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 #: 3319

Measure Title: Long Term Services and Supports (LTSS) Comprehensive Assessment and Update

Measure Steward: Centers for Medicare & Medicaid Services, Centers for Medicaid & CHIP Services

Brief Description of Measure: This measure assesses the percentage of Managed Long Term Services and Support (MLTSS) plan enrollees who have documentation of a comprehensive assessment in a specified timeframe that includes documentation of core and supplemental elements. This measure has two rates:

Rate 1: Percent of MLTSS plan enrollees with documentation of a comprehensive LTSS assessment including nine (9) core elements within 90 days of enrollment or at least annually.

Rate 2: Percent of MLTSS plan enrollees with documentation of a comprehensive LTSS assessment including nine (9) core elements AND at least twelve (12) supplemental elements within 90 days of enrollment or at least annually.

Developer Rationale: Comprehensive assessments serve as the foundation for developing comprehensive person-centered care plans, providing all needed services and supports of high quality, and performing well-coordinated care. However, the tools currently used by states and health plans to conduct assessments, and the performance measures used to evaluate the quality of assessments conducted vary widely. This measure would address the lack of standardization by assessing the percentage of Medicaid MLTSS enrollees who have a comprehensive assessment conducted that includes specific core and supplemental elements. A standardized measure of comprehensive assessment will allow for apples-to-apples comparisons of MLTSS plans across states.

To construct this measure, we followed a multistep process that included evidence review, empiric analysis and stakeholder input. To start, we conducted an environmental scan of assessment domains and elements required for the programs serving adults with LTSS needs (e.g., Special Needs Plans, Medicare-Medicaid Plans, 1915c waivers, Medicare Home Health, Nursing Facility Minimum Data Set, Program for all Inclusive Care of the Elderly). To augment our environmental scan, we conducted one-on-one interviews with key stakeholders and a Technical Expert Panel (TEP) convened in 2013 to solicit additional assessment domains. Through review and voting, the TEP identified 28 data elements as critical for comprehensive LTSS assessment. However, through our field test, we determined that requiring all 28 elements was too stringent and would not result in meaningful performance rates (i.e., four out of five plans had a 0% performance rate when all 28 elements were required).

To balance the need for a measure that is feasible in the current MLTSS environment that would produce non-zero rates and the desire of the TEP to push the field further to do a truly comprehensive assessment, we decided to develop two rates for the measure.

1. "Core" requirements that set a minimum baseline of performance, and

2. “Supplemental” requirements that demonstrate more thorough and comprehensive performance.

Over time, we anticipate that elements from the “supplemental” requirements will move to the “core” requirements as performance improves. In the meantime, the currently proposed “core” rates can fill a long-standing measurement gap while generating results that are both meaningful and usable to stakeholders. Details on how the core and supplemental elements were selected can be found in the Appendix: Additional Testing Results.

The two-rate approach aims to give program administrators and policymakers a much-needed tool to help move the managed long- term services and supports (MLTSS) field from extremely low performance (as demonstrated in our testing) to more comprehensive and standardized documentation of assessment, care planning and care coordination practices. The inclusion of both rates in a single measure signals the importance of this concept.

Numerator Statement: The measure has two rates. The numerators for the two rates are as follows:

Rate 1: MLTSS plan enrollees who had either of the following:

• A comprehensive LTSS assessment completed within 90 days of enrollment for new enrollees, with nine (9) core elements documented, or

• A comprehensive LTSS assessment completed at least once during the measurement year for all other enrollees (established enrollees), with nine (9) core elements documented.

Rate 2: MLTSS plan enrollees who had either of the following:

• A comprehensive LTSS assessment completed within 90 days of enrollment for new enrollees, with nine (9) core and at least twelve (12) supplemental elements documented, or

• A comprehensive LTSS assessment completed at least once during the measurement year for all other enrollees (established enrollees) with nine (9) core and at least twelve (12) supplemental elements documented.

Note: Initial assessment should be completed within 90 days of enrollment, and updated annually thereafter.

Denominator Statement: Medicaid MLTSS plan enrollees age 18 years and older.

Denominator Exclusions: Exclude enrollees in the denominator who were enrolled in the plan prior to September 1 of the year prior to the measurement year (i.e., established enrollees) who left the plan for more than 45 days between January 1 and December 31 of the measurement year.

Exclude enrollees who could not be reached for a comprehensive assessment or who refused a comprehensive assessment.

Measure Type: Process

Data Source: Management Data, Other, Paper Medical Records

Level of Analysis: Health Plan

IF Endorsement Maintenance – Original Endorsement Date: N/A **Most Recent Endorsement Date:** N/A

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

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

3324: LTSS Comprehensive Care Plan and Update

3325: LTSS Shared Care Plan with Primary Care Practitioner

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? This measure is grouped with two other measures that assess the continuum of assessment, care planning and care coordination. This continuum of care is described in greater detail in the accompanying Evidence Attachment form.

- LTSS Comprehensive Assessment and Update

- LTSS Shared Care Plan with Primary Care Practitioner

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *structure, process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured. For measures derived from patient report, evidence also should demonstrate that the target population values the measured process or structure and finds it meaningful.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☐ **Yes** ☒ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☐ **Yes** ☒ **No**
- **Evidence graded?** ☐ **Yes** ☒ **No**

Evidence Summary

- The developer provides a logic model describing the steps between the process of completing a comprehensive assessment and the outcome of improvement in quality of life.
- There is no systematic review of studies of assessment in MLTSS programs. The developer conducted a targeted literature review to gather evidence in support of the measure.
    - The search focused on academic journal articles, gray literature, and federal and state agency reports published in the last 23 years using PubMed, Google, and Google Scholar.
    - The developer also built upon an work done previously in an environmental scan of Assessment and Care Planning measures.
- Medicaid and CHIP Payment and Access Commission (MACPAC) 2016 report highlights variation in assessment elements, mode and timing across states and managed care arrangements limits the ability to make consistent comparisons across states and health plans.
- The developer cites three studies as evidence to support the impact of comprehensive assessment on outcomes for individuals with LTSS needs. The studies are specific to populations who typically need LTSS:
    - Comprehensive Geriatric Assessment (CGA)
    - Geriatric Resources for Assessment and Care of Elders (GRACE)
    - Comprehensive Health Assessment Program (CHAP)
    - Post-Acute Care Tools
- Technical Expert Panel (TEP) provided insight in the development and testing of the measure.

Exception to evidence

- N/A

Questions for the Committee:

- *What is the relationship of this measure to patient outcomes?*
- *How strong is the evidence for this relationship?*
- *Is the evidence directly applicable to the process of care being measured?*

Guidance from the Evidence Algorithm

Process measure no systematic review (box 3) → empirical evidence is submitted without SR and grading (box 7) → Empirical evidence is summarized to include all studies in the body of evidence (box 8) -> High-moderate quality of evidence (box 9) -> Moderate

The highest possible rating is Moderate

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

RATIONALE: N/A

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

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

The developer states that this measure will address the lack of standardization by assessing the percentage of Medicaid MLTSS enrollees who have an assessment conducted with a specified mode (face-to-face, in the home), in a specified timeframe, and addressing specific core and supplemental elements. Performance data is provided from five MTLSS health plans that participated in testing the measure with enrollees representing at least two or more of the major LTSS sub-populations, which include frail older adults age 65 and over, people under age 65 with physical disabilities, people with developmental or intellectual disabilities, and people with serious mental illness.

- On average, 7.9 percent of enrollees across MLTSS plans had documentation of a comprehensive assessment conducted in the specified timeframe, including the specified nine core elements (Rate 1).
    - The range in performance from 0 percent to 26 percent indicates there is substantial room for improvement.
- Only 6.4 percent of enrollees had documentation of nine core elements and at least twelve additional supplemental elements (Rate 2).
    - The range in performance was 0 percent to 22 percent.
- The developer notes that most plans in the sample were regularly conducting assessments with their enrollee population (97 percent of enrollees had documentation of at least one assessment). The low rates are reflective of plans not having documentation of the core elements as defined by this measure.

Rate	Rate 1- Nine core elements documented	Rate 2- Nine core elements documented and twelve supplemental elements documented
Mean	7.9	6.4
Standard Deviation	10.5	8.9
Minimum	0.0	0.0
Maximum	25.5	21.6

- Developer provides additional performance gap rationale for this measure indicating opportunity for improvement:
    - 2013 Commission on Long-Term Care report to Congress "…the development and implementation of a standardized assessment tool that can produce a single care plan across care settings for an individual with cognitive or functional limitations".
    - 2016 CMS final rule for State Medicaid require States to implement "quality assessment and performance improvement programs of services and supports received with those set forth in the enrollee's treatment /service plan".

Disparities

- The developer collected information about race and ethnicity during testing, however due to overall low rates, they did not conduct additional analysis of disparities.
- The developer did not provide any disparities information from the literature regarding the comprehensive assessment addressed in this measure.
- The developer discussed research that identifies racial and ethnic disparities in the need for LTSS. One study from the Congressional Budget Office (CBO) found that older black and Hispanic individuals have higher rates of functional impairment than whites (Congressional Budget Office 2013).
- The developer also cited a report which noted that California Medicaid beneficiaries age 65 and over with disabilities higher instance of complex care needs as well as a greater need for higher instances of care coordination compared to Medicaid beneficiaries under age 65 and non-disables.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Since no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

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

RATIONALE:

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

1a. Evidence

Comments

**The evidence provided by the developer supports the need, importance, and potential value of conducting and updating assessments of the health needs and safety risks of enrollees in LTSS. This goal is directly related to this process measure. Documentation is provided that substantiates the existence of significant variation across state jurisdictions, facilities, and plans for conducting assessments of this type. The measure could be a good starting point for moving the field towards standardizing an assessment process as part of a performance measurement and quality improvement effort.

This is a process measure and the data elements are not directly linked to specific interventions/action items that are intended to achieve specific outcomes.

**The developer presented a combination of empirical data and a targeted literature review, along with information from indicating the MACPAC (Medicaid and CHIP Payment and Access Commission) had indicated a need for standardization of LTSS assessment.

1b. Performance Gap

Comments

**Yes, limited performance data was provided. Documentation indicates a need for standardizing measures and data elements so that an assessment of LTSS documentation can be made across plans. As previously stated there is considerable variability in the type of information that is captured, which prevents the ability to make meaningful comparisons across organizations, plans, states, etc.
While information about race and ethnicity was collected during the testing of the measures, the developer does not include additional analysis of disparities by population group. Data cited from a Congressional Budget Office (CBO) study references disparities in LTSS needs across selected population groups relative to cognitive impairments. The study compared whites to black and Hispanic individuals with cognitive impairments, but data is not available about disparities (racial, ethnic, geographic, age, culture, or other) in the documentation of periodic individual patient need and risk assessments conducted within specified timeframes. Another study was cited, which highlighted the need for more complex care and care coordination services among Medicaid beneficiaries with disabilities who were age 65 years and older in comparison to their younger counter parts.

The developer states the intent of the measures are "to measure the percent of beneficiaries being assessed and the quality of the assessment."
Based on my review, I see where the measures are being used to determine whether individual assessments of enrollee health and safety risks are being consistently conducted and updated within specified timeframes. I questions the measure's focus on the quality of the patient's assessment.
are being consistently conducted and updated within specified timeframes. As such, the measures do not document disparities in care, the quality of the documentation of the assessments, interventions that have been identified to address health or safety risks, or whether there have been any changes in the status of the individual's health or safety risks from baseline to subsequent patient assessment.
**The Gap occurs both in a lack of assessment as well as variability across the states and within the states as to the assessment utilized.
The developer cites three studies as evidence that assessment is impactful on outcomes of persons with LTSS needs.
In my opinion, the assessment alone will not drive outcomes. It is an important step to a care plan - a separate measure. These should be analyzed and implemented in tandem.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications **and** Testing
2b. Validity: Testing; Exclusions**;** Risk-Adjustment**;** Meaningful Differences**;** Comparability**;** Missing Data

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.

