keeping individuals in home and

community settings and minimizing their stays in institutions helps to increase their quality of life and care experience.

Finally, there is some evidence from the general population (not limited to Medicaid or LTSS users) suggesting the duration of institutional stays among LTSS users may vary by state. Among all nursing home admissions of Medicare beneficiaries in 2012 (not limited to those dually eligible for Medicare and Medicaid or LTSS users), the median percentage of stays that lasted 100 days or more was 18.7 in 2012; the highest share was in Louisiana (35 percent), and the lowest was in Arizona (8.9 percent) (Irvin et al. 2016). However, these state-level estimates might mask variation by programs across states, such as differences in 1915(c) HCBS waiver programs. It is also important to note that while this variation has been identified across states, it may be less relevant in an MLTSS environment where managed care plans are all operating within a single state, and subject to the same LTSS requirements.

Citations

Felix, Holly C., Glen P. Mays, M. Kathryn Stewart, Naomi Cottoms, and Mary Olson. "Medicaid Savings Resulted When Community Health Workers Matched Those With Needs to Home and Community Care." Health Affairs, vol. 30, no. 7, 2011, pp. 1366-1374.

Greiner, Melissa A., Laura G. Qualls, Isao Iwata, Heidi K. White, Sheila L. Molony, M. Terry Sullivan, Bonnie Burke, Kevin A. Schulman, and Soko Setoguchi. "Predicting Nursing Home Placement Among Home- and Community-Based Services Program Participants." The American Journal of Managed Care, vol. 20, no. 12, 2014, pp. e535-e536.

Guo, Jing, R. Tamara Konetzka, Elizabeth Magett, and Willian Dale. "Quantifying Long-Term Care Preferences." Medical Decision Making, vol. 35, no. 1, 2015, pp. 106-113.

Irvin, Carol, Noelle Denny-Brown, Eric Morris, and Claire Postman. Table 5. Indicators of performance of state long-term services and supports systems, "Pathways to Independence: Transitioning Adults Under Age 65 from Nursing Home to Community Living (), table 5, Indicators of performance of state long-term services and supports systems. Cambridge, MA: Mathematica Policy Research, 2016, pp. 30-31. Available at https://www.medicaid.gov/medicaid/ltss/downloads/mfpfieldreport19.pdf.

Keenan, Teresa A. "Home and Community Preferences of the 45+ Population." Washington, DC: AARP Public Policy Institute, November 2010. Available at http://assets.aarp.org/rgcenter/general/home-community-services-10.pdf.

Sands, Laura P., Huiping Xu, Joseph Thomas, III, Sudeshna Paul, Bruce A. Craig, Marc Rosenman, Caroline C. Doebbeling, and Michael Weiner. "Volume of Home- and Community-Based Services and Time to Nursing-Home Placement." Medicare & Medicaid Research Review, vol. 2, no. 3, 2012, pp. E1-E21.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

These data are from four health plans, representing 14 health plan product lines (HPPLs) from 10 states that participated in testing these measures in 2017. Participating plans used 24 months of data from either calendar years 2014 and 2015 or 2015 and 2016 (varied by plan). Full details of this testing effort are included in the testing attachment.

Below we present the average, standard deviation, median, minimum and maximum observed and risk-adjusted rates across HPPLs. We do not present deciles given the limited number of plans included in the testing sample. Additional details on the individual plan rates and meaningful difference in performance across plans is available in the testing attachment, section 2b4.

Observed Performance:

Average	Stdev	Median	Min	Max
36.08	19.53	37.01	0.00	75.00

Risk-adjusted performance:

Average	Stdev	Median	Min	Max
33.60	16.92	36.49	0.00	65.88

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

Not applicable, performance data provided above demonstrating gap.

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

We did not test differences in performance by race/ethnicity in the testing data, however age, gender, and dual status are included in the statistical risk model.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4

The Congressional Budget Office identified racial and ethnic disparities in the need for LTSS. More specifically, it found that older black and Hispanic individuals have higher rates of functional impairment than whites (Congressional Budget Office 2013). Another report identified higher incidence of complex care needs, as well as greater need for care coordination, among California Medicaid beneficiaries age 65 and over or with disabilities (excluding Medicare-Medicaid dual eligibles) compared to Medicaid beneficiaries under age 65 and non-disabled, among those who transitioned from FFS to Medicaid managed care covering acute, primary and specialty services (LTSS were carved out) (KFF, 2013).

We were unable to find much evidence on disparities in the quality of LTSS care provided to minority populations. This is likely due to the lack of available valid and reliable quality measure of LTSS care. We believe that use of this measure will help to identify if there are disparities in LTSS quality of care for minority populations. However, the two studies that were identified suggest that disparities do exist. One study of dually eligible Medicaid HCBS recipients age 65 and older found that blacks and Hispanics were less likely than whites to have a nursing home stay; they remained in the community longer and were more physically and cognitively impaired upon admission (Cai and Temkin-Greener 2015). Still another study of Medicaid enrollees found that those newly admitted to nursing homes were more likely to be older white women (Schmitz et al. 2014), which is consistent with the findings for Medicaid HCBS users. Taken together, these two studies suggest there may be racial/ethnic disparities in the use of nursing homes and/or cultural differences that lead to variance in the settings where people receive long-term care.

Citations

Cai, Xueya, and Helena Temkin-Greener. "Nursing Home Admissions Among Medicaid HCBS Enrollees: Evidence of Racial/Ethnic Disparities or Differences?" Medical Care, vol. 53, no. 7, 2015, pp. 566-573.

Congressional Budget Office. (2013). Rising Demand for Long-Term Services and Supports for Elderly People. Washington, DC: Congressional Budget Office.

Kaiser Family Foundation, KFF (2013). Issue Brief. Transitioning Beneficiaries with Complex Care Needs to Medicaid Managed Care: Insights from California. Available at

https://kaiserfamilyfoundation.files.wordpress.com/2013/06/8453-transitioning-beneficiaries-with-complex-care-needs2.pdf.

Schmitz, Robert, Victoria Peebles, Rosemary Borck, and Miller, Dean. "Medicaid-Financed Institutional Services: Patterns of Care for Residents of Nursing Homes and Intermediate Care Facilities for Individuals with Intellectual Disabilities in 2008 and 2009." Washington, DC: Office of the Assistant Secretary for Planning and Evaluation, May 2014. Available at https://aspe.hhs.gov/sites/default/files/pdf/137851/CarePatt.pdf.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. Subject/Topic Area *(check all the areas that apply):*

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

De.7. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*

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.)*

Technical specifications are available at https://www.medicaid.gov/medicaid/managed-care/ltss/index.html. In addition, CMS provides implementation support via a technical assistance mailbox.

S.2a. If this is an eMeasure, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)

This is not an eMeasure **Attachment:**

S.2b. Data Dictionary, Code Table, or Value Sets *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*

Attachment **Attachment:** 4._MinInstit_ValueSet_RiskAdjTables_7.30.18.xlsx

S.2c. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

No, this is not an instrument-based measure **Attachment:**

S.2d. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

Not an instrument-based measure

S.3.1. For maintenance of endorsement: Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

No

S.3.2. For maintenance of endorsement, please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.

Not applicable.

S.4. Numerator Statement *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*

IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

The count of discharges from an institutional facility to the community that occurred within 100 days or less from admission, and resulted in successful discharge to the community (community residence for 60 or more days).

S.5. Numerator Details *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

The count of discharges from an institutional facility to the community between July 1 of the year prior to the measurement year* and October 31 of the measurement year that occurred within 100 days or less of admission. Discharges that result in death, hospitalization or re-admission to the institution within 60 days of discharge from the institution do not meet the numerator criteria.

The measure focuses on discharges within 100 days because it is generally considered a cross-over point in long-term care. After that time, evidence shows that: (1) residence in the institutional facility becomes considered semi-permanent, (2) dual eligible enrollees lose Medicare coverage for their institutional stay, and (3) they often lose any community-based housing they were previously using further limiting the probability they will ever return to the community.

Institutional facility: Medicaid- or Medicare- certified nursing facilities providing skilled nursing/medical care; rehabilitation needed due to injury, illness or disability; and long-term care (also referred to as "custodial care") or Medicaid certified Intermediate Care Facility for Individuals with Intellectual Disabilities (ICF/IID). (see Institutional Facility Value Set).

Community residence: Any residence that is not an institutional facility (see definition above). Note community residence may include assisted living, adult foster care, or other care in another setting that is not defined as an institution.

*The measurement year is January 1 through December 31, i.e., is equivalent to the calendar year.

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

New admissions to an institutional setting for MLTSS enrollees age 18 and older.

S.7. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

New admissions to an institutional setting between July 1 of the year prior to the measurement year and June 30 of the measurement year among MLTSS enrollees age 18 and older who receive both medical and LTSS benefits through the accountable health plan (see value set Institutional Facility).

Include all admissions to the institutional setting directly from the community. Include admissions to the institutional setting from the hospital setting only if the MLTSS enrollee lived in the community prior to the hospital admission. These are considered "new admissions."

Do not include admissions to the institutional setting from the hospital setting if the MLTSS enrollee was residing in an institution prior to the hospital admission. Do not include admissions to the institutional setting that are transfers from another institution. These admissions are NOT considered "new admissions."

Do not include admissions where the MLTSS enrollee dies in the institution, dies within one day of discharge from the institution or is discharged to a hospital and dies in the hospital between July 1 of the year prior to the measurement period. Due to differences in coding practices, death within one day of discharge is considered a death in the institution. These admissions are considered admissions where there was not opportunity for discharge (i.e., death occurred within 100 days of admission) or the individual was near end of life and discharge may not have been clinically appropriate.

Do not include admissions where the MLTSS enrollee was discharged to a hospital between July 1 of the year prior to the measurement year and remained in the hospital until the end of the measurement year.

Do not include admissions for MLTSS enrollees who were not continuously enrolled in the MLTSS plan on the day of the new admission through 160 days following the new admission date.