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

Evaluators: Staff

Evaluation of Reliability and Validity (and composite construction, if applicable): Staff Evaluation of Scientific Acceptability

Questions for the Committee regarding reliability:

- *Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?*
- *Is the Committee satisfied with the reliability analysis for this measures, and 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., multiple rates, exclusions, risk-adjustment approach, etc.)?*

o *Is the Committee satisfied with the validity analysis for this measures, and is a need to discuss and/or vote on validity?*

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

Scientific Acceptability

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 subcriteria for both reliability and validity to pass this criterion.**

Instructions:

- Please complete this form for each measure you are evaluating.
- Please pay close attention to the skip logic directions.
- If you are unable to check a box, please highlight or shade the box for your response.
- You must answer the "overall rating" item for both Reliability and Validity. Also, be sure to answer the composite measure question at the end of the form if your measure is a composite.
- We have provided TIPS to help you answer the questions.
- We've designed this form to try to minimize the amount of writing that you have to do. That said, ***it is critical that you explain your thinking/rationale if you check boxes where we ask for an explanation*** (because this is a Word document, you can just add your explanation below the checkbox)**.** Feel free to add additional explanation, even if an explanation is not requested (but please type this underneath the appropriate checkbox).
- This form is based on Algorithms 2 and 3 in the Measure Evaluation Criteria and Guidance document (see pages 18-24). These algorithms provide guidance to help you rate the Reliability and Validity subcriteria. ***We ask that you refer to this document when you are evaluating your measures***.
- Please contact Methods Panel staff if you have questions (methodspanel@qualityforum.org).

Measure Number: 3319

Measure Title: Long Term Services and Supports (LTSS) Comprehensive Assessment and Update

RELIABILITY

1. Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?
   NOTE: NQF staff will conduct a separate, more technical, check of eMeasure (eCQM) specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.

   TIPS: Consider the following: Are all the data elements clearly defined? Are all appropriate codes included? Is the logic or calculation algorithm clear? Is it likely this measure can be consistently implemented?

   ☒Yes (go to Question #2)

   ☐No (please explain below, and go to Question #2) NOTE that even though ***non-precise specifications should result in an overall LOW rating for reliability***, we still want you to look at the testing results.

2. Was empirical reliability testing (at the data element or measure score level) conducted using statistical tests with the measure as specified?

   TIPS: Check the 2nd "NO" box below if: only descriptive statistics provided; only describes process for data management/cleaning/computer programming; testing does not match measure specifications (i.e. data, eMeasure, level of analysis, patients)

   ☒Yes (go to Question #4)

☐No, there is reliability testing information, but *not* using statistical tests and/or not for the measure as specified OR there is no reliability testing (please explain below then go to Question #3)

3. Was **empirical VALIDITY testing** of patient-level data conducted?

   ☐Yes (use your rating from data element validity testing – Question #16- under Validity Section)

   ☐No (please explain below and rate Question #11: OVERALL RELIABILITY as INSUFFICIENT and proceed to the VALIDITY SECTION)

4. Was reliability testing conducted with computed performance measure scores for each measured entity?

   TIPS: Answer no if: only one overall score for all patients in sample used for testing patient-level data

   ☒Yes (go to Question #5)

   ☐No (go to Question #8)

5. 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.*

   TIPS: Examples of appropriate methods include signal-to-noise analysis (e.g. Adams/RAND tutorial); random split-half correlation; other accepted method with description of how it assesses reliability of the performance score.

   ☒Yes (go to Question #6) Split sample reliability was assessed using ICC

   ☐No (please explain below then go to Question #8)

6. **RATING (score level)** - What is the level of certainty or confidence that the performance measure scores are reliable?

   TIPS: Consider the following: Is the test sample adequate to generalize for widespread implementation? Do the results demonstrate sufficient reliability so that differences in performance can be identified?

   ☒High (go to Question #8)

   ☐Moderate (go to Question #8)

   ☐Low (please explain below then go to Question #7)

7. Was other reliability testing reported?

   ☒Yes (go to Question #8)

   ☐No (rate Question #11: OVERALL RELIABILITY as LOW and proceed to the VALIDITY SECTION)

8. Was reliability testing conducted with patient-level data elements that are used to construct the performance measure?

   TIPS: Prior reliability studies of the same data elements may be submitted; if comparing abstraction to "authoritative source/gold standard" see Validity Section Question #15)

   ☒Yes (go to Question #9)

   ☐No (if there is score-level testing, rate Question #11: OVERALL RELIABILITY based on score-level rating from Question #6; otherwise, rate Question #11: OVERALL RELIABILITY as INSUFFICIENT. Then proceed to the VALIDITY SECTION)

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

   TIPS: For example: inter-abstractor agreement (ICC, Kappa); other accepted method with description of how it assesses reliability of the data elements

Answer no if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)

☒Yes (go to Question #10) Cohen's kappa statistic used to evaluate IRR

☐No (if no, please explain below and rate Question #10 as INSUFFICIENT)

10. **RATING (data element)** – Based on the reliability statistic and scope of testing (number and representativeness of patients and entities), what is the level of certainty or confidence that the data used in the measure are reliable?

TIPS: Consider the following: Is the test sample adequate to generalize for widespread implementation? Can data elements be collected consistently?

☒Moderate (if score-level testing was NOT conducted, rate Question #11: OVERALL RELIABILITY as MODERATE)

☐Low (if score-level testing was NOT conducted, rate Question #11: OVERALL RELIABILITY as LOW)

☐Insufficient (go to Question #11)

Five of the nine data elements designated as "core" elements for the measure met the threshold for moderate reliability ($\hat{\kappa} \geq 0.4$).

11. OVERALL RELIABILITY RATING

OVERALL RATING OF RELIABILITY taking into account precision of specifications and all testing results:

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

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

☐Low (please explain below) [NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete]

☐Insufficient (please explain below) [NOTE: For most measure types, testing at both the score level and the data element level is not required]

VALIDITY

ASSESSMENT OF THREATS TO VALIDITY

1. Were all potential threats to validity that are relevant to the measure empirically assessed?

TIPS: Threats to validity include: exclusions; need for risk adjustment; Able to identify statistically significant and meaningful differences; multiple sets of specifications; missing data/nonresponse.

☒Yes (go to Question #2)

☐No (please explain below and go to Question #2) [NOTE that even if ***non-assessment of applicable threats should result in an overall INSUFFICENT rating for validity***, we still want you to look at the testing results]

2. Analysis of potential threats to validity: Any concerns with measure exclusions?

TIPS: Consider the following: Are the exclusions consistent with the evidence? Are any patients or patient groups inappropriately excluded from the measure? Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)? If patient preference (e.g., informed decisionmaking) is a basis for exclusion, does it impact performance and if yes, is the measure specified so that the information about patient preference and the effect on the measure is transparent?

☐Yes (please explain below then go to Question #3)

☒No (go to Question #3)

☐Not applicable (i.e., there are no exclusions specified for the measure; go to Question #3)

3. Analysis of potential threats to validity: Risk-adjustment (applies to all outcome, cost, and resource use measures; may also apply to other types of measure)

☒Not applicable (e.g., structure or process measure that is not risk-adjusted; go to Question #4)

a. Is a conceptual rationale for social risk factors included? ☐Yes ☐No

b. Are social risk factors included in risk model? ☐Yes ☐No

c. Any concerns regarding the risk-adjustment approach?

*TIPS: Consider the following: If a justification for **not risk adjusting** is provided, is there any evidence that contradicts the developer's rationale and analysis? If the developer asserts there is **no conceptual basis** for adjusting this measure for social risk factors, do you agree with the rationale? **If risk adjusted**: Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented? Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented? Are all of the risk adjustment variables present at the start of care (if not, do you agree with the rationale)? If social risk factors are not included in the risk-adjustment approach, do you agree with the developer's decision? Is an appropriate risk-adjustment strategy included in the measure (e.g., adequate model discrimination and calibration)? Are all statistical model specifications included, including a "clinical model only" if social risk factors are included in the final model?*

☐Yes (please explain below then go to Question #4)

☐No (go to Question #4)

4. Analysis of potential threats to validity: Any concerns regarding ability to identify meaningful differences in performance or overall poor performance?

☒Yes (please explain below then go to Question #5)

☐No (go to Question #5)

Very poor overall performance observed by developer. However, developer indicates that health plan 01 had rates demonstrating a statistically significant difference from the mean. This finding indicates that the measure as specified does have the potential to identify meaningful differences in performance between health plans.

5. Analysis of potential threats to validity: Any concerns regarding comparability of results if multiple data sources or methods are specified?

☐Yes (please explain below then go to Question #6)

☐No (go to Question #6)

☒Not applicable (go to Question #6)

The developer did not provide an analysis of the comparability of results.

6. Analysis of potential threats to validity: Any concerns regarding missing data?

☐Yes (please explain below then go to Question #7)

☒No (go to Question #7)

ASSESSMENT OF MEASURE TESTING

7. Was empirical validity testing conducted using the measure as specified and appropriate statistical test?

Answer no if: face validity; only refer to clinical evidence; only descriptive statistics; only describe process for data management/cleaning/computer programming; testing does not match measure specifications (i.e. data, eMeasure, level, setting, patients).

☐Yes (go to Question #10) [NOTE: If appropriate empirical testing has been conducted, then evaluation of face validity is not necessary. Go to Question #8 **only if** there is insufficient information provided to evaluate data element and score-level testing.]

☒No (please explain below then go to Question #8)

Score level empirical testing was done, but results were inconclusive (e.g. neither validated, nor invalidated the measure). These may be viewed in the testing submission.

8. Was face validity systematically assessed by recognized experts to determine agreement on whether the computed performance measure score from the measure as specified can be used to distinguish good and poor quality?

   TIPS: Answer no if: focused on data element accuracy/availability/feasibility/other topics; the degree of consensus and any areas of disagreement not provided/discussed.

   ☒Yes (go to Question #9)

   ☐No (please explain below and rate Question #17: OVERALL VALIDITY as INSUFFICIENT)

9. **RATING (face validity)** - Do the face validity testing results indicate substantial agreement that the performance measure score from the measure as specified can be used to distinguish quality AND potential threats to validity are not a problem, OR are adequately addressed so results are not biased?

   ☐Yes (if a NEW measure, rate Question #17: OVERALL VALIDITY as MODERATE)

   ☐ Yes (if a MAINTENANCE measure, do you agree with the justification for not conducting empirical testing? If no, rate Question #17: OVERALL VALIDITY as INSUFFICIENT; otherwise, rate Question #17: OVERALL VALIDITY as MODERATE)

   ☐No (please explain below and rate Question #17: OVERALL VALIDITY AS LOW)

Response	Would high performance on this measure indicate that a health plan is providing higher quality care?	In the future, do you think that performance scores on this measure will distinguish between good and poor performance?
Strongly Agree	1	1
Agree	7	8
Disagree	2	2
Strongly Disagree	3	2
No response	0	0
Total % Agree	62%	69%

The developer provides additional feedback from the TEP from Systematic Assessment of Face Validity:

- The TEP noted that as documentation of the core elements improves among plans over time, the other rates are in turn likely to become more meaningful and useful.
- Rate 2, which reports the percentage of enrollees with all nine core elements and at least 12 supplemental elements, appears the most useful as an "aspirational" measure.
- Health plan performance is slightly lower for Rate 2 relative to Rate 1 (focused on just core elements), but still yields non-zero rates for three of the five health plans.
- Participants in the TEP, the workgroup and public commenters all supported moving forward with this measure.

10. Was validity testing conducted with computed performance measure scores for each measured entity?

TIPS: Answer no if: one overall score for all patients in sample used for testing patient-level data.

☐Yes (go to Question #11)

☐No (please explain below and go to Question #13)

11. Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?

    TIPS: For example: correlation of the performance measure score on this measure and other performance measures; differences in performance scores between groups known to differ on quality; other accepted method with description of how it assesses validity of the performance score

    ☐Yes (go to Question #12) assess convergent validity using Spearman Rank Correlations

    ☐No (please explain below, rate Question #12 as INSUFFICIENT and then go to Question #14)

12. **RATING (measure score)** - Based on the measure score results (significance, strength) and scope of testing (number of measured entities and representativeness) and analysis of potential threats, what is the level of certainty or confidence that the performance measure scores are a valid indicator of quality?

    ☐High (go to Question #14)

    ☐Moderate (go to Question #14)

    ☐Low (please explain below then go to Question #13)

    ☐Insufficient

13. Was other validity testing reported?

    ☐Yes (go to Question #14)

    ☐No (please explain below and rate Question #17: OVERALL VALIDITY as LOW)

14. Was validity testing conducted with patient-level data elements?

    TIPS: Prior validity studies of the same data elements may be submitted

    ☐Yes (go to Question #15)

    ☐No (please explain below and rate Question #17: OVERALL VALIDITY as INSUFFICIENT if no score-level testing was conducted, otherwise, rate Question #17: OVERALL VALIDITY based on score-level rating from Question #12)

    Systematic assessment of face validity surveyed 13 member technical expert panel.

15. 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.*

    TIPS: For example: Data validity/accuracy as compared to authoritative source- sensitivity, specificity, PPV, NPV; other accepted method with description of how it assesses validity of the data elements.