An enrollee can be counted more than once in the denominator if the individual had more than one admission to an institutional setting during the measurement year.

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

None.

S.9. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

Not applicable.

S.10. Stratification Information *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*

None.

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

Statistical risk model

If other:

S.12. Type of score:

Rate/proportion

If other:

S.13. Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*

Better quality = Higher score

S.14. Calculation Algorithm/Measure Logic (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*

CALCULATION OF THE OBSERVED RATE DENOMINATOR - STEPS TO IDENTIFY INSTITUTION ADMISSIONS FROM THE COMMUNITY

Step 1: Identify all admissions to institutional facilities between July 1 of the year prior to the measurement period and June 30 of the measurement year.

Step 2: Remove admissions that are transfers from another institution.

Step 3: Remove admissions from the hospital that originated from an institution.

Step 4: Remove admissions that result in death in the institution or death within 1 day of discharge from the institution.

Step 5: Remove admissions for MLTSS enrollees who were not continuously enrolled in the MLTSS plan on the day of the new admission through 160 days following the new admission date. All resulting admissions directly from the community and from the hospital that originated in the community make up the denominator for the observed rate.

CALCULATING THE RISK ADJUSTMENT WEIGHTS FOR THE DENOMINATOR

For each qualified admission in the denominator, use the following steps to identify risk adjustment weights based on dual eligibility, age and gender, diagnoses from the qualified admission, and number of hospital stays and months of enrollment in the classification period. Risk adjustment weights are provided in the attached Excel file.

Step 1: Identify the base weight

Step 2: Link the age and gender weights for each qualified admission

Step 3: For each qualified admission with dual eligibility, link the dual eligibility weight.

Step 4: For each qualified admission with a Chronic Conditions Warehouse (CCW) category, link the qualified CCW category weight.

Step 5: For each qualified admission with 1 or more hospitalizations prior to qualified admission, link the number of hospitalization weight.

Step 6: For each qualified admission with six months or more of enrollment prior to the qualified admission, link the six months enrollment weight.

Step 7: Sum all weights associated with the qualified admission (i.e., base, age and gender, dual eligibility, qualified CCW categories, number of hospitalizations, and six months enrollment weight) to calculate the expected estimated probability of Successful Discharge to the Community for each qualified admission.

Expected Discharge Probability = [exp (sum of weights for IFA)]/[1+exp(sum of weights for IFA)]

Note: “Exp” refers to the exponential or antilog function.

Step 8: Calculate the count of successful discharges to the community. The count of expected discharges is the sum of the estimated discharge probability calculated in Step 7 for each qualified admission.

Count of Expected Discharges = Sum of (estimated discharge probability for each qualified admission)

CALCULATION OF OBSERVED RATE NUMERATOR – STEPS TO IDENTIFY DISCHARGES TO THE COMMUNITY WITH A LENGTH OF STAY (LOS) OF LESS THAN OR EQUAL TO 100 DAYS

Step 1: Identify all qualified admissions (see Denominator criteria above).

Step 2: Look for location of the first discharge for each qualified admission between July 1 of the year prior to the measurement year and October 31 of the measurement year.

? If the enrollee is discharged to the community, calculate LOS as the date of institution discharge minus the index admission date.

? If there is no discharge, calculate LOS as the date of the last day of the measurement year minus index admission date.

? If the enrollee is discharged to the hospital, look for the hospital discharge and location of discharge. If the enrollee is discharged from the hospital to the community, calculate LOS as the date of institution discharge minus the qualified index admission date.

? If the enrollee is discharged to the hospital and dies in the hospital, exclude the admission from the qualified index admission.

? If the enrollee is discharged to the hospital and remains in the hospital at the end of the measurement year, exclude the admission from the qualified index admission.

? If the enrollee is discharged from the hospital to the institution, repeat step 2 until there is a discharge to the community or the end of the measurement period.

? If the enrollee is discharged to a different institution (i.e. a transfer), repeat step 2 until there is a discharge to the community or the end of the measurement period.

? When counting the duration of each stay within a measurement period, include the day of entry (admission) but not the day of discharge unless the admission and discharge occurred on the same day in which case the number of days in the stay is equal to 1.

Step 3: Using information from step 2, identify all qualified admissions with length of stay of less than or equal to 100 days. This should include only discharges to the community (either directly from the institution or from the institution to the hospital to the community).

Step 4: Remove discharge if the MLTSS enrollee was hospitalized, died or was re-admitted to the institution within 60 days of the day of discharge.

CALCULATION OF OBSERVED PERFORMANCE RATE

Calculate the observed discharge rate by dividing the the numerator (step 4 under numerator) by the denominator (step 5 under denominator).

CALCULATION OF RISK-ADJUSTED RATE: EXPECTED RATE

Calculate the expected discharge rate by dividing the expected count of successful discharges by the denominator (count of new admissions). Report the expected discharge rate as the expected performance rate of the Minimizing Institutional Length of Stay measure.

Plans can understand their results by calculating the ratio of their observed to expected (O/E) rates. A ratio of greater than 1 implies a higher than expected rate of successful discharges, whereas a ratio of less than 1 implies lower than expected rate of successful discharges.

CALCULATION OF THE RISK ADJUSTED RATE: RISK ADJUSTED PERFORMANCE RATE

Reporting of a risk-adjusted rate requires standardization of the O/E ratio using a multi-plan, population rate.

States should calculate the multi-plan population rate by taking the sum of all observed numerator events and dividing by the sum of all observed denominator events.

The risk-adjusted rate of Minimizing Institutional Length of Stay for each plan is calculated by multiplying the plan O/E ratio by the multi-plan population rate.

Plan Risk Adjusted Rate = O/E Ratio x Multi-plan population rate

S.15. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*

IF an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.

Not applicable.

S.16. Survey/Patient-reported data *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*

Specify calculation of response rates to be reported with performance measure results.

Not applicable.

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

If other, please describe in S.18.

Claims, Enrollment Data

S.18. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data are collected.)*

IF instrument-based, identify the specific instrument(s) and standard methods, modes, and languages of administration.

Not applicable.

S.19. Data Source or Collection Instrument *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

No data collection instrument provided

S.20. Level of Analysis *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*

Health Plan

S.21. Care Setting *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*

Home Care, Other, Post-Acute Care

If other: Nursing Home/Skilled Nursing Facility, ICF/IID, Community Settings

S.22. COMPOSITE Performance Measure - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*

Not applicable.

2. Validity – See attached Measure Testing Submission Form

3._MinInstit_TestingAttachment_7.30.18.docx

2.1 For maintenance of endorsement

Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

No

2.2 For maintenance of endorsement

Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

No

2.3 For maintenance of endorsement

Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes social risk factors is not prohibited at present. Please update sections 1.8, 2a2, 2b1,2b4.3 and 2b5 in the Testing attachment and S.140 and S.11 in the online submission form. NOTE: These sections must be updated even if social risk factors are not included in the risk-adjustment strategy. You MUST use the most current version of the Testing Attachment (v7.1) -- older versions of the form will not have all required questions.

Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:** N/A
Measure Title: Minimizing institutional length of stay
Date of Submission: 8/1/2018

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.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: **(*must be consistent with data sources entered in S.17*)**	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ claims	☒ claims
☐ registry	☐ registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other:	☐ other:

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

Not applicable. There are no existing, nationally standardized datasets for Medicaid beneficiaries enrolled in managed long-term services and supports (MLTSS) plans, which is the target population for this measure. Therefore, we worked directly with health plans to obtain the enrollment and claims data needed to support measure testing. These data represented 4 parent health plan organizations, and 14 different health plan product lines (HPPLs) from 10 states, located in geographically diverse regions of the country. Health plans are anticipated to calculate this measure utilizing their own data, similar to reporting for the Healthcare Effectiveness Data and Information Set (HEDIS) plan-level meaures.

1.3. What are the dates of the data used in testing? MLTSS plans participating in field testing provided 24 months of data from either calendar years 2014 and 2015 or 2015 and 2016 (varied by plan) to support measure testing.

1.4. What levels of analysis were tested? (*testing must be provided for <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: **(*must be consistent with levels entered in item S.20*)**	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☒ health plan	☒ health plan
☐ other:	☐ other:

1.5. How many and which <u>measured entities</u> 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*)

Background on Managed Long-Term Services and Supports (LTSS) Plans: LTSS includes a wide range of care offered in both the institutional (e.g., nursing facility) and community (e.g., home and community based services) setting for adults who needs some assistance with daily tasks due to disability, aging or chronic illness. LTSS is primarily non-medical in nature – many of the services provided are intended to maximize individual's independence living in their preferred environment. The majority of older adults and people with disabilities prefer to receive LTSS in home or community settings, rather than institutions (Guo et al. 2015; Keenan 2010), and states have an obligation under the Supreme Court *Olmstead* decision to provide LTSS in

the most integrated setting appropriate to the needs of qualified beneficiaries (Olmstead v. L.C. 1195.Ct. 2176, 1999).[1] [2]

Medicaid is the largest payer for LTSS in the U.S. and most states have traditionally provided LTSS under fee-for-service (FFS) arrangements with providers, facilities and community-based organizations. In recent years, however, state Medicaid agencies have shifted LTSS from FFS to managed care delivery models, in which states contract with managed care plans to deliver LTSS. States may contract with managed LTSS (MLTSS) plans that only provide LTSS, or comprehensive managed care plans that provide both LTSS and medical care. Almost half of all states (24) provide LTSS through either type of MLTSS arrangement in 2017, 50 percent more than the 16 states that did so in 2012 (Lewis et al. 2018).[3] States have adopted MLTSS to achieve several goals, including improved participant outcomes and quality of care, increased access to home- and community-based services (HCBS), and improved care coordination. However, if not well-designed, MLTSS could disrupt longstanding relationships (e.g. if enrollees' providers are not part of the managed care plan's network) and create barriers to obtaining needed care (e.g., through gatekeeping or coverage restrictions). Consequently, it is important to systematically measure the quality of care delivered to people in MLTSS plans and their effectiveness in helping individuals with disability to live in the community.

There are currently no NQF endorsed measures of MLTSS quality, and states with MLTSS programs generally do not use reliable, validated LTSS quality measures. People who receive LTSS typically have chronic conditions and their functional ability is likely to decline over time due to the nature of their disability or age. Thus, outcomes such as improvements in health status and function are not applicable to MLTSS enrollees; instead, outcomes such as improvement in quality of life, community integration, and avoidance or delay of institutionalization are more relevant and important (MACPAC 2018).[4]

The intended use of the measure is to allow states to compare and evaluate the quality of LTSS care being provided by the MLTSS plans with which they contract. This specific measure focuses on a critical component of high quality MLTSS care – ensuring that care can be provided to the greatest extent possible in the community setting as opposed to the institutional setting. This measure evaluates how well MLTSS plans can minimize length of stay in institutions for MLTSS enrollees (i.e., keep stays to less than 100 days). This is important to MLTSS enrollees who are newly admitted to institutional facilities for temporary skilled nursing needs, but could live independently in a community setting with appropriate support.

The 100-day marker is generally considered a cross-over point at which permanent residence in the institutional facility becomes more likely because individuals often lose community-based supports and housing after that time (Arling et al. 2010).[5] In addition, MLTSS enrollees who are dually eligible for Medicare

---

[1] Guo, Jing, R. Tamara Konetzka, Elizabeth Magett, and Willian Dale. "Quantifying Long-Term Care Preferences." Medical Decision Making, vol. 35, no. 1, 2015a, pp. 106-113.

[2] Keenan, Teresa A. "Home and Community Preferences of the 45+ Population." Washington, DC: AARP Public Policy Institute, November 2010. Available at http://assets.aarp.org/rgcenter/general/home-community-services-10.pdf.

[3] Lewis, E., S. Eiken, A. Amos, and P. Saucier. 2018. The growth of managed long-term services and supports programs: 2017 update. Cambridge, MA: Truven Health Analytics, IBM Watson Health. https://www.medicaid.gov/medicaid/managed-care/downloads/ltss/mltssp-inventory-update-2017.pdf

[4] Medicaid and CHIP Payment and Access Commission (MACPAC). "Managed Long-Term Services and Supports: Status of State Adoption and Areas of Program Evolution." Report to Congress on Medicaid and CHIP, June 2018. Chapter 3. https://www.macpac.gov/publication/managed-long-term-services-and-supports-status-of-state-adoption-and-areas-of-program-evolution/

[5] Arling, Greg, Robert L. Kane, Valerie Cooke, and Teresa Lewis. "Targeting Residents for Transitions from Nursing Home to Community." Health Services Research, vol. 45, no. 3, June 2010, pp. 691-711.

and Medicaid lose Medicare coverage for post-acute care in a nursing facility after 100 days, so MLTSS plans (as Medicaid payers for MLTSS enrollees) become wholly responsible for paying for LTSS and have substantial ability to determine where enrollees receive such services and supports.

Testing Sample of MLTSS Plans: Unlike Medicare FFS and Medicaid FFS data, there is no national data source for MLTSS data. Therefore, to test measures, we had to recruit MLTSS plans to provide data for the purposes of testing. We recruited plans through national outreach to individual MLTSS plans, outreach to states operating MLTSS programs, and selected outreach to MLTSS plans that represented needed variation in our sample (i.e., we specifically targeted small MLTSS plans to ensure we had a mix of large and small MLTSS plans in our sample).

The four participating test plans included a mix of large and small plans (three national plans and one local plan), representing 14 Health Plan Product Lines (HPPLs) from 10 states, located in geographically diverse regions of the country (see Table 1).[6] Two HPPLs covered rural regions, 2 HPPLs covered an urban area, and 10 covered a mixed rural-urban region. The four plans submitted data from the following types of product lines: 6 Medicare-Medicaid Plans (MMPs), 1 Fully Integrated Dual Eligible Special Needs Plan (FIDE SNP), 1 Dual Eligible SNP (D-SNP) with a linked Medicaid Managed Care Organization (MCO) contract, and 6 comprehensive Medicaid MCOs for Medicaid-only beneficiaries, covering medical and LTSS benefits (see Table 2).[7]

Table 1. Health Plan Product Line Characteristics

Health Plan Product Line (HPPL)[8]	Program Type	Region	Population Type (Rural/Urban/Mix)	Number of enrollees
HPPL-01	FIDE-SNP	Midwest	Rural	1,277
HPPL-02	Medicaid MCO + LTSS	Midwest	Rural	797
HPPL-03	Medicaid MCO + LTSS	South	Mix	2,546
HPPL-04	MMP	West	Mix	30,152
HPPL-05	Medicaid MCO+LTSS	West	Mix	36,807
HPPL-06	Medicaid MCO+LTSS	Midwest	Mix	9,196
HPPL-07	MMP	Midwest	Mix	7,274
HPPL-08	MMP	Midwest	Urban	12,192
HPPL-09	Medicaid MCO+LTSS	West	Mix	4,241
HPPL-11	MMP	Midwest	Mix	5,771
HPPL-12	MMP	South	Mix	271
HPPL-13	Medicaid MCO+LTSS	South	Mix	51,555
HPPL-14	MMP	South	Mix	20,227

[6] U.S. Census regions were used to assess geographical diversity. https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf.

[7] Federal rules define a Medicaid comprehensive MCO as a risk contract between the State and an MCO that covers inpatient hospital services and nine other types of services, including nursing facility services, but not long-term home and community-based services. We use the term "comprehensive MCO + LTSS" to clarify that the plans participating in this measure testing cover acute care, as well as institutional and HCBS LTSS. §42 CFR 438.2

[8] The HPPL IDs are not consecutive. A fifth plan agreed to participate and was assigned HPPL-15 and HPPL-16, but this plan was unable to provide usable data in time for inclusion in this report. One plan was also assigned HPPL-10, but it was determined that this HPPL was not eligible to participate in testing because it did not have complete Medicare data for dually eligible beneficiaries.

Health Plan Product Line (HPPL)[8]	Program Type	Region	Population Type (Rural/Urban/Mix)	Number of enrollees
HPPL-17	MLTSS+D-SNP	South	Urban	7,416

Source: Mathematica analysis of MLTSS enrollees from four health plans and fourteen health plan product lines.

Notes: D-SNP = Medicare Advantage Dual Eligible Special Needs Plan; FIDE-SNP = Fully Integrated Dual Eligible Special Needs Plan; HPPL = Health Plan Product Line; LTSS = Long-term Services and Supports; MCO = Managed Care Organization; MMP = Medicare-Medicaid Plan

Table 2. Health Plan Product Line Information

Characteristic	Percentage of enrollees in the testing sample (n=189,722)
MLTSS Program	
Integrated **Medicare-Medicaid Plan (MMP)** for dual enrollees	43.5
Medicare Advantage Dual Eligible Special Needs Plan (D-SNP) for dual enrollees	0.0
Fully Integrated Dual Eligible (FIDE) Special Needs Plan (SNP) for dual enrollees	0.7
Medicaid comprehensive MCO, including medical and long-term services and supports (LTSS) benefits, for Medicaid-only beneficiaries	51.9
D-SNP with linked MLTSS plan	3.9

Source: Mathematica analysis of MLTSS enrollees from four health plans and fourteen health plan product lines.

1.6. How many and which <u>patients</u> 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*)

MLTSS enrollees are, by definition, Medicaid enrolled adults who are eligible for receiving long-term services and supports[9]. Each state defines criteria for eligibility for LTSS differently, but in general enrollees must meet some criteria for being at a nursing-home level of care. Since the intended use of this measure is evaluation of MLTSS plans within a single state, the differences between states' Medicaid and LTSS eligibility should not impact comparison of results.

Testing Sample of MLTSS Enrollees: Table 3 provides the enrollee demographic information for the 14 HPPLs (representing four plans) that participated in testing. The 14 participating HPPLs were asked to provide data on all MLTSS enrollees age 18 and older in their plan. The 14 participating HPPLs collected data on 189,730 unique MLTSS enrollees who were 18 or older. Of these entries, 8 were missing information on sex, ethnicity, and/or race and were thus excluded from our data analyses. Table 3 summarizes the enrollees' characteristics for the remaining 189,722 enrollees in the sample. Of the total enrollees, 53.8 percent were female and 34.4 percent were age 65 or older. Nearly half (45.0 percent) of the enrollees had a reported race of "Other" in the HPPL data (however this is likely reflective of HPPL missing data on race), 28.0 percent were white, and 20.2 percent were black or African American. Other races (American Indian or Alaska Native, Asian, Native Hawaiian or

[9] Some state MLTSS programs include children with special needs. This population is not included in the measure or in testing.

Other Pacific Islander, and Unknown) each accounted for less than three percent of the sample. A little over a fifth of the enrollees (22.5 percent) were Hispanic or Latino.

Table 3. Analytic Sample Demographic Information

Characteristics	Total
Number of Enrollees (n)	189,722
Sex	
Female	53.8
Male	46.2
Age	
18-39	20.9
40-64	44.7
65–74	17.6
75-84	10.8
85 and older	6.1
Race	
American Indian or Alaska Native	0.5
Asian	2.0
Black or African American	20.2
Native Hawaiian or Other Pacific Islander	2.8
Other Race	45.0
Unknown	1.4
White	28.0
Ethnicity	
Hispanic or Latino	22.5
Non-Hispanic or Latino	74.5
Unknown	3.0

Source: Mathematica analysis of MLTSS enrollees from four health plans and fourteen health plan product lines

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

No difference in the sample size used for testing.