    Answer No if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)

    ☐Yes (go to Question #16)

    ☐No (please explain below and rate Question #16 as INSUFFICIENT)

16. **RATING (data element)** - Based on the data element testing results (significance, strength) and scope of testing (number and representativeness of patients and entities) and analysis of potential threats, what is the level of certainty or confidence that the data used in the measure are valid?

    ☐Moderate (if score-level testing was NOT conducted, rate Question #17: OVERALL VALIDITY as MODERATE)

    ☐Low (please explain below) (if score-level testing was NOT conducted, rate Question #17: OVERALL VALIDITY as LOW)

☐Insufficient (go to Question #17)

17. OVERALL VALIDITY RATING

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

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

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

☐Low (please explain below) [NOTE: Should rate LOW if you believe that there are threats to validity and/or threats to validity were not assessed]

☐Insufficient (if insufficient, please explain below) [NOTE: For most measure types, testing at both the score level and the data element level is not required] [NOTE: If rating is INSUFFICIENT for all empirical testing, then go back to Question #8 and evaluate any face validity that was conducted, then reconsider this overall rating.]

Empirical validity provided was inconclusive (neither validated, nor invalidated the measure). Rating was based on face validity results (which are lower than desired -- most likely due to the extremely low performance rate of the plans in the testing sample).

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

2a1. Reliability- Specifications

Comments

** The developer made some refinements to the list of 9 core and 12 supplemental data elements, which appears to have reduced the level of interpretation among some of the data elements. In other words, such refinements have appeared to clarify how certain elements are defined, making it clearer to utilize the measure.

** My main concern was how information would be gathered. Is this an observational assessment on the patient. The patient, in many cases, may not be able to participate verbally or cognitively in the assessment process. That suggests information to complete the assessment must be gathered from other sources. How will that be accomplished and documented?

2a2. Reliability- Testing

Comments

** No, especially since the refinements were made to clarify the definitions of the data elements.

** No.

2b1. Validity—Testing

2b4-7. Threats to Validity

2b4. Meaningful Differences

Comments

**Face validity was assessed through the use of surveying stakeholder panels. Developer indicates that these surveys helped to validate the intent of the measure, which is “to measure the percent of beneficiaries being assessed and the quality of the assessment.”

I am not clear on what data elements are being considered as a measure of the quality of the assessment. I feel comfortable that the measure does provide an indication of whether a beneficiary was assessed, but not necessarily how well they were assessed.

** Satisfied with validity.

2b2-3. Other Threats to Validity

2b2. Exclusions

2b3. Risk Adjustment

Comments

** No, I did not determine in inappropriate omission of any patient groups.

** N/A

Criterion 3. Feasibility

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 developer provided the following information:

- This measure is primarily collected from health plan and case management records, many of which are electronic.
    - Some data elements are in defined fields in electronic sources
    - The data elements needed for this measure are not standardized (i.e. may be in free text rather than structured fields).
- The developer noted that as the LTSS fields move forward, more of the data elements used in the measure will become structured fields.
- No fees or licensing are currently required.
- This is a multi-rate measure.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Does the Committee agree that measurement in this area will drive standardization?*
- *Does the Committee believe the use of multi-rate for this measure is the best approach?*

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

RATIONALE: Data elements needed for this measure are not currently standardized.

Committee Pre-evaluation Comments: Criteria 3: Feasibility

3. Feasibility
Comments
**I think there may be some variability in the types of data elements that are being collected based on the developer's initial testing of the 28 elements, hence the reason for reducing the number of elements to be measured and avoid a zero percent performance rate among the plans. However, I think the core and supplemental elements identified for the measures are reasonable elements to use in an assessment of an individual's health safety risks. Further, I think it is feasible to collect this information, particularly if plans are informed of clearly defined data elements that will be collected.
**Not an eComm. Records are drawn from health plan and case management records, which may be electronic. Data collection is not currently standardized. A date for standardization and electronic collectivity would be helpful.

Criterion 4: Usability and Use

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 is a new measure and is not currently in use.
- This measure is intended for use by states and health plans to monitor and improve quality of care provided for the Medicaid MLTSS enrollee population.
- This measure is included in the National MLTSS Health Plan Association recommended LTSS performance measure set model.

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

- N/A

Additional Feedback:

- N/A

***Questions for the Committee**:*

- o *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- o *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 is a new measure and improvement information was not provided

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

Unexpected findings (positive or negative) during implementation

- The developer reported that no unintended consequences were identified during testing.

Potential harms

- N/A

Additional Feedback:

- N/A

***Questions for the Committee**:*

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

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

RATIONALE:

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

4a1. Use- Accountability and Transparency

Comments

**This is a new measure that is currently not in use.

**High on both. Although current assessment seems to be below 10%. Uptake may take awhile.

Criterion 5: Related and Competing Measures

Related or competing measures

Related measures include:

- 3324 LTSS Comprehensive Care Plan and Update
- 3325 LTSS Shared Care Plan with Primary Care Practitioner
- 3326 LTSS Re-Assessment/Care Plan Update after Inpatient Discharge

Harmonization

N/A

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

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: January 18, 2018

- **NQF received zero public comments on this measure.**
- **Zero NQF members have submitted a support/non-support choice.**

Developer Submission

Measure Information

This document contains the information submitted by measure developers/stewards, but is organized according to NQF's measure evaluation criteria and process. The item numbers refer to those in the submission form but may be in a slightly different order here. In general, the item numbers also reference the related criteria (e.g., item 1b.1 relates to sub criterion 1b).

Brief Measure Information

NQF #: 3319

Corresponding Measures:

De.2. Measure Title: Long Term Services and Supports (LTSS) Comprehensive Assessment and Update

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

De.3. Brief Description of Measure: This measure assesses the percentage of Managed Long Term Services and Support (MLTSS) plan enrollees who have documentation of a comprehensive assessment in a specified timeframe that includes documentation of core and supplemental elements. This measure has two rates:

Rate 1: Percent of MLTSS plan enrollees with documentation of a comprehensive LTSS assessment including nine (9) core elements within 90 days of enrollment or at least annually.

Rate 2: Percent of MLTSS plan enrollees with documentation of a comprehensive LTSS assessment including nine (9) core elements AND at least twelve (12) supplemental elements within 90 days of enrollment or at least annually.

1b.1. Developer Rationale: Comprehensive assessments serve as the foundation for developing comprehensive person-centered care plans, providing all needed services and supports of high quality, and performing well-coordinated care. However, the tools currently used by states and health plans to conduct assessments, and the performance measures used to evaluate the quality of assessments conducted vary widely. This measure would address the lack of standardization by assessing the percentage of Medicaid MLTSS enrollees who have a comprehensive assessment conducted that includes specific core and supplemental elements. A standardized measure of comprehensive assessment will allow for apples-to-apples comparisons of MLTSS plans across states.

To construct this measure, we followed a multistep process that included evidence review, empiric analysis and stakeholder input. To start, we conducted an environmental scan of assessment domains and elements required for the programs serving adults with LTSS needs (e.g., Special Needs Plans, Medicare-Medicaid Plans, 1915c waivers, Medicare Home Health, Nursing Facility Minimum Data Set, Program for all Inclusive Care of the Elderly). To augment our environmental scan, we conducted one-on-one interviews with key stakeholders and a Technical Expert Panel (TEP) convened in 2013 to solicit additional assessment domains. Through review and voting, the TEP identified 28 data elements as critical for comprehensive LTSS assessment. However, through our field test, we determined that requiring all 28 elements was too stringent and would not result in meaningful performance rates (i.e., four out of five plans had a 0% performance rate when all 28 elements were required).

To balance the need for a measure that is feasible in the current MLTSS environment that would produce non-zero rates and the desire of the TEP to push the field further to do a truly comprehensive assessment, we decided to develop two rates for the measure.

1. "Core" requirements that set a minimum baseline of performance, and

2. "Supplemental" requirements that demonstrate more thorough and comprehensive performance.

Over time, we anticipate that elements from the "supplemental" requirements will move to the "core" requirements as performance improves. In the meantime, the currently proposed "core" rates can fill a long-standing measurement gap while generating results that are both meaningful and usable to stakeholders. Details on how the core and supplemental elements were selected can be found in the Appendix: Additional Testing Results.

The two-rate approach aims to give program administrators and policymakers a much-needed tool to help move the managed long-term services and supports (MLTSS) field from extremely low performance (as demonstrated in our testing) to more comprehensive and standardized documentation of assessment, care planning and care coordination practices. The inclusion of both rates in a single measure signals the importance of this concept.

S.4. Numerator Statement: The measure has two rates. The numerators for the two rates are as follows:

Rate 1: MLTSS plan enrollees who had either of the following:

• A comprehensive LTSS assessment completed within 90 days of enrollment for new enrollees, with nine (9) core elements documented, or

• A comprehensive LTSS assessment completed at least once during the measurement year for all other enrollees (established enrollees), with nine (9) core elements documented.

Rate 2: MLTSS plan enrollees who had either of the following:

• A comprehensive LTSS assessment completed within 90 days of enrollment for new enrollees, with nine (9) core and at least twelve (12) supplemental elements documented, or

• A comprehensive LTSS assessment completed at least once during the measurement year for all other enrollees (established enrollees) with nine (9) core and at least twelve (12) supplemental elements documented.

Note: Initial assessment should be completed within 90 days of enrollment, and updated annually thereafter.

S.6. Denominator Statement: Medicaid MLTSS plan enrollees age 18 years and older.

S.8. Denominator Exclusions: Exclude enrollees in the denominator who were enrolled in the plan prior to September 1 of the year prior to the measurement year (i.e., established enrollees) who left the plan for more than 45 days between January 1 and December 31 of the measurement year.

Exclude enrollees who could not be reached for a comprehensive assessment or who refused a comprehensive assessment.

De.1. Measure Type: Process

S.17. Data Source: Management Data, Other, Paper Medical Records

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:

3320:LTSS Comphrensive Assessment, Care Planning, and Coordination

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? This measure is grouped with two other measures that assess the continuum of assessment, care planning and care coordination. This continuum of care is described in greater detail in the accompanying Evidence Attachment form.

• LTSS Comprehensive Care Plan and Update

• LTSS Shared Care Plan with Primary Care Practitioner

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

LTSS_Comp_Assess_Evidence_Attachment.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*)**:**

Measure Title: LTSS Comprehensive Assessment and Update

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

Date of Submission: 11/7/2017

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:

☐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*):

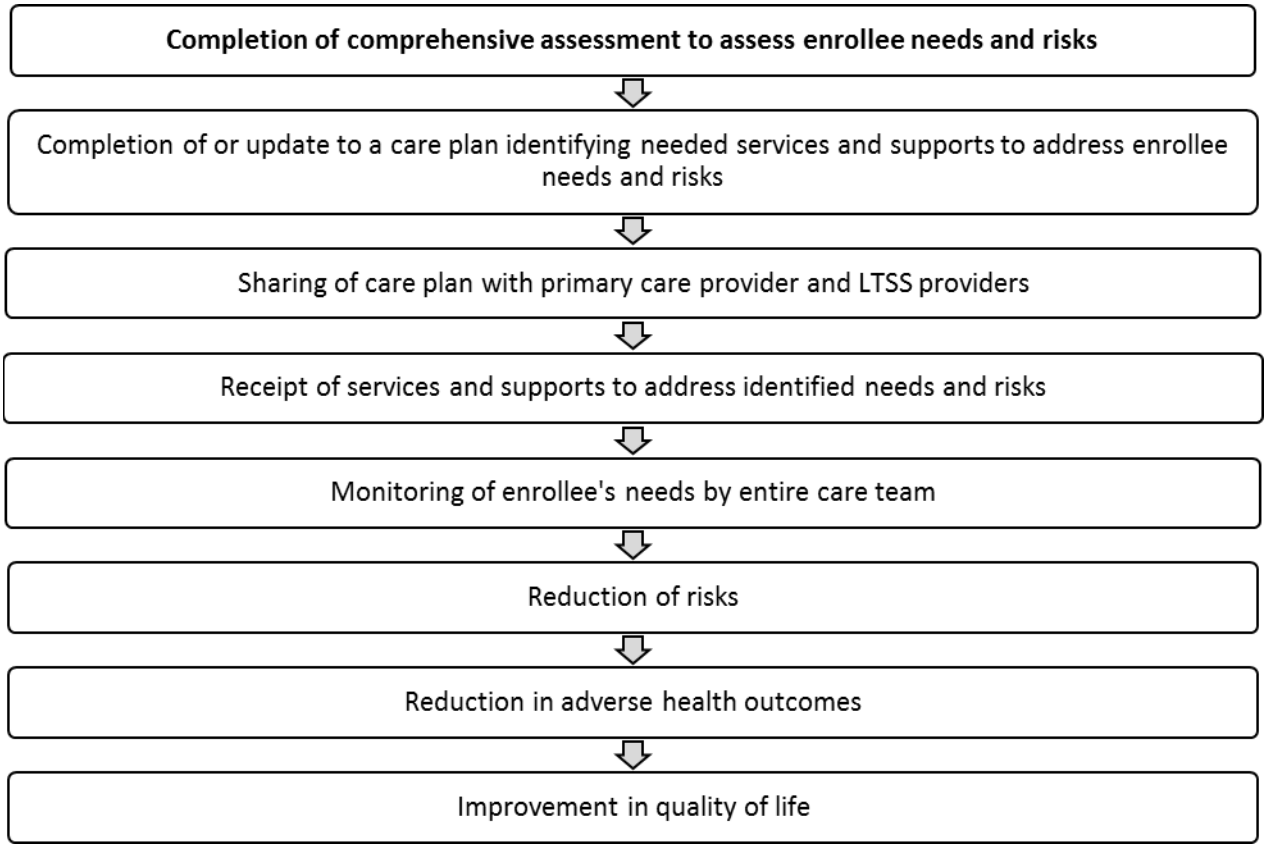
☒ Process: : This measure assesses the extent to which Managed Long Term Services and Support (MLTSS) enrollees receive a comprehensive assessment for provision of long term services and supports.

☐ 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.



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 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.

Not applicable. Not an outcome measure.

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

Not applicable. Evidence is not based on a systematic review

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.

In the absence of a systematic review of studies of assessment in MLTSS programs, the project team conducted a targeted literature review to gather evidence in support of this measure. We searched for academic journal articles, gray literature, and federal and state agency reports published in the last 23 years using PubMed, Google, and Google Scholar.

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

The Medicaid MLTSS enrollee population includes individuals with complex health and social support needs, such as individuals with physical, cognitive, and mental disabilities and older adults with multiple functional limitations and chronic conditions (MACPAC, 2016; KFF, 2015). Given their complex needs, MLTSS enrollees often require a wide range of services and supports and high levels of care coordination (Saucier & Burwell, 2015). Delivering effective care coordination for complex populations, such as MLTSS enrollees, begins with conducting and regularly updating comprehensive assessments to identify a wide array of enrollee needs and potential health and safety risks (Rich et al., 2012).

Comprehensive assessments serve as the foundation for developing comprehensive person-centered care plans, providing all needed services and supports of high quality, and performing well-coordinated care. However, the tools currently used by states and health plans to conduct assessments, and the performance measures used to evaluate the

quality of assessments conducted vary widely (MACPAC, 2016b). This measure would address the lack of standardization by assessing the percentage of Medicaid LTSS enrollees who have a comprehensive assessment conducted that includes specific core and supplemental elements. A standardized measure of comprehensive assessment will allow for apples-to-apples comparisons of LTSS plans across states.

Variation in How Comprehensive Assessment is Defined and Conducted

State Medicaid agencies have implemented numerous MLTSS care coordination models, and most require an assessment at initial enrollment and on a regular basis thereafter (Saucier & Burwell, 2015). However, the tools used to conduct assessments and the performance measures used to evaluate the quality of assessments conducted vary widely (MACPAC, 2016; KFF, 2015; Atkins & Gage, 2014).

A recent analysis by Medicaid and CHIP Payment and Access Commission (MACPAC) identified at least 124 assessment tools currently in use by states to assess functional status (MACPAC 2016). An environmental scan conducted in 2012 under a previous CMS contract (Prime Contract No. HHSM-500-2010-00026I/HHSM-500-T0011) also highlighted this variation particularly for MLTSS plans. In some states MLTSS plans use a state-mandated assessment instrument, in other states MLTSS plans conduct their own assessment in addition to a state "level-of-care" assessment. Some states require assessments to be in person and others do not specify the mode or location of assessment. The variation in assessment elements, mode and timing across states and managed care arrangements limits the ability to make apples-to-apples comparisons across states and health plans.

Evidence to Support Impact of Comprehensive Assessment on Outcomes

We were unable to find a systematic review evaluating the impact of comprehensive assessment on outcomes for individuals with LTSS needs. However, several studies of assessments for populations who typically need LTSS, including older adults and adults with intellectual and developmental disabilities, demonstrate the critical importance of conducting comprehensive assessments as a precursor to the development of person-centered care plans and the coordination of care across providers and settings, and when performed together in care coordination interventions improve health outcomes.

Example 1: Comprehensive Geriatric Assessment (CGA)

CGA is defined as a "multidisciplinary diagnostic and treatment process that identifies medical, psychosocial, and functional limitations of a frail older person in order to develop a coordinated plan to maximize overall health with aging," (Ward, & Reuben 2016). A meta-analysis of 28 controlled trials found that CGA programs linking geriatric evaluation with strong long-term management were effective for improving survival and function in older adults (Stuck, et al., 1993). More recent studies have found that, when used in the hospital setting, CGA can also lead to increased in-home residence up to 12 months post-discharge , and, in the ambulatory care setting, to reduced length of hospital stays and increased sense of security in care interactions (Ellis, et al., 2011; Avelino-Silvia et al., 2014; Ekdahl, et al., 2015).

Example 2: Geriatric Resources for Assessment and Care of Elders (GRACE)

GRACE is an integrated care model that targets low-income seniors, many dually eligible and most with multiple chronic conditions. The model uses in-home assessments by a team consisting of a nurse practitioner and social worker to develop individualized care plans (Bielaszka-DuVernay, 2011; Counsell, et al., 2006; Counsell, et al., 2007; Counsell et al., 2009). A randomized controlled trial found that high-risk patients enrolled in GRACE had fewer emergency department visits, hospitalizations, and readmissions and reduced hospital costs compared to a control group. In addition, the GRACE model saved $1,500 per enrolled high-risk patient by its second year. Finally, the GRACE model received higher care satisfaction ratings by physicians and quality of life reports by patients compared to a control group.

Example 3: Comprehensive Health Assessment Program (CHAP)

Similar to older adults, persons with intellectual disabilities often have unrecognized health conditions, impaired communication, and cognition and recall difficulties and benefit from comprehensive health assessments (Cooper et al., 2006; Lennox et al., 2001; Webb & Rogers, 1999).

CHAP is a comprehensive assessment tool developed and tested in New Zealand and used to evaluate medical histories, conduct targeted examinations, assess for syndrome-specific comorbidities, and develop action plans for persons with

intellectual disabilities. A randomized controlled trial found that CHAP increased provider awareness of health needs of persons with intellectual disabilities and disease detection (Lennox, et al., 2007). A more recent study including interviews and focus groups with various stakeholders (i.e., physicians, nurse practitioners, support workers, and families) determined that the CHAP was beneficial for persons with intellectual disabilities, including greater continuity of care, and was strongly supported for use in Canada (Shooshtari, et al., 2016).

Example 4: Post-Acute Care Tools

Assessments are also used routinely in the post-acute care setting (home health agencies, skilled nursing facilities, inpatient rehabilitation facilities, and long-term care hospitals) to identify patient needs, potential health risks, and monitor outcomes.

- Long-Term Care Minimum Data Set (MDS): A standardized screening and assessment tool of health status that serves as the basis of a comprehensive assessment for all residents in a Medicare and/or Medicaid-certified long-term care facility.
- Outcome and Assessment Information Set (OASIS): A group of data elements that dictates core items of a comprehensive assessment for adult home health care patients and serves as the basis for measuring patient outcomes.
- Continuity Assessment Record and Evaluation (CARE): A tool developed by the U.S. Department of Health and Human Services to assess patients' needs for post-acute services in the four settings listed above. The CARE item set builds on the MDS and OASIS instruments.

All of these assessment tools have been shown to be reliable, valid, and useful for identifying patients' health care and social support needs and developing individualized care plans (Centers for Medicare & Medicaid Services, 2012; CMS, 2012; CMS, 2015).

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

In the absence of a systematic review, the project team conducted a targeted literature review to gather evidence in support of this measure. We searched for academic journal articles, gray literature, and federal and state agency reports published in the last 23 years using PubMed, Google, and Google Scholar. We convened a TEP in 2013 to provide insight into the priority areas for measurement and the usefulness and feasibility of the identified measures for MLTSS plans. The 2013 TEP was comprised of individuals representing multiple perspectives from the MLTSS community including consumers, practitioners, health plans, the federal government, and state governments. A second TEP was convened in 2015 with a similar composition to inform the continued development and testing of this measure.

We also built upon an environmental scan of Assessment and Care Planning measures conducted under a CMS contract (Contract No. HHSM-500-2010-00026I/HHSM-500-T0011).

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

Atkins, G. L., & B. Gage. (2014). The Need to Standardize Assessment Items for Persons in Need of LTSS. Available at http://www.ltqa.org/wp-content/themes/ltqaMain/custom/images/LTQA-The-Need-to-Standardize-Assessment-Items-4-14-1.pdf.

Avelino-Silva, T. J., Farfel, J. M., Curiati, J. A., Amaral, J. R., Campora, F., & Jacob-Filho, W. (2014). Comprehensive geriatric assessment predicts mortality and adverse outcomes in hospitalized older adults. BMC Geriatrics, 14, 129.

Bielaszka-DuVernay, C. (2011). The 'GRACE' Model: In-Home Assessments Lead to Better Care for Dual-eligibles. Health Affairs, 30(3), 431-434.

Centers for Medicare & Medicaid Services (CMS). (2012). Long Term Care Minimum Data Set (MDS). Available at https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/IdentifiableDataFiles/LongTermCareMinimumDataSetMDS.html.

CMS. (2012). Outcome and Assessment Information Set (OASIS). Available at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/OASIS/index.html

CMS. (2015). CARE Item Set and B-CARE. Available at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/CARE-Item-Set-and-B-CARE.html.

Cooper, S. A., J. Morrison, C. Melville, et al. (2006). Improving the health of people with intellectual disabilities: outcomes of a health screening programme after 1 year. Journal of Intellectual Disability Research, 50, 667-677.

Counsell, S. R., C. M. Callahan, A. B. Buttar, D. O. Clark, & K. I. Frank. (2006). Geriatric Resources for Assessment and Care of Elders (GRACE): a new model of primary care for low-income seniors. Journal of the American Geriatrics Society, 54(7), 1136-1141.

Counsell, S. R., C. M. Callahan, D. O. Clark, W. Tu, A. B. Buttar, T. E. Stump, & G. D. Ricketts. (2007). Geriatric care management for low-income seniors: a randomized controlled trial. Journal of the American Medical Association, 298(22), 2623-2633.

Counsell, S. R., C. M. Callahan, W. Tu, T. E. Stump, & G. W. Arling. (2009). Cost analysis of the Geriatric Resources for Assessment and Care of Elders care management intervention. Journal of the American Geriatrics Society, 57(8), 1420-1426.

Ekdahl, A. W., Wirehn, A. B., Alwin, J., Jaarsma, T., Unosson, M., Husberg, M.,Carlsson, P. (2015). Costs and Effects of an Ambulatory Geriatric Unit (the AGe-FIT Study): A Randomized Controlled Trial. Journal of the American Medical Directors Association.

Ellis, G., Whitehead, M. A., O'Neill, D., Langhorne, P., & Robinson, D. (2011). Comprehensive geriatric assessment for older adults admitted to hospital. Cochrane Database of Systematic Reviews, 7, CD006211.

Kaiser Family Foundation (KFF). (2015). Medicaid and Long-Term Services and Supports: A Primer. Available at http://kff.org/medicaid/report/medicaid-and-long-term-services-and-supports-a-primer/.

Lennox, N. G., M. Green, J. Diggens, & A. Ugoni. (2001). Audit and comprehensive health assessment programme in the primary healthcare of adults with intellectual disability: a pilot study. Journal of Intellectual Disability Research, 45, 226-232.

Lennox, N., C. Bain, T. Rey-Conde, D. Purdie, R. Bush, & N. Pandeya. (2007). Effects of a comprehensive health assessment programme for Australian adults with intellectual disability: a cluster randomized trial. International Journal of Epidemiology, 36(1), 139-146.

MACPAC. (2016). Chapter 4. Functional Assessments for Long-Term Services and Supports. Report to Congress on Medicaid and CHIP. Available at https://www.macpac.gov/wp-content/uploads/2016/06/Functional-Assessments-for-Long-Term-Services-and-Supports.pdf.