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 did not analyze social risk factors due to three factors: (1) this measure focuses exclusively on a population with social risk (i.e., Medicaid beneficiaries eligible for LTSS due to aging, disability or chronic illness), (2) patient-reported data and patient community characteristics were not available in the testing data source of administrative claims and (3) findings from a recent two-year National Quality Forum (NQF) effort indicated that the inclusion of area-level socio-economic status (SES) indicators did not improve the predictive capacity of risk-adjustment algorithms of hospital-based care measures developed for Medicare beneficiaries. These Medicare hospital measures were endorsed without SES indicators, although NQF directed the measure

developers to evaluate whether SES indicators should be included in the future as part of the annual update process.[10] .

Although some states collect supplemental data to MLTSS plan-level data that could be used to identify social risk factors, this data is not uniform across states and is not uniformly available to MLTSS plans. As the quality and integration of data between MLTSS plans and states improves it may be possible to examine the impact of social risk factors on this measure in the future.

_____

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*)

This measure is a risk-adjusted outcome measure assessing how well MLTSS plans are able to minimize length of stay in the institutional setting to less than 100 days for all new admissions and ensure the discharge to the community is successful for at least 60 days (i.e., no readmission to the institution, hospitalization or death).

To assess reliability, we used a signal-to-noise ratio (SNR) analysis of the performance measure score, which quantifies the degree to which variation is the result of differences in performance versus random measurement error. This type of assessment addresses whether differences in measure results between reporting entities are attributable to either differences in their underlying performance or chance or other sources of variation. The signal variance characterizes the magnitude of differences in underlying performance between reporting entities, or the between-entity variance.

The SNR statistic, R (ranging from 0 to 1), summarizes the proportion of the variation between entity scores that is due to real differences in underlying entity characteristics (such as differences in population demographics or medical care) as opposed to background-level or random variation (for example, due to measurement or sampling error). If R=0, there is no variation on the measure across entities, and all observed variation is due to sampling variation. In this case, the measure is not useful to distinguish between entities with respect to healthcare quality. Conversely, if R=1, all entity scores are free of sampling error, and all variation represents real differences between entities in the measure result.

We estimated the noise variance of each HPPL's risk-adjusted rate using a logistic regression model, and then used the Morris method to iteratively estimate the between-plan variation (i.e., the signal) using a maximum likelihood approach. (Morris, 1983). We then computed SNR as the ratio of signal variance to the sum of the signal and noise variances (total variance in the measure):

[10] National Quality Forum. 2017. All-Cause Admissions and Readmissions 2015–2017. Technical Report. Available at http://www.qualityforum.org/Publications/2017/04/All-Cause_Admissions_and_Readmissions_2015-2017_Technical_Report.aspx

$$R = \frac{\sigma^2_{between}}{\sigma^2_{between} + \sigma^2_{within}}$$

In general, a high SNR implies that differences in plans' measure results are meaningful to distinguish their performance.

References

1. Morris, C. (1983). Parametric Empirical Bayes Inference: Theory and Applications. *Journal of the American Statistical Association, 78*(381), 47-55. doi:10.2307/2287098

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*)

This measure had a mean reliability score of 0.8575 for all 14 HPPLs. However, one plan had 10 MLTSS enrollees in the denominator and thus was not eligible for reporting based on CMS standards. [11] We performed additional analyses in which we excluded this plan from the SNR calculations. Excluding this plan led to the increase in the mean reliability from 0.86 to 0.92 (Table 4). For twelve of the thirteen remaining HPPLs the SNR exceeded 0.7.

Table 4. Signal-To-Noise Reliability [a]

HPPL	Mean SNR
HPPL-01	0.9511
HPPL-02	0.8454
HPPL-03	0.6326
HPPL-04	0.9907
HPPL-05	0.9799
HPPL-06	0.9237
HPPL-07	0.9861
HPPL-08	0.8592
HPPL-09	0.9309
HPPL-11	0.9809
HPPL-12	N/A
HPPL-13	0.9862
HPPL-14	0.8736
HPPL-17	0.9870
Average	**0.9175**

Source: Mathematica analysis of data from four MLTSS plans and fourteen health plan product lines.

[a] SNR values excluding plans with fewer than 11 enrollees in the numerator.

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, we found the reliability of the *Minimizing institutional length of stay measure* to be high, based on the reliability of the 13 HPPLs meeting minimum reporting standards. In general, a high SNR implies that differences in plans' measure results are meaningful to distinguish their performance. The ideal standard for SNR is 0.7 or greater based on the work of Adams et al (2010). The mean reliability score for this measure was

[11] This policy stipulates that no cell (e.g. admittances, discharges, patients) less than 11 may be displayed. Also, no use of percentages or other mathematical formulas may be used if they result in the display of a cell less than 11 (https://www.cms.gov/Medicare/CMS-Forms/CMS-Forms/downloads/cms-r-0235l.pdf).

0.92 when excluding HPPLs with a low sample size (<11 enrollees), as shown in Table 4. Twelve of 13 HPPLs exceeded a threshold of 0.7 for reliability; the remaining HPPL had a reliability of 0.6326. These results suggest the *Minimizing institutional length of stay* measure was highly reliable in distinguishing performance between HPPLs, according to the guidance published by Adams for the application of provider profiling (Adams, 2010). Of note, these measures are expected to be used for external or internal quality improvement purposes.

References

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

2b1. VALIDITY TESTING

2b1.1. What level of validity testing was conducted? (*may be one or both levels*)
☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☒ **Empirical validity testing**

☐ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*) **NOTE**: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

2b1.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

This measure is a risk-adjusted outcome measure assessing how well MLTSS plans are able to minimize length of stay in the institutional setting to less than 100 days for all new admissions and ensure the discharge to the community is successful for at least 60 days (i.e., no readmission to an institution, hospitalization, or death).

Validity was tested empirically using convergent validity (described below). We also evaluated face validity throughout the measure development process through the input of four technical expert panels and a public comment period. We did not conduct a formal assessment of face validity, however the process of ensuring stakeholders agreed with the measure specification choices and supported the overall measure concept is described below for additional context.

Empiric Validity Testing Method: Measure validity indicates that the measure accurately represents the concept being measured and achieves the purpose for which it is intended. Convergent measure validity can be demonstrated by examining the relationship between measure results and other measures of quality we expect to correlate with performance.

In the absence of any existing, publicly reported plan-level measures of LTSS quality, we examined the convergent validity with two related measures undergoing endorsement consideration, by examining their relationship to one another among the 13 HPPLs meeting minimum reporting standards:

- Admission to an institution from the community: This measure assesses the rate of utilization of institutional care in three categories (Short stay - 1 to 20 days; Medium stay – 21-100 days; Long stay – 101 days or more) per 1,000 MLTSS enrollee months. This measure is risk-stratified by age. We hypothesize that MLTSS plans that are successful at minimizing length of stay (i.e. higher rates on this measure) will have lower rates of utilization of long-stay institutional care (i.e. lower rates on the long-stay rate on the comparison measure) resulting in a negative correlation. The quality improvement actions to reduce the number of admissions that become long-stay would improve performance on both measures (e.g., provision of adequate home and community-based services in the community). We did not hypothesize a clear relationship with the short and medium-stay rates on this measure (i.e., a plan that is good at minimizing length of stay might have higher short term stays in relation to

long term stays, but the direct relationship of absolute number of short term stays to this measure is less clear).

- Successful transition after long-term institutional stay: This measure assesses the proportion of long-stay institutional residents (101 days or more) who successfully transition to the community (no readmission to the institution, hospitalization or death in 60 days post-discharge). This measure is risk adjusted for age, gender and comorbid conditions. We hypothesize that health plans that are successful at coordinating care to minimize length of stay in an institutional setting and ensure successful discharge to the community, will also be successful at ensuring MLTSS enrollees who are long-stay residents can successfully transition to the community. The quality improvement actions to improve performance across these measures are similar (i.e., careful monitoring of MLTSS enrollee needs in the institutional setting and provision of adequate home- and community-based services in the community).

We assessed convergent validity based on the Spearman correlation as poor (less than 0.4), moderate (0.4 – 0.69), good (0.7 – 0.89), and high (greater than 0.9).

Process of Expert Input and Public Comment: In addition to empiric testing of validity, we solicited feedback on the measures' importance and usability, clarity of the specifications, and measure specification issues from all interested parties, through an open public comment process held from September 15 through October 6, 2016. We also obtained input throughout the measure development process from multiple Technical Expert Panels (TEPs) (see Ad.1 of measure submission form for TEP rosters) through multiple meetings from 2013-2018. The TEPs specifically advised on the importance of the measure concept to MLTSS plans, states and MLTSS-enrollees, the choices made in measure development, the measure testing results and the risk-adjustment approach.

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

Table 5. Results of Correlation Analyses

Measure	Strata	Correlation Coefficient	p-value
Admission to an institution from the community: Long Stay (101+ days)**	Age 18-64	-0.12321	0.6747
	Age 65-74	-0.54125*	0.0456
	Age 75-84	-0.47965	0.0826
	Age 85+	-0.56292*	0.0361
Successful transition after long-term (101+ days) institutional stay	Risk-adjusted for age, gender and co-morbid conditions	0.89091**	0.0005

Source: Mathematica analysis of data from four MLTSS plans and fourteen health plan product lines.

*Correlation was significant at p<.05

**Correlations between the proposed measure and Long-stay admissions was hypothesized to be negative because the measure of long-stay admissions is a "lower is better" measure.

As hypothesized, we saw a moderate, negative correlation between a measure of utilization of long-stay institutional care and performance on this measure of minimizing institutional length of stay. This correlation was consistently negative across age groups and was significant (p<0.05) for the 65-74 age group (correlation = -0.54) and 85+ age group (-0.56). These results support the hypothesis that MLTSS HPPLs that provide high quality high quality home- and community-based services will show better performance across these measures (increased in performance on *Minimizing length of stay* measure and decreased performance on *Admission to an institution from the community* measure long stay rates).

We saw an even stronger positive correlation between a measure of successful transition to the community after long-stay institutional stay and this measure of minimizing institutional length of stay (correlation=0.89, p = 0.0005). This result is also consistent with our hypothesis that health plans which are successful at coordinating care to minimize length of stay in an institutional setting and ensure successful discharge to the community, will also be successful at ensuring MLTSS enrollees who are long-stay residents can successfully transition to the community. The quality improvement actions to improve performance across these measures are similar (i.e., careful monitoring of MLTSS enrollee needs in the institutional setting and provision of adequate home and community based services in the community).

Expert Input and Public Comment

We also received feedback from a three-week public comment period hosted on CMS's online public comment system during September – October 2016. The public comment period was open and broadcast to all interested parties. Overall, commenters supported the measure and the efforts to measure transitions between the community and institutional settings for MLTSS beneficiaries. Commenters noted the gap in measures for MLTSS beneficiaries and applauded this measure as a first step that could provide valuable information about the ability of beneficiaries to live safely in the community. We received specific comments questioning the overlap with measures of post-acute care facilities developed under the IMPACT act. It is important to note the similar IMPACT measures only include Medicare FFS beneficiaries and therefore exclude all Medicaid-only beneficiaries and Medicare managed care enrollees. None the less, the measurement team reviewed these measures and aligned the approach where appropriate for the level of accountability.

Over the course of multiple years of development and testing, multiple technical expert panels provided input on the development of these measures. They provided input on specific questions on importance of the measure concept to MLTSS plans, states and MLTSS-enrollees, the choices made in measure development, the measure testing results, and the risk-adjustment approach. With regard to the measure specification, the TEPs provided valuable confirmation of our choices for the definition of the denominator, the numerator, and the risk-adjustment approach. While not formally assessed, the TEP confirmed these measures are important and demonstrate a significant gap in performance across MLTSS plans, confirming that the performance score from this measure can be used to compare quality across MLTSS plans.

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?*)

Empiric validity of performance measure score

We assessed convergent validity based on Spearman correlations with other measures of similar quality constructs. We interpreted correlation coefficients as poor (less than 0.4), moderate (0.4 – 0.69), good (0.7 – 0.89), and high (greater than 0.9). Overall the results showed moderate to good associations for hypothesized relationships suggesting that the *Minimizing institutional length of stay* measure meets the test of validity, based on the direction and strength of observed associations.

2b2. EXCLUSIONS ANALYSIS

NA ☒ no exclusions — *skip to section* 2b3

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*)

Not applicable.

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*)

Not applicable.

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

_____

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 2b4.

2b3.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with 18 risk factors**

☐ **Stratification by risk categories**

☐ **Other,**

2b3.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

Risk Model: A logistic regression was used to model the log-odds of an admission being successfully discharged within 100 days of admission.

Risk Factors: Age, gender, dual-eligibility for Medicaid and Medicare, hospital utilization in the period prior to admission to the institutional facility, enrollment in the MLTSS plan for 6 months, and comorbid conditions (Alzheimer's disease and related disorders, asthma, intellectual disabilities, mental health conditions, stroke).

Comorbid conditions for chronic medical conditions, physical, intellectual or developmental disabilities, mental health conditions, or substance use disorders were assigned based on the 62 condition category Chronic Conditions Warehouse (CCW) algorithms. These algorithms are maintained by CMS and publicly available at https://www.ccwdata.org/web/guest/. The diagnosis codes used to calculate the conditions included in the final model are provided in the Minimizing Length of Stay Value Set and Risk Adjustment Tables Excel file.

Coefficients (including codes and definitions): See Table 6

Table 6. Logit Model Specification: Risk Factor Weights

Risk factor	Beta	Risk factor	Beta
Dual	0.1157	**Number of hospital stays in the pre-period**	
Sex/Age Category		1	0.0935
Female, age 18-44	0.8471	2+	-0.4930
Female, age 45-64	0.3157	**Chronic or potentially disabling condition**	
Female, age 65-74	0.5872	Alzheimer's disease and related disorders	-0.5862
Female, age 75-84	0.6513	Asthma	-0.7522
Female, age 85+	0.5065	Intellectual disabilities[a]	-0.4861
Male, age 45-64	0.4356	Mental health conditions[b]	-0.4468
Male, age 65-74	0.4109	Stroke	-0.5140
Male, age 75-84	0.6856	**Had 6 months of enrollment**	0.4212
Male, age 85+	0.4395	Intercept	-0.9966

Source: Mathematica analysis of MLTSS enrollees from four health plans and fourteen health plan product lines.

a Includes intellectual disability, learning disability, other developmental delays, traumatic brain injury, and autism spectrum disorders.

b Includes conduct disorders, ADHD, hyperkinetic syndrome, bipolar disorder, depressive disorders (including depression), personality disorders, post-traumatic stress disorder, schizophrenia, and other psychotic disorders.

Equations: Calculate the expected performance rate for each plan based on the case mix of the denominator admissions. To calculate the expected rate for each plan k, let $E_k = \sum_{i \in k} p_i$, where n_i is the number of admissions for patient i in plan k and p_i is indicator of successful discharge to the community within 100 days of admission for patient i. p_i is calculated by using the following formula:

$$p_i = \frac{exp\left(\sum_j c_j \cdot x_j\right)}{\left(1 + exp\left(\sum_j c_j \cdot x_j\right)\right)}$$

where c_j is the coefficient estimate for risk factor j and x_j is patient *i*'s value for risk factor.

Calculate the multi-plan population rate Y by taking the sum of all observed numerator events and dividing by the sum of all observed denominator events.

The risk-adjusted rate (r_k) for each plan *k* is equal to:

$$r_k = \frac{Observed\ Rate_k}{Expected\ Rate\ _k} \times\ Y.$$

2b3.2. If an outcome or resource use component measure is <u>not risk adjusted or stratified</u>, provide <u>rationale and analyses</u> 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 <u>and</u> 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?

<u>Conceptual Method for Selecting Patient Factors Used in the Model</u>

The choice of risk factors was guided by Andersen's Behavioral Model of Health Services Use (Andersen, 1995), which frames the determinants of health care utilization into the following: (1) factors including demographic characteristics such as age and sex that predispose individuals to use care; (2) factors such as income and distance to a clinic that enable individuals to seek care (e.g., income); and (3) factors such as the presence of chronic conditions or functional limitations that drive individuals to need care. Andersen's model also treats health care utilization as a measure of "realized access" to care, signifying that individuals were able to overcome any perceived barriers to care receipt.

We used the following three criteria to assess the appropriateness for potential inclusion as predictors in the risk-adjustment model:

1. Likely predictive importance - Assessed by reviewing the related health services literature.
2. Feasibility for testing - Our assessment was limited to measures that can be constructed by using data available in our testing sample (administrative claims data on hospital and institutional facility utilization and enrollment data). The project team assessed whether constructing the predictor was possible with existing data and feasible, given time and resource constraints.
3. Appropriate accountability incentives – Assessment of whether the inclusion of a measure was consistent with the aim of rewarding accountable entities for good performance on the measure in a dynamic context.

Table 7 shows the candidate predictors for inclusion rated by the three assessment criteria. The subsequent model development and testing was limited to the following variables that the project team rated "high" across all three categories (highlighted in bold in Table 6 below): sex, age, eligibility category, and the presence of chronic conditions. We did not include race in the model in keeping with guidance from "A Blueprint for the CMS Measures Management System," which outlines that the inclusion of race can potentially mask important disparities across racial or ethnic groups (Centers for Medicare & Medicaid Services 2016).

Table 7. Candidate risk factors by assessment criteria

	Evidence of association in literature	**Feasibility of MLTSS plans to report**	**Appropriate accountability incentives**
Predisposing variables			
Sex	**No**	**Yes**	**Yes**
Age	**Yes**	**Yes**	**Yes**
Marital status	Yes	No	Yes
Dual status	**Yes**	**Yes**	**Yes**
Enabling variables			
Length of MLTSS plan enrollment	No	Yes	Yes
Housing availability	No	No	Yes
Institutional characteristics (e.g., occupancy levels)	Yes	No	No
Availability of community-based LTSS providers	Yes	No	No
Socioeconomic status	No	No	No
Need variables			
Diagnoses for chronic medical conditions	**Yes**	**Yes**	**Yes**
Diagnoses for physical disabilities	**Yes**	**Yes**	**Yes**
Diagnoses for intellectual/ developmental disabilities	**Yes**	**Yes**	**Yes**
Diagnoses for mental health conditions	**Yes**	**Yes**	**Yes**
Assessment data: ADLs/IADLs	Yes	No	Yes
Assessment data: Social supports	Yes	No	Yes
Realized access			
Prior institutionalizations	No	Yes	No
Prior hospitalizations	**No**	**Yes**	**Yes**

Source: Based on review of Arling et al, 2011; Gaugler et al. 2007; Irvin et al, 2016; Dominiak and Bohl, 2016.

Note: Bold font indicates variables that were included in the final model specification

To group chronic conditions, we used the Chronic Condition Warehouse (CCW) grouper to align with other Medicaid measures in development[12]. All CCW were included in the modeling process. We also explored the effect of modeling combined groupings of conditions that were clinically related based on feedback from staff clinicians to determine if predictive power was improved. Both the individual CCWs and groups were considered as potential predictors in the model development stage.

We were unable to include marital status or functional status because these data were not maintained by plans in standardized, easily extractable formats. Although we did have information on the number of institutional admissions during the six months prior to the measurement period, including this as a candidate risk factor would have created a disincentive for improvement by adjusting for prior poor performance (higher rates of institutional admissions reflect poor quality of MLTSS coordination and or services, which has a conceptual relationship to facilitating a successful discharge to the community after a short-term stay).