MACPAC. (2016b). Chapter 4. Functional Assessments for Long-Term Services and Supports. Report to Congress on Medicaid and CHIP. Available at https://www.macpac.gov/wp-content/uploads/2016/06/Functional-Assessments-for-Long-Term-Services-and-Supports.pdf.

Medicaid and CHIP Payment and Access Commission (MACPAC). (2016). Users of long-term services and supports. Available at https://www.macpac.gov/subtopic/long-term-services-and-supports-population/.

Rich, E., D. Lipson, J. Libersky, and M. Parchman. (2012). Coordinating Care for Adults with Complex Care Needs in the Patient-Centered Medical Home: Challenges and Solutions. Agency for Healthcare Research and Quality (AHRQ). Available at https://pcmh.ahrq.gov/sites/default/files/attachments/coordinating-care-for-adults-with-complex-care-needs-white-paper.pdf

Saucier, P., & B. Burwell. (2015). Care Coordination in Managed Long-Term Services and Supports. Washington, DC: AARP Public Policy Institute. Available at http://www.aarp.org/content/dam/aarp/ppi/2015/care-coordination-in-managed-long-term-services-and-supports-report.pdf.

Shooshtari, S., B. Temple, C. Waldman, S. Abraham, H. Ouellette-Kuntz, & N. Lennox. (2016). Stakeholders' Perspectives towards the Use of the Comprehensive Health Assessment Program (CHAP) for Adults with Intellectual Disabilities in Manitoba. Journal of Applied Research in Intellectual Disabilities.

Stuck, A. E., A. L. Siu, G. D. Wieland, J. Adams, & L. Z. Rubenstein. (1993). Comprehensive geriatric assessment: a meta-analysis of controlled trials. Lancet, 342(8878), 1032-1036.

Ward, K. T., & D. Reuben. (2016). Comprehensive geriatric assessment. Available at http://www.uptodate.com/contents/comprehensive-geriatric-assessment.

Webb, O., & L. Rogers. (1999). Health screening for people with intellectual disability: the New Zealand experience. Journal of Intellectual Disability Research, 43, 497-503.

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)*

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

Comprehensive assessments serve as the foundation for developing comprehensive person-centered care plans, providing all needed services and supports of high quality, and performing well-coordinated care. However, the tools currently used by states and health plans to conduct assessments, and the performance measures used to evaluate the quality of assessments conducted vary widely. This measure would address the lack of standardization by assessing the percentage of Medicaid MLTSS enrollees who have a comprehensive assessment conducted that includes specific core and supplemental elements. A standardized measure of comprehensive assessment will allow for apples-to-apples comparisons of MLTSS plans across states.

To construct this measure, we followed a multistep process that included evidence review, empiric analysis and stakeholder input. To start, we conducted an environmental scan of assessment domains and elements required for the programs serving adults with LTSS needs (e.g., Special Needs Plans, Medicare-Medicaid Plans, 1915c waivers, Medicare Home Health, Nursing Facility Minimum Data Set, Program for all Inclusive Care of the Elderly). To augment our environmental scan, we conducted one-on-one interviews with key stakeholders and a Technical Expert Panel (TEP) convened in 2013 to solicit additional assessment domains. Through review and voting, the TEP identified 28 data elements as critical for comprehensive LTSS assessment. However, through our field test, we determined that requiring all 28 elements was too stringent and would not result in meaningful performance rates (i.e., four out of five plans had a 0% performance rate when all 28 elements were required).

To balance the need for a measure that is feasible in the current MLTSS environment that would produce non-zero rates and the desire of the TEP to push the field further to do a truly comprehensive assessment, we decided to develop two rates for the measure.

1. “Core” requirements that set a minimum baseline of performance, and

2. “Supplemental” requirements that demonstrate more thorough and comprehensive performance.

Over time, we anticipate that elements from the “supplemental” requirements will move to the “core” requirements as performance improves. In the meantime, the currently proposed “core” rates can fill a long-standing measurement gap while generating results that are both meaningful and usable to stakeholders. Details on how the core and supplemental elements were selected can be found in the Appendix: Additional Testing Results.

The two-rate approach aims to give program administrators and policymakers a much-needed tool to help move the managed long-term services and supports (MLTSS) field from extremely low performance (as demonstrated in our

testing) to more comprehensive and standardized documentation of assessment, care planning and care coordination practices. The inclusion of both rates in a single measure signals the importance of this concept.

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 five MLTSS health plans that participated in testing these measures with enrollees representing at least two or more of the major LTSS sub-populations, which include frail older adults age 65 and over, people under age 65 with physical disabilities, people with developmental or intellectual disabilities, and people with serious mental illness. For purposes of testing, the eligible population included MLTSS members who were age 18 and older as of January 1, 2015 and enrolled on or prior to August 31, 2015, allowing for at least 120 days of enrollment during the measurement year of CY 2015.

On average, 7.9 percent of enrollees across MLTSS plans had documentation of a comprehensive assessment conducted in the specified timeframe, including the specified nine core elements (Rate 1). The range in performance from 0 percent to 26 percent indicates there is substantial room for improvement. Only 6.4 percent of enrollees had documentation of nine core elements and at least twelve additional supplemental elements. The range in performance was 0 percent to 22 percent.

It is important to note that although these data show low rates of performance, most plans in the sample were regularly conducting assessments with their enrollee population (97 percent of enrollees had documentation of at least one assessment). The low rates are reflective of plans not having documentation of the core elements as defined by this measure.

Percent of enrollees with Rate 1. Nine (9) core elements documented:

Mean: 7.9

Standard Deviation: 10.5

Minimum: 0.0

Maximum: 25.5

Percent of enrollees with Rate 2. Nine (9) core elements documented and twelve (12) supplemental elements documented:

Mean: 6.4

Standard Deviation: 8.9

Minimum: 0.0

Maximum: 21.6

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.

Comprehensive assessments serve as the foundation for developing comprehensive person-centered care plans, providing all needed services and supports of high quality and performing well-coordinated care. While almost all MLTSS plans perform initial and annual assessment of members, the lack of a standardized measure to assess the degree to which assessments among the MLTSS enrollee population are comprehensive has precluded the collection of comparable data across plans.

A central challenge to measuring the rate of assessment is the variation in the way assessments are conducted across states and health plans. The tools used to conduct assessments and the performance measures used to evaluate the quality of assessments conducted vary widely. A recent review by MACPAC in 2016 found that over 124 tools are

currently in use (MACPAC, 2016). On average, states use three different tools each, as they generally use separate tools for different populations.

In its 2013 report to Congress, the Commission on Long-Term Care called for "…the development and implementation of a standardized assessment tool that can produce a single care plan across care settings for an individual with cognitive or functional limitations," (Atkins & Gage, 2014). More recently in the May 6, 2016 Federal Register, CMS issued a final rule that requires State Medicaid agencies that operate MLTSS programs to implement "mechanisms to detect both underutilization and overutilization of services and the quality and appropriateness of care furnished to enrollees with special health care needs," (CMS, 2016). In addition, the rule requires States to implement "quality assessment and performance improvement programs for plans offering LTSS [which] must include assessments of care between care settings and comparisons of services and supports received with those set forth in the enrollee's treatment/service plan," (ICRC, 2016). Both sets of requirements, which go into effect for rating periods for contracts starting on or after July 1, 2017, rely on a comprehensive assessment of MLTSS enrollees' needs.

This measure would address the lack of standardization by assessing the percentage of Medicaid MLTSS enrollees who have an assessment conducted with a specified mode (face-to-face, in the home), in a specified timeframe, and addressing specific core and supplemental elements.

Atkins, G. L., & B. Gage. (2014). The Need to Standardize Assessment Items for Persons in Need of LTSS. Available at http://www.ltqa.org/wp-content/themes/ltqaMain/custom/images/LTQA-The-Need-to-Standardize-Assessment-Items-4-14-1.pdf.

CMS. (2016). 42 CFR Parts 431, 433, 438, et al. Medicaid and Children's Health Insurance Program (CHIP) Programs; Medicaid Managed Care, CHIP Delivered in Managed Care, and Revisions Related to Third Party Liability; Final Rule. Available at https://www.gpo.gov/fdsys/pkg/FR-2016-05-06/pdf/2016-09581.pdf.

Integrated Care Resource Center (ICRC). (2016). Spotlight: CMS Medicaid Managed Care Final Rule – Provisions Related to Integrated Programs for Medicare-Medicaid Enrollees. Available at http://www.integratedcareresourcecenter.com/PDFs/2016%2005%2012%20Medicaid%20Managed%20Care%20Regulations.pdf

MACPAC. (2016). Chapter 4. Functional Assessments for Long-Term Services and Supports. Report to Congress on Medicaid and CHIP. Available at https://www.macpac.gov/wp-content/uploads/2016/06/Functional-Assessments-for-Long-Term-Services-and-Supports.pdf.

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.*

Not applicable. We collected information about race and ethnicity during testing. However, due to the overall low rates, we did not believe additional analysis of disparities would provide meaningful information.

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

We could not find any research on disparities in performing comprehensive assessments among the MLTSS enrollee population. However, studies have identified disparities in the need for and use of LTSS more broadly, which highlight the need for more comprehensive and well-documented assessments.

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). It also found that fewer than 60 percent of newly transitioned seniors and persons with disabilities were successfully contacted and administered a health risk assessment, which is much less intensive than the comprehensive assessment required by this measure.

We collected information about race and ethnicity during testing. However, due to the overall low rates, we did not believe additional analysis of disparities would provide meaningful information.

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.

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.)*

Measure specifications have been drafted, and are anticipated to be publicly posted on CMS's MLTSS website in early 2018. However, currently there is no link.

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)*

No data dictionary **Attachment:**

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.

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

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.*

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

The measure has two rates. The numerators for the two rates are as follows:

Rate 1: MLTSS plan enrollees who had either of the following:

- A comprehensive LTSS assessment completed within 90 days of enrollment for new enrollees, with nine (9) core elements documented, or
- A comprehensive LTSS assessment completed at least once during the measurement year for all other enrollees (established enrollees), with nine (9) core elements documented.
- Rate 2: MLTSS plan enrollees who had either of the following:
- A comprehensive LTSS assessment completed within 90 days of enrollment for new enrollees, with nine (9) core and at least twelve (12) supplemental elements documented, or
- A comprehensive LTSS assessment completed at least once during the measurement year for all other enrollees (established enrollees) with nine (9) core and at least twelve (12) supplemental elements documented.

Note: Initial assessment should be completed within 90 days of enrollment, and updated annually thereafter.

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)*

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

Time Period for Data:

16 months (September 1 of the year prior to the measurement year to December 31 of the measurement year).

The numerator details for the two rates are as follows:

Rate 1: MLTSS plan enrollees who had either of the following:

- A comprehensive LTSS assessment (face-to-face, in the home) with nine (9) core elements documented within 90 days of enrollment for enrollees newly enrolled in the plan between September 1 of the year prior to the measurement year and August 31 of the measurement year, or

- A comprehensive LTSS assessment (face-to-face, in the home) with nine (9) core elements documented during the measurement year for enrollees enrolled in the plan prior to September 1 of the year prior to the measurement year.

The comprehensive assessment must document current enrollee status on nine (9) core elements. Documentation of "no change" is not sufficient to meet numerator criteria. The date of the comprehensive assessment must be documented.

Core Elements

1. Limitations in activities of daily living (ADLs): Any difficulty in performing ADLs without assistance (i.e., walking, toileting, bathing, dressing, eating, and transferring) must be documented. Ability to perform all six ADLs must be documented.

2. Acute and chronic health conditions

3. List of current medications (The medication list may include medication names only)

4. Cognitive function assessed using a standardized validated tool (e.g., AD8 = Eight-item Informant Interview to Differentiate Aging and Dementia; AWV = Annual Wellness Visit; GPCOG = General Practitioner Assessment of Cognition; HRA = Health Risk Assessment; MIS = Memory Impairment Screen; MMSE = Mini Mental Status Exam; MoCA = Montreal Cognitive Assessment; SLUMS = St. Louis University Mental Status Exam; Short IQCODE = Short Informant Questionnaire on Cognitive Decline in the Elderly)

5. Mental health status (e.g., Patient Health Questionnaire (PHQ9), Beck Depression Inventory (BDI or BDI-II), Center for Epidemiologic Studies Depression Scale (CES-D), Depression Scale (DEPS), Duke Anxiety-Depression Scale (DADS), Geriatric Depression Scale (GDS), Cornell Scale Screening,and PRIME MD-PHQ2, Generalized Anxiety Disorder 7-Item Scale (GAD7)

6. Home safety risks (e.g., home fall risks, bathroom safety, chemical hazards, food preparation safety)

7. Living arrangement: Documentation of whether member lives in a nursing facility, institution, assisted living, general community or other setting.