Statistical Method for Selecting Patient Factors Used in the Model

We used a modified version of the forward stepwise regression modeling to determine which risk factors should be included in the model for each measure. The procedure first estimates parameters for risk factors forced into the model. These effects are the intercept and the 13 explanatory variables (including nine age by sex interaction dummy variables, two hospital stay dummy variables, dual status, and the six-month enrollment status). Chronic condition risk factors are then entered into and removed from the model in such a way that each forward selection step can be followed by one or more backward elimination steps. Specifically, the procedure computes a chi-square statistic for each risk factor not in the model and examines the largest of these statistics. If it is significant at the specified level for entry,[13] the corresponding risk factor is added to the model. Risk factors already in the model do not necessarily stay; as new factors enter into the model, the least significant of these risk factors is removed from the model and the algorithm proceeds to the next step. The stepwise selection process terminates once none of the risk factors outside the model have a statistically significant level for entry and every effect in the model is significant at the level for stay.

Results of Risk Factor Selection:

After further imposing the criterion of only retaining risk factors with a p-value lower than 0.10, the final model included 18 risk factors and an intercept term. The prevalence, coefficient and p-value for each risk factor is described in Table 8 below.

Overall, there are similarities in the magnitude of odds ratios for the risk factors between the development and validation models, which indicates models are well-calibrated (see Table 8). The relationship between age and successful discharge within 100 days of admission was positive across nearly all age and sex categories, suggesting relative to males age 18-44, being older was associated with higher odds of transition. We attribute this finding to the heterogeneous nature of the MLTSS population, with varying needs and functional limitations.

The relationship between having a single hospital stay occurring prior to the measurement period was not a significant predictor but having 2 or more hospital stays was associated with a significant reduction in the odds of successful discharge within 100 days of admission. Conversely, having been enrolled in MLTSS for 6 months or more prior to the measurement period was positively associated with a successful outcome. In models predicting successful transition after a short stay, all three chronic or potentially disabling conditions (Alzheimer's disease and related conditions, asthma, and stroke), and one group chronic health risk factor

[12] We evaluated alternative chronic condition groupers, including Hierarchical Condition Categories (HCCs) and the AHRQ Clinical Classifications Software (CCS) but determine the CCW was the best approach.

[13] The significance levels we set up for a risk factor to enter or to stay in a model is 0.30 (*Minimizing Institutional Length of Stay*) and 0.35 (*Successful Transition after Long-Term Stay*). Note that to prevent cycling, we deliberately set the level-to-enter slightly smaller than the level-to-stay.

(mental health conditions) were statistically significant and negatively associated with discharge. Dual status was not an indicator of a successful outcome.

Table 8. Final model specification: Risk factor prevalence and odds ratios

	Development Sample			Validation Sample			Full Sample		
Risk factor	**Percentage of enrollees with this risk factor**	**Odds ratio**	**(95% confidence interval)**	**Percentage of enrollees with this risk factor**	**Odds ratio**	**(95% confidence interval)**	**Percentage of enrollees with this risk factor**	**Odds ratio**	**(95% confidence interval)**
Dual	68.03	1.12	(0.9215, 1.3676)	66.43	1.16	(0.9499, 1.4236)	67.23	1.14	(0.9901, 1.3126)
Sex/Age Category									
Female, age 18-44	1.80	2.33	(1.1671, 4.6633)	1.98	1.15	(0.5774, 2.2947)	1.89	1.63	(1.0036, 2.6560)
Female, age 45-64	17.94	1.37	(0.8442, 2.2271)	17.41	1.31	(0.8164, 2.1133)	17.67	1.35	(0.9616, 1.8950)
Female, age 65-74	9.69	1.80	(1.0719, 3.0191)	10.52	1.87	(1.1296, 3.0809)	10.11	1.85	(1.2897, 2.6489)
Female, age 75-84	13.59	1.92	(1.1545, 3.1867)	13.36	1.33	(0.8081, 2.1959)	13.48	1.61	(1.1265, 2.2947)
Female, age 85+	14.57	1.66	(0.9963, 2.7641)	13.24	1.55	(0.9346, 2.5613)	13.90	1.61	(1.1229, 2.2985)
Male, age 45-64	21.37	1.55	(0.9580, 2.4947)	21.75	1.28	(0.8025, 2.0479)	21.56	1.42	(1.0128, 1.9770)
Male, age 65-74	8.36	1.51	(0.8887, 2.5592)	7.98	1.54	(0.9205, 2.5908)	8.17	1.53	(1.0604, 2.2206)
Male, age 75-84	6.53	1.99	(1.1517, 3.4211)	6.71	1.23	(0.7181, 2.1194)	6.62	1.58	(1.0757, 2.3142)
Male, age 85+	3.07	1.55	(0.8268, 2.9131)	3.99	1.43	(0.7927, 2.5731)	3.53	1.50	(0.9790, 2.3082)
Number of hospital stays in the pre-period									
1	5.47	1.10	(0.7891, 1.5279)	5.73	1.12	(0.8110, 1.5505)	5.60	1.11	(0.8774, 1.3916)
2+	5.38	0.61	(0.4261, 0.8754)	5.70	0.43	(0.2977, 0.6116)	5.54	0.51	(0.3941, 0.6547)
Chronic or potentially disabling condition									
Alzheimer's disease and related disorders	23.82	0.56	(0.4583, 0.6756)	22.34	0.53	(0.4325, 0.6493)	23.08	0.54	(0.4735 , 0.6266)
Asthma	6.06	0.47	(0.3259, 0.6817)	5.44	0.50	(0.3360, 0.7346)	5.75	0.49	(0.3731 , 0.6376)
Intellectual disabilities[a]	3.37	0.62	(0.3832, 0.9870)	2.36	0.92	(0.5538, 1.5275)	2.87	0.73	(0.5206 , 1.0357)
Mental health conditions[b]	27.51	0.64	(0.5332, 0.7673)	27.51	0.66	(0.5473, 0.7884)	27.51	0.65	(0.5671, 0.7336)
Stroke	6.21	0.60	(0.4263, 0.8393)	5.76	0.68	(0.4812, 0.9517)	5.98	0.64	(0.5018, 0.8105)

	Development Sample			Validation Sample			Full Sample		
Risk factor	Percentage of enrollees with this risk factor	Odds ratio	(95% confidence interval)	Percentage of enrollees with this risk factor	Odds ratio	(95% confidence interval)	Percentage of enrollees with this risk factor	Odds ratio	(95% confidence interval)
Had 6 months of enrollment data	21.96	1.52	(1.2386, 1.8748)	22.84	1.76	(1.4342, 2.1616)	22.40	1.64	(1.4187, 1.8975)

Source: Mathematica analysis of MLTSS enrollees from four health plans and fourteen health plan product lines.

[a] Includes intellectual disability, learning disability, other developmental delays, traumatic brain injury, autism spectrum disorders.

[b] Includes conduct disorders, ADHD, hyperkinetic syndrome, bipolar disorder, depressive disorders (including depression), personality disorders, post-traumatic stress disorder, schizophrenia, and other psychotic disorders.

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?

No social risk factors were evaluated in the risk-adjustment analysis for two reasons:

1. The focus of this measure is a population with increased social risk (e.g., Medicaid MLTSS enrollees) who are primarily low income. By defining a measure and risk-adjustment approach specific to this population, the measure is acknowledging and accounting for the unique risk that this population faces. Because the measure is not intended for use on an overall population with more diverse social risk (e.g., a Medicare population that may include dual-eligible and non-dual eligible beneficiaries), the inclusion of risk factors that account for the existing medical and disability conditions within the population of Medicaid beneficiaries utilizing LTSS is more appropriate.
2. Our assessment was also limited to risk factors that can be constructed by using claims and encounter files available to MLTSS plans. Therefore, we were limited in the social risk factors that could be calculated from the existing data. Findings from a recent two-year National Quality Forum (NQF) effort indicated that the inclusion of area-level SES indicators does not improve the predictive capacity of risk-adjustment algorithms of hospital-based care measures developed for Medicare beneficiaries (NQF, 2017).

References

1. National Quality Forum. (2017). "All-Cause Admissions and Readmissions 2015–2017." Technical Report. April 2017. Available at http://www.qualityforum.org/Publications/2017/04/All-Cause_Admissions_and_Readmissions_2015-2017_Technical_Report.aspx.

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.**

Not applicable. We did not have access to 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*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

If stratified, skip to 2b3.9

We used stratified random sampling to assign enrollees to the development and validation samples. The development sample supported our model-building and exploration work; the other served as the validation sample against which we assessed the final model's performance.

We performed a number of tests of model performance.

First, we tested model discrimination using receiver operating characteristic (ROC) curves and area under the curve (AUC) statistics. The ROC curve assesses how well a model discriminates, or separates, individuals into two classes (in our case, the ROC curve assesses whether there is a successful discharge/transition to the community after a short/long institutional stay). Visually, the further up and away from the 45-degree line the curve is, the better the model fit (Figures 1-2 below).

Second, we conducted a Hosmer-Lemeshow test of goodness of fit. This test examines whether there is significant evidence of a poor fit; in that respect p-values exceeding 0.05 reflect better fitting models (Hosmer and Lemeshow, 2000).

Third, we evaluated model calibration, measured by the decile plots of observed risk versus predicted risk. In decile assessment, we should see similar numbers of observations in each decile group and increasing observed rates when we move from low to high deciles (see Table 9 below).

References

1. Hosmer, DW and Lemeshow, S. Applied Logistic Regression. New York: Wiley; 2000.

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

AUC estimates indicate relatively moderate discrimination[14] (AUC=0.6346 on the development sample). However, we note that since the goal is to use our predictive risk-adjustment models to stratify score data into risk groups, in which case proper model calibration is more important than model discrimination.

Figure 1. ROC curve, developmental sample

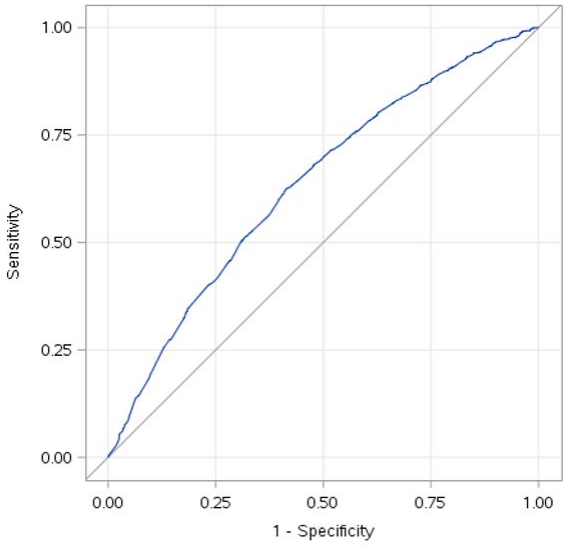
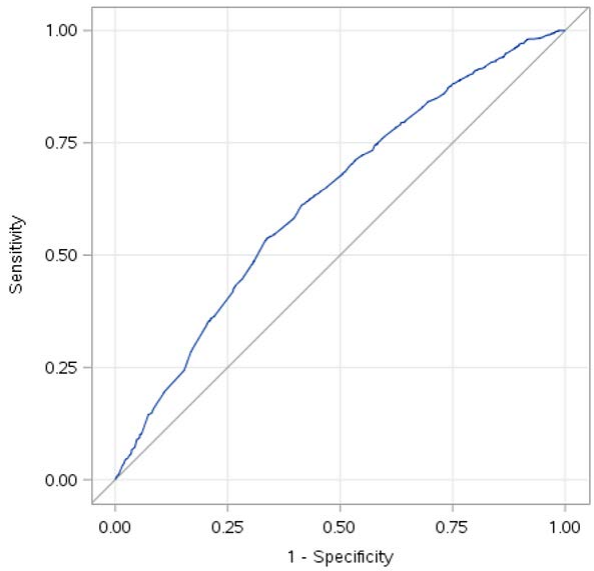


Figure 2. ROC curve, validation sample



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

Our Hosmer-Lemeshow goodness of fit results (p-values from the Chi-squared tests) was 0.69 on the development samples and 0.37 on the validation samples, reinforcing that models are well-calibrated.

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

Table 9. Model-type decile table (validation)

Decile	Number of stays	Observed mean count per 1,000	Expected mean count per 1,000
Short-stay model			
1 (lowest)	352	16.48	15.86
2	336	22.62	22.93
3	335	27.76	26.91
4	355	28.45	29.70
5	318	30.50	33.37
6	335	31.94	36.36
7	443	42.89	39.58
8	288	47.57	42.75
9	292	38.36	44.81
10 (highest)	330	50.00	50.43

Source: Mathematica analyses of enrollee data from 4 health plans representing 14 health plan product lines.

[14] AUC thresholds are domain specific, and we follow the classification system from Hosmer & Lemeshow, where AUC can be interpreted as follows: 0.90-1.00 = outstanding discrimination; 0.80- 0.89 = excellent discrimination; 0.70-0.79 = acceptable discrimination; 0.60–0.70 = able to discriminate; 0.50–0.60 = no discrimination.

Note: The sample is composed of enrollee stays in the validation sample (n =3,384).

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*)

Based on the testing results described above we determined the model adequately controlled for differences in the patient population and was the best possible model given the limited data available. We received input from an independent panel of risk-adjustment experts who suggested alternative to the model that the team investigated (see question 2b3.11). After testing alternatives, the risk-adjustment advisory workgroup (members listed below) agreed the proposed model was adequate.

Risk Adjustment Advisory Workgroup Members

- Marguerite Burns, Ph.D., assistant professor, University of Wisconsin School of Medicine and Public Health
- Ezra Golberstein, Ph.D., associate professor, University of Minnesota School of Public Health
- Lisa Iezzoni, M.D., M.Sc., professor, Harvard Medical School (Technical Expert Panel member)
- Joanna Jiang, Ph.D., senior social scientist, Agency for Healthcare Research and Quality
- Zhenqiu Lin, Ph.D., director of data management and analytics, Center for Outcomes Research and Evaluation, Yale University
- Patrick Romano, M.D., professor, University of California, Davis, School of Medicine
- Jonathan Shaw, M.D., M.S., clinical assistant professor, Stanford University School of Medicine

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*)

The risk adjustment advisory workgroup provided the following suggestions relating to the risk-adjustment model for this measure, and the results of our subsequent analyses:

- *Analyze the distribution of enrollees having at least 6 months of pre-period enrollment by HPPL. T*he group felt that if there was marked variation across HPPLs in the number of enrollees with at least six months of enrollment, this risk factor might show a spurious statistical association with the outcomes (successful discharge within 100 days of admission). We explored this issue, and found that this did vary markedly by plan: seven of 14 plans had between 45 and 72 percent of enrollees with at least six months of enrollment, while the remaining seven had 15 percent or less of enrollees meeting this criteria. However, we feel the inclusion of this risk factor in our models, and the observed association between at least six months of enrollment and greater likelihood of successful discharge or transition, is more likely a reflection of HPPLs' greater ability to impact and manage outcomes among enrollees for whom they have a longer relationship. In addition, more than 40% of enrollees were part of a Medicare-Medicaid Plan (MMP), which were introduced in 2014, and thus 0% of data from MMPs have 6 months or more of enrollment data. Consequently, it is an important variable to include in the risk adjustment model.
- *Consider including a variable for total days in the hospital during the pre-period rather than the count of hospital stays.* The workgroup thought the length of a hospital stay, rather than a categorical indicator of number of hospital stays, might be a better predictor of discharge or transition. Based on this suggestion, we created a variable capturing total days in the hospital. We then included both this variable and the original categorical variable capturing the count of hospital stays, in the pool of candidate risk factors considered in the stepwise regression models. We found that the stepwise regression, which prioritizes the most statistically significant associations, selected the original categorical variable for the count of

hospital stays, rather than the variable with the count of hospital days. We also found that HPPLs with the highest average number of hospital days also had the highest average number of stays, indicating the number of hospital days variable was not adding explanatory power, so we do not recommend that it be included in the model.

- *Analyze the differences in number of hospitalizations across plans.* The concern expressed by the work group is that variation in hospitalization rates might signify variation in quality. We analyzed the average number of hospitalizations by HPPL and observed differences across plans but not as many differences within plans. In other words, the HPPLs with the highest average number of hospitalizations were part of the same health plan, suggesting the differences are driven by geographic differences in inpatient rates rather than differences in quality.
- *Identify the diagnoses/conditions statistically associated with successful discharge or transition, and then evaluate whether including it may create an inappropriate incentive (i.e., allows plans to send enrollees to poorer quality institutions).* We carefully considered each condition category included in the model to determine whether it had face validity and created inappropriate incentives for MLTSS plans to deliver poor quality care to enrollees in institutional settings. Although it is possible – and perhaps even likely – that some of the conditions included in our final recommended models developed during the institutional stay, such as stroke and pressure ulcers, ultimately we felt it was more important to fully reflect the complexity of these enrollees prior to discharge or transition. By including these conditions in our model, we are able to give credit to MLTSS plans that attempt to ensure these individuals are discharged or transitioned successfully back to the community.

The workgroup met again on May 18, 2018, to review risk-adjustment results. After the meeting, they suggested we examine the following:

- *Estimate models using age and sex variables separately (not age by sex interaction terms).* The workgroup expressed concern that the age by sex interaction terms might be too narrowly defined. We followed the workgroup's suggestion and estimated additional models that did not interact age and sex (i.e., instead of females aged 65-74 as one term, we had a term for females and another term for age 65-74). Not only did we find the same set of condition variables enter into our models, but model performance using these alternative risk factor specifications was worse than our original models.

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)*

To evaluate whether the measures demonstrated statistically significant variation in performance across HPPL and/or showed sub-optimal performance (suggesting room for improvement), we calculated the overall sample rate and compared to the 95% confidence interval around each HPPL's risk-adjusted rate.

HPPL-level adjusted performance rates were calculated as the ratio of the observed to the expected (O/E) rate multiplied by the total (multi-HPPL) sample's unadjusted rate. An HPPL-specific O/E ratio documents whether a given HPPL's actual performance is higher or lower than expected, given the underlying sociodemographic and health profiles of its population. An O/E greater than 1 indicates that an HPPL has higher than expected utilization, indicating poorer performance. Conversely, an O/E less than 1 indicates that an HPPL has lower than expected utilization, indicating stronger performance. Multiplying each HPPL's O/E ratio by the multi-HPPL benchmark generates performance scores with the same scale as the outcome measure, a proportion per 100 new admissions, which facilitates cross-HPPL comparisons along the measure's original scale.

To allow for plausible assessment of each HPPL's relative performance, we conducted significance testing of the risk-adjusted results, through comparing each HPPL-specific risk adjusted rate to the overall rate (which is

the average outcome in the entire sample and is considered as the estimate of the reference population rate). We used the Ratio Method[15] for estimating the Risk Adjusted Rate (RAR) and standard error[16] (SE) of indirect standardization for each HPPL. We then compared the lower and upper bounds of the 95 percent confidence interval (CI) around the RAR to the overall rate. If the upper bound of the CI was lower than the overall rate, the outcome of HPPL *i* (RARi) was considered as significantly lower than the overall performance. If the lower bound of the CI was higher than the overall rate, the performance of HPPL *i* was considered as significantly higher than the overall performance.

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

The risk-adjusted rates also showed significant variation between HPPLs, as shown in Table 10. Five HPPLs had significantly higher rates (indicating better performance) and four HPPLs had significantly lower rates (indicating worse performance) relative to the overall sample performance rate (33.57%).