8. Family and Friend Caregiver Availability: Documentation of whether any family or friend caregivers are providing paid or unpaid assistance to the enrollee (assistance with activities of daily living, instrumental activities of daily living, health care related tasks, or emotional support). The availability of a friend or a family caregiver (paid or unpaid) to provide caregiving support in the future must be documented along with the contact information for said caregivers. If there is no friend or family caregiver, the lack of informal caregiver availability must be documented to meet this element.

9. Current providers including primary care practitioner

Rate 2: MLTSS plan enrollees who had either of the following:

- A comprehensive LTSS assessment (face-to-face, in the home) with nine (9) core and at least twelve (12) supplemental elements within 90 days of enrollment for enrollees newly enrolled in the plan between September 1 of the year prior to the measurement year and August 31 of the measurement year, or

- A comprehensive LTSS assessment (face-to-face, in the home) with nine (9) core and at least twelve (12) supplemental elements during the measurement year for enrollees enrolled in the plan prior to September 1 of the year prior to the measurement year.

The assessment must document current enrollee status on the nine (9) core elements and at least twelve (12) of the supplemental elements described below. Documentation of “no change” is not sufficient to meet numerator criteria. The date of the comprehensive assessment must be documented.

Supplemental Elements

1. Instrumental activities in daily living (IADLs): Any difficulty in performing IADLs without assistance (i.e., using the telephone, managing money, preparing meals, doing light or heavy housework, and shopping for personal items). Documentation of at least five IADLs is necessary to meet this item.

2. Current use of accommodations related to the physical disability, such as use of assistive technology

3. Enrollee’s self-reported health status using a standardized validated tool or question (e.g., Would you say your health in general is.., Short-Form Survey -12 (SF-12), Patient Reported Outcome Measurement Information System (PROMIS) Global 10).

4. Behavior difficulties (e.g., wandering, aggression)

5. Patient activation or self-efficacy assessed using a standardized validated tool (e.g., Patient Activation Measure (PAM), Stanford Chronic Disease Self-Efficacy Scale)

6. Vision needs. Documentation must include whether an individual has an impairment in vision and whether they use any devices to address that need.

7. Hearing needs. Documentation must include whether an individual has an impairment in hearing and whether they use any devices to address that need.

8. Speech needs. Documentation must include whether an individual has an impairment in speech and whether they use any devices to address that need.

9. Physical/occupational therapy needs. Documentation must include whether there is a need for physical or occupational therapy.

10. Falls risk (e.g., documentation of history of falls, problem with gait or balance, or other falls risk factors)

11. Alcohol and other drug use

12. Smoking status

13. Availability of public and plan benefits (e.g., eligibility for Medicare, Medicaid, Supplemental Security Income, transportation services, food subsidies, electric/gas subsidies, or housing subsidies)

14. Availability of social support in community (e.g., support from friends, community based services, or other non-medical based services provided to the individual)

15. Assessment of social isolation, loneliness or other social issues

16. Cultural and linguistic preferences (e.g., preferred language, communication style)

17. Advance care plan (e.g., living will, health care power of attorney, health care proxy, Physician Orders for Life Sustaining Treatment [POLST], Five Wishes, documented preferences for life-sustaining treatment and end-of-life care, or documented surrogate decision maker) or enrollee refusal of advance care planning.

18. Preference for participating in work or volunteer activities

19. Recent use of services that may include emergency department, hospitalization, home health, skilled nursing facility, paid home care, homemaker, or other services.

Rate 1 & 2: Additional Notes

A comprehensive assessment must include a face-to-face discussion with the enrollee in the home using a structured or semi-structured tool that assesses the enrollee's health status and needs. Home is defined as the location where the member is currently residing and considers their long-term residence including assisted living facilities and long-term care facilities. The requirement to have in-home assessment is waived if the member refuses to have the assessment conducted in their home or the member is residing temporarily in an inpatient facility at the time of assessment.

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

Medicaid MLTSS plan enrollees age 18 years 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).

A systematic sample drawn from the eligible population, which includes enrollees:

- Who are 18 years and older as of the first day of the measurement year.

- Who are enrolled in a Medicaid MLTSS plan for at least 120 days between September 1 of the year prior to the measurement year and December 31 of the measurement year. This timeframe allows for assessment within 90 days of enrollment and development of a care plan within 30 days of assessment for new enrollees; and at least an annual MLTSS re-assessment for established enrollees.

- Who have either of the following benefits: 1) long-term services and supports: home- and community-based or 2) long-term services and supports: institution based.

Note: For individuals who have multiple distinct continuous enrollment periods during the measurement year, plans should look at the assessment completed in the last continuous enrollment period of 120 days or greater during the measurement year. This denominator is aligned with the denominator of a paired measure, LTSS Comprehensive Care Plan and Update, to allow MLTSS plans to use a single sample for assessing both measures.

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

Exclude enrollees in the denominator who were enrolled in the plan prior to September 1 of the year prior to the measurement year (i.e., established enrollees) who left the plan for more than 45 days between January 1 and December 31 of the measurement year.

Exclude enrollees who could not be reached for a comprehensive assessment or who refused a comprehensive assessment.

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.)*

Exclude enrollees in the denominator who were enrolled in the plan prior to September 1 of the year prior to the measurement year (i.e., established enrollees) who left the plan for more than 45 days between January 1 and December 31 of the measurement year. These are enrollees who may have left the plan before their annual assessment was conducted.

Exclude enrollees who could not be reached for a comprehensive assessment or who refused a comprehensive assessment. Enrollees who refuse an in-home assessment are excluded from the numerator requirement of in-home assessment but are not excluded from the other measure elements. Refusal of an in-home assessment must be documented in the record to qualify for this numerator exclusion.

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.)*

Not Applicable, no stratification.

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

No risk adjustment or risk stratification

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.)*

Step 1a. Determine the eligible population.

Step 1b. From the eligible population, draw a systematic sample.

Exclusion – Could Not Be Reached

Step 1c. From the systematic sample, identify all enrollees who were newly enrolled in the plan between September 1 of the year prior to the measurement year and August 31 of the measurement year. These are “new enrollees.”

Step 1d. Identify enrollees who could not be reached for a comprehensive assessment within 90 days of enrollment.

Step 1e. From the systematic sample, identify all enrollees who were enrolled prior to September 1 of the year prior to the measurement year. These are “established enrollees.”

Step 1f. Exclude enrollees who were enrolled in the plan prior to September 1 of the year prior to the measurement year and left the plan for more than 45 days between January 1 and December 31 of the measurement year.

Step 1g. Identify enrollees who could not be reached for a comprehensive assessment during the measurement year.

Step 1h. Add the number of enrollees from Steps 1d and 1g.

Step 1i. Divide the total number of enrollees from Step 1h by the number of enrollees from Step 1b to calculate the rate. This is the exclusion rate of enrollees who could not be reached for a comprehensive assessment.

Exclusion – Refused Comprehensive Assessment

Step 2a. From the systematic sample from Step 1b, identify all enrollees who were newly enrolled in the plan between September 1 of the year prior to the measurement year and August 31 of the measurement year. These are "new enrollees."

Step 2b. Identify enrollees who refused a comprehensive assessment within 90 days of enrollment.

Step 2c. From the systematic sample, identify all enrollees who were enrolled prior to September 1 1 of the year prior to the measurement year. These are "established enrollees."

Step 2d. Exclude enrollees who were enrolled in the plan prior to September 1 of the year prior to the measurement year and left the plan for more than 45 days between January 1 and December 31 of the measurement year.

Step 2e. Identify enrollees who refused a comprehensive assessment during the measurement year.

Step 2f. Add the number of enrollees from Steps 2b and 2e.

Step 2g. Divide the total number of enrollees from Step 2f by the number of enrollees from Step 1b to calculate the rate. This is the exclusion rate of enrollees who refused comprehensive assessment.

Numerator Rate 1

Step 3a. From the systematic sample from Step 1b, identify all enrollees who were newly enrolled in the plan between September 1 of the year prior to the measurement year and August 31 of the measurement year. These are "new enrollees."

Step 3b. Exclude enrollees who could not be reached (Step 1d + Step 1g) or refused a comprehensive assessment (Step 2b + Step 2e).

Step 3c. Identify enrollees who have documentation of a comprehensive assessment with 9 core elements within 90 days of enrollment.

Step 3d. From the systematic sample, identify all enrollees who were enrolled prior to September 1 of the year prior to the measurement year. These are "established enrollees."

Step 3e. Exclude enrollees who were enrolled in the plan prior to September 1 of the year prior to the measurement year and left the plan for more than 45 days between January 1 and December 31 of the measurement year.

Step 3f. Exclude enrollees who could not be reached (Step 1d + Step 1g) or refused a comprehensive assessment (Step 2b + Step 2e).

Step 3g. Identify enrollees who have documentation of a comprehensive assessment with 9 core elements during the measurement year.

Step 3h. Add the number of enrollees from Steps 3c and 3g.

Step 3i. Divide the total number of enrollees from Step 3h by the number of enrollees from Step 1b to calculate the rate. This is the rate of enrollees with a comprehensive assessment completed in the appropriate time frame with nine (9) core elements.

Numerator Rate 2

Step 4a. From enrollees identified in Step 3c (new enrollees with a completed assessment of the core elements), identify enrollees who have documentation of at least twelve (12) supplemental elements.

Step 4b. From enrollees identified in Step 3g (established enrollees with a completed assessment of the core elements), identify enrollees who have documentation of at least twelve (12) supplemental elements.

Step 4c. Add the number of enrollees from Steps 4a and 4b.

Step 4d. Divide the total number of enrollees from Step 4c by the number of enrollees from Step 1b to calculate the rate. This is the rate of enrollees with a comprehensive assessment completed in the appropriate time frame with nine (9) core and at least twelve (12) supplemental elements.

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.

MLTSS plans should identify a systematic sample of 411 enrollees who meet the eligible population criteria. The same sample may be used to calculate three paired measures:

- Comprehensive LTSS Assessment and Update

- Comprehensive LTSS Care Plan and Update

- Shared LTSS Care Plan with Primary Care Practitioner.

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.

Management Data, Other, Paper Medical Records

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.

Other = Case Management Records. Records are reviewed to determine if assessment elements were documented during the required timeframe.

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

If other: Long-term non-acute care, home- and community-based services, health plan case management.

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.)*

2. Validity – See attached Measure Testing Submission Form

LTSS_Comp_Assess_Testing_Attachment_Nov28.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.

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.

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.

No - This measure is not risk-adjusted

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

Measure Number (*if previously endorsed*)**:**
Measure Title: LTSS Comprehensive Assessment and Update
Date of Submission: 11/7/2017

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 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: abstracted from case management records	☒ other: abstracted from case management records

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.

1.3. What are the dates of the data used in testing? September 1, 2014 to December 31, 2015

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

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

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Five MLTSS plans representing both large national plans and small local plans from five states, located in geographically diverse regions of the country were included in the testing and analysis. All five participating health plans enrolled two or more of the major LTSS sub-populations, which include frail older adults age 65 and over, people under age 65 with physical disabilities, people with developmental or intellectual disabilities, and people with serious mental illness. All five participating plans covered Medicare benefits as well as Medicaid LTSS and had at least one of the following types of contracts: Medicare-Medicaid Plan (MMP, Dual Eligible Special Needs Plan (D-SNP), or Fully Integrated Dually Eligible Special Needs Plan (FIDE SNP).

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

Each of the five plans identified a random sample of 150 enrollees from an eligible population of MLTSS members age 18 and older as of January 1, 2015. Two groups of enrollees were included in the sample identified by each plan (i.e., "new" and "established" enrollees). "New" enrollees were members who were newly enrolled between September 1, 2014 and August 31, 2015, without any gaps in enrollment during the first 120 days during the measurement year of CY 2015. "Established" enrollees were members who were enrolled prior to September 1, 2014 and enrolled continuously with no more than one 45-day gap throughout the measurement year. To ensure that the final sample of 150 enrollees included adequate data from both subgroups, each health plan was asked to include at least 40 "New" enrollees in the sample. Forty-six enrollees were removed from the sample for inconsistent information, and the total sample included in the testing and analysis contained 715 enrollees. Table 1 summarizes the enrollees' characteristics for the sample. Table 2 summarizes their place of residence, type of plan, whether they were new or established enrollees, and presence of chronic conditions or activities of daily living (ADL) limitations.