Table 10. Unadjusted and risk-adjusted measure results by HPPL

HPPL	Observed-to-expected ratio (O/E)	Unadjusted rate	Adjusted rate	
HPPL-01	1.21	52.3%	40.6%	H
HPPL-02	1.40	55.0%	47.0%	*H
HPPL-03	1.96	52.3%	40.6%	H
HPPL-04	1.34	55.0%	47.0%	H
HPPL-05	1.36	75.0%	65.9%	H
HPPL-06	1.18	43.9%	44.8%	
HPPL-07	0.93	48.9%	45.5%	
HPPL-08	0.25	34.8%	39.6%	L
HPPL-09	0.99	26.2%	31.2%	
HPPL-11	0.64	8.6%	8.3%	L
HPPL-12	0	39.2%	33.3%	
HPPL-13	0.70	23.7%	21.7%	L
HPPL-14	1.26	0.0%	0.0%	
HPPL-17	0.79	21.0%	23.6%	L
Total	**1.00**	43.2%	42.3%	

Source: Mathematica analysis of MLTSS enrollees from four health plans and fourteen health plan product lines

Sample: Full sample (N=6,768)

L = HPPL rate is significantly lower (worse) than the overall rate (95% CI)

[15] Risk-adjusted rate = (Observed rate / Expected Rate) * Reference Population Rate

[16] According to AHRQ quality indicator calculation tools, Var(RARi) = (α/Ei)2(1/ni)2Σ [Pij(1-Pij)] for HPPL i. where Pij is the predicted probability for jth individual within ith HPPL from our fitted logistic model. See *http://www.qualityindicators.ahrq.gov/Downloads/Resources/Publications/2011/Calculating_Confidence_Intervals_for_the_AHRQ_QI.pdf* for details.

H = HPPL rate is significantly higher (better) than the overall rate (95% CI)

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?*)

These results demonstrate statistically significant differences in performance across MLTSS HPPLs. We conclude that this measure can demonstrate meaningful differences in performance.

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*)
In our testing sample exactly 8 enrollees who were eligible for the measure (18 years and older) and enrolled in one of the 14 HPPLs were dropped due to invalid sex, race, or ethnicity data. This represents less than 0.01 percent of our total eligible sample of 189,730. The extent of missing data is too small to impact performance rates.

In implementation the intent of this measure is that MLTSS plans will calculate the measure based on their entire MLTSS population with no missing data. As states develop systems for implementing these measures we will encourage a process to audit the calculation of the measure to ensure plans are systematically removing certain observations.

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)

Not applicable; see 2b6.1.

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; see 2b6.1.

3. Feasibility

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

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. Data Elements Generated as Byproduct of Care Processes.

Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. To what extent are the specified data elements available electronically in defined fields (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*)
Update this field for **maintenance of endorsement**.

ALL data elements are in defined fields in electronic claims

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

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

Attachment:

3c. Data Collection Strategy

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

3c.1. Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

IF instrument-based, consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

Not applicable.

3c.2. Describe any fees, licensing, or other requirements to use any aspect of the measure as specified *(e.g., value/code set, risk model, programming code, algorithm)*.

The Institutional Facility Value Set containing UBTOB and UBREV codes is provided with the measure. NUBC Official UB-04 Specifications (UB-04 Data) is copyrighted by the American Hospital Association.

The American Hospital Association holds a copyright to the Uniform Bill Codes ("UB") contained in the measure specifications. The UB Codes in this specification are included with the permission of the AHA. The UB Codes contained in this specification may be used by health plans and other health care delivery organizations for the purpose of calculating and reporting the measure results or using the measure results for their internal quality improvement purposes. All other uses of the UB Codes require a license from the AHA. Anyone desiring to use the UB Codes in a commercial Product(s) to generate results, or for any other commercial use, must obtain a commercial use license directly from the AHA. To inquire about licensing, contact ub04@healthforum.com.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

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

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Specific Plan for Use	Current Use (for current use provide URL)
Quality Improvement (external benchmarking to organizations) Quality Improvement (Internal to the specific organization)	

4a1.1 For each CURRENT use, checked above (update for maintenance of endorsement), provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included
- Level of measurement and setting

Not applicable.

4a1.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

The measure is under initial endorsement review and is not currently used in an accountability program. A measure implementation plan will be developed by, or in conjunction with, CMS.

4a1.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3

years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

This measure is intended for use by states and health plans to monitor and improve the quality of care provided for the Medicaid MLTSS enrollee population. States and health plans may choose to begin implementing the measures based on their programmatic needs.

The intended use of the measure is to allow states to compare and evaluate the quality of LTSS care being provided in MLTSS plans with which they contract using a common, standardized and validated measure. The National Quality Forum confirmed the importance of developing measures of the degree to which people who need LTSS are served in home and community settings rather than in institutions, citing current state-specific measures that are similar to this measure, but vary by state and have not been rigorously tested (NQF 2016, p. 34-35).[17] This specific measure focuses on one critical outcome of high quality MLTSS care – minimizing institutional stays and ensuring successful return to the community.

In May 2017, the National MLTSS Health Plan Association recommended a set of model LTSS performance measures and network adequacy standards in an effort to assist states in complying with the 2016 final rule on managed care in Medicaid and Children's Health Insurance (CHIP). The LTSS minimizing institutional length of stay measure is included in the set of recommended measures that assesses person-centered planning and coordination.

4a2.1.1. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.

Not applicable. This measure has not been implemented yet. Unlike Medicare measures, there is no formal process by which draft results for Medicaid measures are shared with measured entities, such as a Dry Run used in the Hospital Inpatient Quality Reporting (IQR) and Outpatient Quality Reporting (OQR) programs. Feedback on the measure will be available after the measure has been implemented by states in their MLTSS programs.

4a2.1.2. Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

Not applicable.

4a2.2.1. Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.

Describe how feedback was obtained.

Not applicable.

4a2.2.2. Summarize the feedback obtained from those being measured.

Not applicable.

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

Not applicable.

[17] One of the priority gap areas, in system performance and accountability domain, is financing and service delivery structures that serve: "to increase the proportion of people served in home and community settings and to meet the needs of consumers." NQF provided an example of a measure in one state that is similar to the one proposed for NQF endorsement: "percent of new members meeting nursing facility level of care criteria who opt for HCBS over institutional placement."

4a2.3. Describe how the feedback described in 4a2.2.1 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

Not applicable.

Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)

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

This measure is being considered for initial endorsement. Adoption of this performance measure has the potential to improve the quality of care for Medicaid managed care enrollees who are receiving long-term services and supports.

4b2. Unintended Consequences

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

4b2.1. Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.

This measure has not been implemented yet. There were no unexpected findings identified during testing of this measure.

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

This measure has not been implemented yet. There were no unexpected findings identified during testing of this measure.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

No

5.1a. List of related or competing measures (selected from NQF-endorsed measures)

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

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

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

Are the measure specifications harmonized to the extent possible?

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

Not applicable.

5b. Competing Measures

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

OR

Multiple measures are justified.

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

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

Not applicable.

Appendix

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

No appendix **Attachment:**

Contact Information

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

Co.2 Point of Contact: Roxanne, Dupert-Frank, Roxanne.Dupert-Frank@cms.hhs.gov, 410-786-9667-

Co.3 Measure Developer if different from Measure Steward: Mathematica Policy Research

Co.4 Point of Contact: Jessica, Ross, jross@mathematica-mpr.com, 617-674-8384-

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

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

2015-2018 Technical Expert Panel

Carol Raphael, Manatt Health Solutions (Chair)

Ann Hwang, MD, Community Catalyst

Ari Houser, PhD, AARP Public Policy Institute

Dennis Heaphy, MPH, Disability Policy Consortium

Joe Caldwell, PhD, National Council on Aging

Lauren Murray, BA, National Partnership for Women and Families

Maggie Nygren, EdD, American Association on Intellectual and Developmental Disabilities

RoAnne Chaney, MPA, Michigan Disability Rights Coalition

Mary Lou Bourne, National Association of State Directors for Developmental Disabilities Services

Raina Josberger, MS, New York State Department of Health

Jason Rachel, PhD, Virginia Department of Medical Assistance Services

Balu Gadhe, MD, CareMore

Patricia Kirkpatrick, MJ, RN, CPHQ, Amerigroup Corporation

Cheryl Phillips, MD, SNP Alliance

Diane McComb, MSEd, QLarant

Steve Guenthner, BS, Almost Family, Inc.

Bonnie Marsh, RN, BSN, MA, Health Services Advisory Group

Brian Abery, PhD, University of Minnesota

Lisa Iezzoni, MD, Harvard Medical School

Pamela Parker, MPA, Independent Consultant-Integrated Care

Valerie Bradley, MA, Human Services Research Institute

2013 Technical Expert Panel

Anne Cohen, Health and Disability Policy Consultant, Disability Health Access, LLC

Patti Killingsworth, Assistant Commissioner and Chief of LTSS, Bureau of TennCare

Jennifer Lenz, Executive Director, State and Corporate Services, Health Services Advisory Group

Bonnie Marsh, Executive Director, State and Corporate Services, Health Services Advisory Group

Diane McComb, ANCOR Liaison with State Associations

Margaret A. Nygren, Executive Director and CEO, American Association on Intellectual and Developmental Disabilities

Joseph Ouslander, Professor of Clinical Biomedical Science, Florida Atlantic University

Pamela J. Parker, Manager, Special Needs Purchasing, State of Minnesota Department of Human Services

Cheryl Phillips, Senior VP Public Policy and Advocacy, Leading Age

D.E.B. Potter, Senior Survey Statistician, Agency for Healthcare Research and Quality

Juliana Preston, Utah Executive Director, HealthInsight

Genie Pritchett, Sr. Vice President Medical Services, Colorado Access

Alice Lind, Aging and Long Term Support Division, Washington State Department of Social and Health Services

The 2015-2018 TEP reviewed feedback obtained during public comment, as well as alpha and beta testing results, and advised on the refinements of the technical specifications. The 2013 TEP advised on the development of the initial measure concept and preliminary specifications.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released:

Ad.3 Month and Year of most recent revision:

Ad.4 What is your frequency for review/update of this measure?

Ad.5 When is the next scheduled review/update for this measure?

Ad.6 Copyright statement: Not applicable.

Ad.7 Disclaimers: Not applicable.

Ad.8 Additional Information/Comments: Please include Jessica Ross (jross@mathematica-mpr.com) and Henry Ireys (hireys@mathematica-mpr.com) on any communications about these measures, as well as Roxanne Dupert-Frank. Thank you.