Table 1. Analytic Sample Demographic Information

Characteristics	Percentage of enrollees in the testing sample (n=715)
Sex	
Female	68.5
Male	31.0
Missing	0.4
Age	
Under 18	0.7
18-40	6.9
41-64	33.3
65 and older	59.0
Missing	0.1
Race	
White	37.9
Black/African American	27.0
Asian	3.5
American Indian/Alaskan Native	0.3
Multi-race	0.1
Hawaiian/Pacific Islander	0.1
Unknown	19.2
Other	11.2
Ethnicity	
Non-Hispanic	55.9
Hispanic	17.3
Unknown	22.0
Primary language	
English	66.7
Spanish	10.4
Missing	17.1
Other	5.9

Source: Mathematica analysis of MLTSS enrollees from five health plans. Information from 46 enrollees was excluded due to data quality issues.

Table 2. Analytic Sample LTSS Information

Characteristics	Percentage of enrollees in the testing sample (n=715)
Place of residence	
Home or community residence	77.6
Nursing facility	14.3
Assisted living facility	1.1
Other institution	0.7
Missing	6.3
MLTSS program	
Integrated plan (MMP, D-SNP, FIDE-SNP)	68.3
Non-integrated	31.6
Missing	0.1
Type of enrollee	
New	48.6
Established	51.3
Missing	0.1
Chronic conditions present by end of measurement year	
Arthritis	42.8
Asthma	13.2
Cancer	8.0
Cardiac conditions (e.g., CAD, arrhythmia)	42.1
Dementia	17.5
Depression	34.3
Diabetes	35.4
Gastrointestinal (GI) disorders (e.g., IBD, cirrhosis)	26.0
Chronic Heart Failure	16.9
HIV	1.5
Neurological Disorders	20.7
Other Pulmonary Conditions (e.g., COPD)	24.2
Psychotic Disorder	11.9
Renal Disease	13.4
Stroke	16.1
ADL Limitations present by end of measurement year	
Walking	69.5
Toileting	57.2
Bathing	61.5
Eating	26.9
Transferring	61.0
Dressing	59.0

Source: Mathematica analysis of MLTSS enrollees from five health plans. Information from 46 enrollees was excluded due to data quality issues.

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.

No patient-level social risk factors were analyzed. All patients in the sample were Medicaid-eligible.

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*)

Data presented here reflects the final measure specifications. Additional details on the selection of the core and supplemental data elements can be found in the Appendix: Additional Testing Data.

Reliability of Data Elements

We calculated reliability of the critical data elements used in the measure with Cohen's kappa statistic to evaluate inter-rater reliability. Four participating health plans contributed thirty paired assessments and one plan contributed twenty-four paired assessments for analysis by two independent assessors. In total, the records of 144 MLTSS health plan enrollees were used for this analysis.

When comparing observations made by two individuals, Cohen's kappa statistic, or $\hat{\kappa}$ (ranging from 0 to 1) measures the percentage of agreement between individuals. If the observed agreement is greater than or equal to chance agreement, then $\hat{\kappa} \geq 0$, with $\hat{\kappa} = 1$ signifying perfect agreement. If the observed agreement is less than or equal to chance agreement, then $\hat{\kappa} \leq 0$ (Fleiss, Levin and Paik, 2003). We calculated the $\hat{\kappa}$ statistic reflecting the amount of agreement among key data elements as:

$$\hat{\mathrm{K}} = \frac{\rho_{\mathrm{a}} - \rho_{\mathrm{e}}}{1 - \rho_{\mathrm{e}}}$$

Where ρe is the expected percent chance agreement and ρa is the observed agreement.

Reliability of Measure Rates

To examine the reliability of measure performance rates, we evaluated split-sample reliability using the Inter-class correlation coefficient (ICC) to determine agreement. Enrollees were randomly assigned to each of the two groups, and rates were calculated separately for each group and then compared at the health plan-level.

When evaluating the reliability of rates or proportions, the inter-class correlation coefficient, $\hat{\rho}$, summarizes the estimated agreement among observations. The inter-class correlation coefficient (ICC) is the ratio of the subject variance to the total variance (the subject variance plus the error variance). Higher values of $\hat{\rho}$ indicate that the subject variance exceeds the error variance by a wide margin (Gwet, 2014). We calculated the $\hat{\rho}$ statistic reflecting the amount of agreement among measure results as:

$$\hat{\rho} = \frac{\sigma_s^2}{\sigma_s^2 - \sigma_e^2}$$

where $\sigma s2$ is the subject variance, and $\sigma e2$ is the error variance.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e*.g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Reliability of Data Elements

Nineteen of the thirty-three potential data elements met the threshold for moderate or higher reliability ($\hat{\kappa} \geq 0.4$), as shown in Table 3. Data elements with the highest reliability were documented dates of assessment ($\hat{\kappa}$= 0.7780), assessment of smoking, alcohol and other drug use, and behavioral difficulties ($\hat{\kappa}$= 0.6629, 0.6531, and 0.6031). Data elements with the lowest reliability were assessments of cognitive function, social isolation, and living arrangements ($\hat{\kappa}$= 0.1814, 0.1295, and 0.1467). Other data elements with low reliability were the availability of friend or family caregiver support, public and plan benefits, and preference for participation in care planning ($\hat{\kappa}$= 0.1706, 0.1503, and 0.1987).

Five of the nine data elements designated as "core" elements for the measure met the threshold for moderate reliability ($\hat{\kappa} \geq 0.4$). Core data elements with the highest reliability were mental health status, home safety risks, activities of daily living and overall health status ($\hat{\kappa}$=0.5659, 0.5257, 0.4277 and 0.4391). Core data elements with the lowest reliability were current providers, cognitive function, availability of friend or family caregiver support, and living arrangements ($\hat{\kappa}$=0.3201, 0.1814, 0.1706 and 0.1467).

A total of ten of the eighteen data elements designated as "supplemental" elements for the measure met the threshold for substantial ($\hat{\kappa}$=0.6), or moderate reliability ($\hat{\kappa} \geq 0.4$). Supplemental data elements with substantial reliability included smoking status, alcohol and other drug use, and behavior difficulties ($\hat{\kappa}$=0.6629, 0.6531, and 0.6031). Supplemental data elements with moderate reliability included hearing needs, instrumental activities in daily living, vision needs, speech needs, cultural and linguistic preferences, physical/occupational therapy needs, and recent use of services ($\hat{\kappa}$=0.5851, 0.5832, 0.5150, 0.5141, 0.4510, 0.4430, and 0.4019). Supplemental data elements with the lowest reliability were physical disability accommodations, patient activation or self-efficacy, preference for routine activities, and availability of social support in community ($\hat{\kappa}$=0.2495, 0.2783, 0.2395, 0.3668).

We did not include one element in the supplemental or core set, Preference for Participating in Care Planning, due to the lack of clarity on the item's definition and its absence on any of the assessments reviewed by the team.

Table 3. Reliability of Key Data Elements

Data element	Kappa statistic	Interpretation
Comprehensive Assessment Completed	0.6868	Substantial
Setting of Assessment (Face-to-Face, Phone, Other)	0.6868	Substantial
Assessment Date	0.7780	Substantial
Elements Documented:		
ADLs*	0.4277	Moderate
IADLs	0.5832	Moderate
Physical Disability Accommodations	0.2495	Fair
Overall Health Status*	0.4391	Moderate
Cognitive Function*	0.1814	Slight
Behavioral Difficulties	0.6031	Substantial
Mental Health Status*	0.5659	Moderate
Patient Activation/Self-Efficacy	0.2783	Fair
Vision Needs	0.5150	Moderate
Hearing Needs	0.5851	Moderate
Speech Needs	0.5141	Moderate
PT/OT Needs	0.4430	Moderate
Home Safety Risks*	0.5257	Moderate
Smoking	0.6629	Substantial
Alcohol and other drug use	0.6531	Substantial
Availability of social support in community	0.3668	Fair
Availability of friend or family caregiver support*	0.1706	Slight
Availability of public and plan benefits	0.1503	Slight
Assessment of Social Isolation	0.1295	Slight
Living Arrangements*	0.1467	Slight
Preference for Routine Activities	0.2395	Fair
Preferences for Advance Care Planning	0.5044	Moderate
Preference for Participating in Care Planning**	0.1987	Slight
Cultural and Linguistic Needs	0.4510	Moderate
Current Providers*	0.3201	Fair
Recent use of Services	0.4019	Moderate
Identification of Family/Friend Caregiver	0.4892	Moderate
Contact information for at least one Family/Friend Caregiver	0.5141	Moderate

* Core data elements

** Removed data elements

Reliability of Measure Rates

ICCs for Rate 1 and Rate 2 exceed 0.9, indicating almost perfect agreement between the samples, and showing a significant association at p< 0.5 or less (Table 4). Reliability of the exclusion rates was not available as plans indicated that none of the enrollees who did not receive a comprehensive assessment refused an assessment and plans did not record any additional reason why an assessment was not completed.

Table 4. Reliability of Recommended Measure Rates

Measure	ICC statistic	Interpretation
Rate 1: Core Elements Documented Rate	0.9499*	Almost Perfect
Rate 2: Core Elements + 12 or More Supplemental Documented Elements	0.9166*	Almost Perfect

Source: Mathematica analysis of data from 715 MLTSS enrollees, representing five health plans.

Notes: Interpretation of ICC used the following standards: < 0 – Less than chance agreement; 0 – 0.2 Slight agreement; 0.21 – 0.39 Fair agreement; 0.4 – 0.59 Moderate agreement; 0.6 – 0.79 Substantial agreement; 0.8 – 0.99 Almost Perfect agreement; 1 Perfect agreement. (Landis and Koch, 1977)

*Significantly associated at the p<0.05 level.

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?*)

There was a mix in the inter-rater reliability of data elements. While many of the core data elements had high reliability, some core data elements had low reliability, specifically cognitive function, living arrangements, and the availability of friend or family caregiver support. To address this limitation, we revised the measure specifications to include greater specifics in the definition of these elements and reduce inter-rater variation in interpretation. For cognitive function, we limited the item to only assessment of cognitive function using a validated tool and provided examples. For living arrangement we clarified that the documentation must identify the individual as living in the community, nursing facility or other institution, assisted living facility or other setting. For availability of family and friend caregiver support we clarified what level of documentation was necessary. Some elements with extremely low reliability were dropped (i.e., preference for participating in care planning), others were renamed (i.e., preference for routine activities was renamed preference for participating in volunteer or paid work activities), and others were revised to provide additional examples (i.e., availability of public and plan benefits).

The Interclass Correlation Coefficient for both Rate 1 and 2 were high indicating the subject variance exceeds the error variance by a wide margin indicating good measure score reliability.

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)*

Empiric Validity of Performance Measure Score

Empiric validity was assessed by analyzing the convergent validity (whether the measure results correlate with other theoretically related quality measure results) of this assessment measure. Comparison with established health plan quality measures was not possible due to the lack of publicly available standardized measures for the MLTSS population. Therefore, we analyzed correlation between this assessment measure and four measures being tested for the MLTSS population in this project using the Spearman Rank Correlations:

- LTSS Comprehensive Care Plan and Update Measure (MLTSS-2)
- LTSS Shared Care Plan with Primary Care Practitioner (MLTSS-3)
- LTSS Re-Assessment/Care Plan Update after Inpatient Discharge (MLTSS-4)
- Falls: Screening, Risk-Assessment and Plan of Care to Prevent Future Falls (MLTSS-5)

Systematic Assessment of Face Validity

By surveying several panels of stakeholders, we were able to assess face validity, which indicates whether the measure accurately represents the concept being measured and achieves the purpose for which it is intended (i.e., to measure the percent of beneficiaries being assessed and the quality of the assessments conducted).

In November 2017, 13 members of the 2017 Technical Expert Panel (see Ad.1 for member list) voted on the face validity of the measure via a web survey. The TEP voted on whether they "strongly agree", "agree", "disagree", or "strongly disagree" with the following survey items:

1. Denominator is appropriate given the intent of the measure
2. Numerator Rate 1 is appropriate given the intent of the measure
3. Numerator Rate 2 is appropriate given the intent of the measure
4. Exclusion 1 is appropriate given the intent of the measure
5. Exclusion 2 is appropriate given the intent of the measure
6. Would high performance on this measure indicate that a health plan is providing higher quality care?
7. In the future, do you think that performance scores on this measure will distinguish between good and poor performance?

TEP members had the opportunity to provide written comments on all the survey items above, as well as an additional open response question:

8. Do you have any recommendations that would help strengthen the Comprehensive Assessment and Update Measure?

See Appendix 1 for summaries of the written comments for each survey item. While the majority of TEP members agreed with the measure, the comments included in the summaries are largely from those who disagreed or suggested revisions to the measure.

Additional Face Validity Feedback

Under a previous CMS contract (Contract No. HHSM-500-2010-00026I/HHSM-500-T0011), a multi-stakeholder technical expert panel (TEP) of 20 individuals representing home and community based services organizations, disability service organizations, organizations providing services for aging populations, state Medicaid and quality organizations, and consumer advocacy groups was convened to provide input on the MLTSS measure development and testing processes (see Ad 1 for TEP member list – 2013 TEP). Under the current contract, we convened a new TEP with a similar size and composition to inform the continued development and testing of this measure (see Ad 1 for TEP member list – 2017 TEP). Six members of the 2017 TEP also participated in a workgroup that provided in-depth feedback on the MLTSS measures specifically (2017 MLTSS Workgroup).

We also received feedback from a three-week public comment period hosted on CMS's online public comment system. The public comment period was open and broadcast to all interested parties. Overall, commenters offered general support for the measure. Some commenters noted concern that the measure focuses on completing the assessments on time, rather than the quality of the assessments. Several comments stressed the importance of harmonizing the measures with existing measures and tools. Numerous comments offered ways that the measure specifications could be

furthered clarified, including denominator and numerator criteria, timing, construction of required elements, setting of the assessment and exclusion criteria. A few comments asked us to explore the importance of the measure and the burden for health plans to report during testing.

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

Empiric Validity of Performance Measure Score

Among the recommended measure rates, we observed very few positive, significant relationships. The Spearman Rank Correlation coefficient, $\hat{\rho}$, showed a significant, strong positive relationship between two rates in this measure (Rate 1: Core Elements vs. Rate 2: Core Elements and at least 12 Supplemental elements) and the two rates in a paired measure *Comprehensive LTSS Care Plan* (Rate 1: Core Elements versus Rate 2: Core Elements and at least 4 Supplemental elements), as shown in Table 5. The remaining relationships ranged from slight to moderate relationships, some positive and some negative, but none were significant.

Table 5. Correlation of recommended measure rates

Measures	MLTSS-1, Rate 1	MLTSS-1, Rate 2	MLTSS-2, Rate1	MLTSS-2, Rate 2	MLTSS-3	MLTSS-4, Rate 1	MLTSS-4, Rate 2	MLTSS-5 Screening	MLTSS-5 Assess	MLTSS-5 Care Plan
MLTSS-1, Rate 1: Core Elements	--	1.000**	-0.8603	-0.8603	-0.0574	-0.2368	-0.0513	-0.0513	0.3341	NA
MLTSS-1, Rate 2: Core Elements + 12+ Supplemental Elements	1.000**	--	-0.8603	-0.8603	-0.0574	-0.2368	-0.0513	-0.0513	0.3341	NA

Source: Mathematica analysis of data from 715 MLTSS enrollees, representing five health plans.

*Significant association, at $p < 0.05$

**Significant association, at $p < 0.01$

MLTSS-1 = LTSS Comprehensive Assessment and Update

Rate 1 = Assessment with 9 core elements documented

Rate 2 = Assessment with at least 12 supplemental elements documented

MLTSS-2 = LTSS Comprehensive Care Plan and Update

Rate 1 = Care plan with 7 core elements documented

Rate 2 = Care plan with at least 4 supplemental elements documented

MLTSS-3 = LTSS Shared Care Plan with Primary Care Practitioner

Rate = Care plan transmitted to provider within 30 days of the plan's development or update

MLTSS-4 = LTSS Re-Assessment/Care Plan Update after Inpatient Discharge

Rate 1 = Re-assessment and care plan update, no face to face requirement

Rate 2 = Re-assessment only, no face to face requirement

MLTSS-5 = Falls: Screening, Risk-Assessment, and Plan of Care to Prevent Future Falls

Rate 1 = Screened for falls risk

Rate 2 = Falls risk assessment

Rate 3 = Falls risk plan of care

Systematic Assessment of Face Validity

Table 6 contains voting results from the survey. Overall, most TEP members supported the denominator, numerators, and exclusions for Comprehensive Assessment and Update measure. The majority of TEP members also agreed that the measure is reflective of quality and had the potential to distinguish performance.

Table 6. TEP Face Validity Survey Results

Response	Denominator is appropriate given the intent of the measure	Numerator Rate 1 is appropriate given the intent of the measure	Numerator Rate 2 is appropriate given the intent of the measure	Exclusion 1 is appropriate given the intent of the measure (enrollees who could not be reached)	Exclusion 2 is appropriate given the intent of the measure (enrollees who refused)	Would high performance on this measure indicate that a health plan is providing higher quality care?	In the future, do you think that performance scores on this measure will distinguish between good and poor performance?
Strongly Agree	4	2	1	2	2	1	1
Agree	9	9	11	6	10	7	8
Disagree	0	2	1	2	0	2	2
Strongly Disagree	0	0	0	3	1	3	2
No response	0	0	0	0	0	0	0
Total % Agree	100%	85%	92%	62%	92%	62%	69%

Additional Face Validity Feedback

The TEP noted that as documentation of the core elements improves among plans over time, the other rates are in turn likely to become more meaningful and useful. Specifically, Rate 2, which reports the percentage of enrollees with all nine core elements and at least 12 supplemental elements, appears the most useful as an "aspirational" measure. Health plan performance is slightly lower for Rate 2 relative to Rate 1 (focused on just core elements), but still yields non-zero rates for three of the five health plans. Participants in the TEP, the workgroup and public commenters all supported moving forward with this measure.

Additionally, feedback from the public comment period was generally supportive of the measure. Some commenters noted concern that the measure focuses on completing the assessments on time, rather than the quality of the assessments. Several comments stressed the importance of harmonizing the measures with existing measures and tools. Numerous comments offered ways that the measure specifications could be furthered clarified, including denominator and numerator criteria, timing, construction of required elements, setting of the assessment and exclusion criteria. A few comments asked us to explore the importance of the measure and the burden for health plans to report during testing.

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

Because all of the MLTSS measure under development are expected to reflect the quality of care provided to MLTSS enrollees, ideally we would expect to see significant, strong positive relationships correlations. However, given the overall sub-optimal and incongruent results among health plans, these results are not surprising. For example, the Core Element rates reported in the Long Term Services and Supports Comprehensive Assessment and Update measure and the Core Elements reported in the Long Term Services and Supports Comprehensive Care Plan and Update measure have a substantial, negative relationship. This relationship reflects the fact that for one measure two health plans have zero rates, while for the other measure, the other three health plans have zero rates. As reporting improves the internal validity of the measures should also improve accordingly.

Systematic Assessment of Face Validity

The voting results suggest that this is a valid measure. TEP members who did not support the measure cited that this measure is a process measure, and it would be better to have an outcome measure. The measurement team agrees, however before outcome measures can be collected, organizations must be assessing and documenting the needs of beneficiaries using a standard set of guidelines. One cannot measure an outcome that is not being assessed in the first place. This measure is a first step and is not meant to be used alone to determine quality; rather, it is one piece of critical care delivery in determining whether individuals' needs and preferences are being assessed and addressed - including non-medical needs.

Additional Face Validity Feedback

Stakeholder input suggest that this measure is valid for assessing the need for long term services and supports for plan enrollees. The findings from the TEPs, workgroup, and public comment suggest that modifications made to the measure specification produced valid results.

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*)

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*)

N/A

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*)

N/A

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 _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.

N/A

2b3.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

N/A

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

N/A

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?

N/A

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.**

N/A

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*)

N/A

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

If stratified, skip to 2b3.9

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

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

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

2b3.9. Results of Risk Stratification Analysis:

N/A

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*)

N/A

2b3.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

N/A

2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b4.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

To evaluate whether the measures demonstrated statistically significant variation in performance across health plans and/or showed sub-optimal performance (suggesting room for improvement), we calculated the sample mean, minimum and maximum values, and standard deviations for the five health plans-level results. We utilized two-tailed T-tests to evaluate whether each health plan's results differed significantly from the sample mean.

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*)

MLTSS plan performance is presented in Table 6. Health plan 03 demonstrated rates that differed significantly from the mean at the .05 level.

Table 7. Long Term Services and Supports Comprehensive Assessment and Update (MLTSS-1) Performance Rates by Health Plans with Significant Difference Noted

Health Plan Identifier	Rate 1: 9 Core elements	Rate 2: 9 Core Elements + 12 supplemental elements
HP 01	0.0	0.0
HP 02	0.0	0.0
HP 03	25.5*	21.6*
HP 04	8.5	6.3
HP 05	5.7	4.3
Minimum	0.0	0.0
Mean	7.9	6.4
Maximum	25.5	21.6
Standard deviation	10.5	8.9

Source: Mathematica analysis of enrollee data from 5 MLTSS plans.

Note: Health plan identifiers "HP 01" through "HP 05" are used to protect the confidentiality of health plans participating in beta testing.

NA = Not applicable (no enrollees had all the 9 core elements documented)

*Significantly different from the mean at the .05 level, two-tailed test

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?*)

Although we observed very low (in some cases zero) performance rates, we do see that certain plans consistently performed better than others. Additionally, health plan 03 had rates that demonstrated a statistically significant

difference from the mean. These findings indicate that the measure as specified does have the potential to identify meaningful differences in performance between health plans.

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

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

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

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

N/A

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

N/A

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

N/A

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

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

The primary focus of this measure is an assessment of missing data – i.e., whether MLTSS enrollees show documentation of information relating to completion of a comprehensive assessment. When required elements are missing, the enrollee is considered not to have met the numerator requirements. Therefore, the measure results themselves reflect the extent and distribution of missing data overall and by plan. In addition, the extent of missing data for the core and supplemental elements is described in further detail in the Additional Testing Results Appendix.

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*)

Please see details in the Additional Testing Results Appendix.

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*)

The low rates for this measure reflect the lack of standardization in the data elements that should always be documented in a comprehensive assessment for MLTSS enrollees. This measure assesses the percentage of Medicaid MLTSS enrollees who have an assessment conducted with a specified mode (face-to-face, in the home), in a specified timeframe, and addressing specific core and supplemental elements, and in doing so, should help address this lack of standardization.

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.

Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

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**.

Some data elements are in defined fields in electronic sources

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).

This measure is primarily collected from health plan and case management records, many of which are electronic. However, the data elements needed for this measure are not currently standardized (i.e., data elements may be in free text instead of structured fields). As the LTSS field moves forward, we anticipate the data elements needed to calculate this measure will become structured allowing for an eMeasure in the future.

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)*.

Not applicable, no fees or licensing are currently required.

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)
Public Reporting 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?*)

Not applicable; 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.

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 Comprehensive Assessment and Update measure is included in the set of recommended measures that assesses person-centered planning and coordination.

http://mltss.org/wp-content/uploads/2017/05/MLTSS-Association-Quality-Framework-Domains-and-Measures-042117.pdf

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.

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.

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 enrollees who are receiving long-term services and supports. Comprehensive assessments serve as the foundation for developing comprehensive person-centered care plans, providing all needed services and supports of high quality, and performing well-coordinated care. However, the tools currently used by states and health plans to conduct assessments, and the performance measures used to evaluate the quality of assessments conducted vary widely. This measure would address the lack of standardization by assessing the percentage of Medicaid LTSS enrollees who have a comprehensive assessment conducted that includes specific core and supplemental elements. A standardized measure of comprehensive assessment will allow for apples-to-apples comparisons of LTSS plans across states.

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.

Not applicable. 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.

Not applicable. 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.

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.)

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.

Attachment Attachment: LTSS_Comp_Assess_Additional_Testing_Results_Nov28.docx

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: Henry, Ireys, hireys@mathematica-mpr.com, 202-554-7536-

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.

2017 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 for People with 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 for Health

Jason Rachel, PhD, Virginia Department of Medical Assistance Services

Balu Gadhe, MD, CareMore

Patricia Kirkpatrick, MJ, RN, CPHQ, Amerigroup Corporation

Cheryl Phillips, MD, LeadingAge

Diane McComb, MSEd, American Network of Community Options and Resources

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

Quality Measure Development (QMD) – Medicare-Medicaid (Dual) and Medicaid-Only Enrollees MLTSS Workgroup, January 2017

Laura Brannigan, GuildNet

Jennifer Clark, Centene Corporation

Camille Dobson, NASUAD

Patricia Kirkpatrick, Amerigroup

Michael Monson, Centene Corporation

Lauren Murray, National Partnership for Women and Families

Pamela Parker, Independent Consultant-Integrated Care

Carol Raphael, Manatt Health Solutions

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 2017 TEP and Workgroup 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? Not applicable

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

Ad.6 Copyright statement:

Ad.7 Disclaimers:

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