Patient Experience & Function Measure Worksheets

1. 1741: Consumer Assessment of Healthcare Providers and Systems (CAHPS)® Surgical Care Survey Version 2.0

2. 3319: Long Term Services and Supports (LTSS) Comprehensive Assessment and Update

3. 3324: Long Term Services and Supports (LTSS) Comprehensive Care Plan and Update

4. 3325: Long Term Services and Supports (LTSS) Shared Care Plan with Primary Care Practitioner

5. 3326: Long Term Services and Supports (LTSS) Re-Assessment/Care Plan Update after Inpatient Discharge

NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 1741
Measure Title: Consumer Assessment of Healthcare Providers and Systems (CAHPS)® Surgical Care Survey Version 2.0
Measure Steward: American College of Surgeons, Division of Advocacy and Health Policy
Brief Description of Measure: The following 6 composites and 1 single-item measure are generated from the Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Surgical Care Survey. Each measure is used to assess a particular domain of surgical care quality from the patient's perspective.

Measure 1: Information to help you prepare for surgery (2 items)
Measure 2: How well surgeon communicates with patients before surgery (4 items)
Measure 3: Surgeon's attentiveness on day of surgery (2 items)
Measure 4: Information to help you recover from surgery (4 items)
Measure 5: How well surgeon communicates with patients after surgery (4 items)
Measure 6: Helpful, courteous, and respectful staff at surgeon's office (2 items)
Measure 7: Rating of surgeon (1 item)

The Consumer Assessment of Healthcare Providers and Systems (CAHPS) Surgical Care Survey (S-CAHPS) is a standardized survey instrument that asks patients about their experience before, during and after surgery received from providers and their staff in both inpatient and outpatient (or ambulatory) settings. S-CAHPS is administered to adult patients (age 18 and over) that had an operation as defined by CPT codes (90 day globals) within 3 to 6 months prior to the start of the survey.

The S-CAHPS expands on the CAHPS Clinician & Group Survey (CG-CAHPS), which focuses on primary and specialty medical care, by incorporating domains that are relevant to surgical care, such as sufficient communication to obtain informed consent, anesthesia care, and post-operative follow-up and care coordination. Other questions ask patients to report on their experiences with office staff during visits and to rate the surgeon.

The S-CAHPS survey is sponsored by the American College of Surgeons (ACS). The survey was approved as a CAHPS product in early 2010 and the Agency for Healthcare Research and Quality (AHRQ) released version 1.0 of the survey in the spring of 2010. The S-CAHPS survey Version 2.0 was subsequently endorsed by NQF in June 2012 (NQF #1741). The survey is part of the CAHPS family of patient experience surveys and is available in the public domain at https://cahps.ahrq.gov/surveys-guidance/cg/about/index.html. Surgeons may customize the S-CAHPS survey by adding survey items that are specific to their patients and practice. However, the core survey must be used in its entirety in order to be comparable with other S-CAHPS data. The S-CAHPS survey is available in English and Spanish.

The 6 composite measures are made up of the following items:

The 1 single item measure (Measure 7) is (Q35): Using any number from 0 to 10, where 0 is the worst surgeon possible and 10 is the best surgeon possible, what number would you use to rate all your care from this surgeon?

Measure 1: Information to help you prepare for surgery (2 items)
Q3. Before your surgery, did anyone in this surgeon´s office give you all the information you needed about your surgery?
Q4. Before your surgery, did anyone in this surgeon's office give you easy to understand instructions about getting ready for your surgery?

Measure 2: How well surgeon communicates with patients before surgery (4 items)
Q9. During your office visits before your surgery, did this surgeon listen carefully to you?
Q10. During your office visits before your surgery, did this surgeon spend enough time with you?
Q11. During your office visits before your surgery, did this surgeon encourage you to ask questions?
Q12. During your office visits before your surgery, did this surgeon show respect for what you had to say?

Measure 3: Surgeon's attentiveness on day of surgery (2 items)
Q15. After you arrived at the hospital or surgical facility, did this surgeon visit you before your surgery?
Q17. Before you left the hospital or surgical facility, did this surgeon discuss the outcome of your surgery with you?

Measure 4: Information to help you recover from surgery (4 items)
Q26. Did anyone in this surgeon's office explain what to expect during your recovery period?
Q27. Did anyone in this surgeon's office warn you about any signs or symptoms that would need immediate medical attention during your recovery period?
Q28. Did anyone in this surgeon's office give you easy to understand instructions about what to do during your recovery period?
Q29. Did this surgeon make sure you were physically comfortable or had enough pain relief after you left the hospital or surgical facility where you had your surgery?

Measure 5: How well surgeon communicates with patients after surgery (4 items)
Q31. After your surgery, did this surgeon listen carefully to you?
Q32. After your surgery, did this surgeon spend enough time with you?
Q33. After your surgery, did this surgeon encourage you to ask questions?
Q34. After your surgery, did this surgeon show respect for what you had to say?

Measure 6: Helpful, courteous, and respectful staff at surgeon's office (2 items)
Q36. During these visits, were clerks and receptionists at this surgeon's office as helpful as you thought they should be?
Q37. During these visits, did clerks and receptionists at this surgeon's office treat you with courtesy and respect?

Developer Rationale: All of the measures submitted to NQF for endorsement share the main objective of the S-CAHPS survey, which is to measure aspects of surgical quality that are important to consumers and for which consumers are the best source of information. At the same time, consistent with the survey, the measures aim to directly benefit a variety of users, including: patients, practice groups, health plans, insurers, and specialty boards. Each group has a need for information regarding the quality of surgeons and surgical care.

The S-CAHPS survey measures patients' perceptions of their experiences with pre-, during and post-surgical care. The survey addresses issues relevant to surgical care, such as informed consent, communication, clear and helpful information, anesthesia care, office staff helpfulness, and post-operative follow-up.

To offer surgical patients and surgeons valid and reliable information on patient experience of care, the American College of Surgeons (ACS), in partnership with other surgical and anesthesia organizations, sponsored the development of the S-CAHPS survey. The S-CAHPS survey is a patient experience-of-care survey measure specifically tailored for surgical patients. The S-CAHPS survey was developed by working with patients to report on the full experience of surgical care, including their experience with the surgeon, the anesthesiologist, and the facility. Qualitative research (i.e., focus groups, cognitive testing, literature review) prior to field testing and focus groups following the main field test contributed to the development of the final composite measures. The data gathered through S-CAHPS survey data can assist consumers in identifying a high-quality surgeon and help surgeons to better understand and ultimately improve patient care.

To capture and ascertain the major domains surrounding the consumers' view of quality surgical care, the American Institutes for Research performed a literature review prior to developing the surgical CAHPS instrument. Using four literature databases: Medline, PsychInfo, CINAHL, and Evidence-Based Medicine Review, AIR reviewed 930 abstracts. These abstracts were narrowed from certain search terms and limitations. AIR asked the Surgical Quality Alliance (SQA) Technical Advisory Panel (TAP) members to provide input and guidance in these terms and limitations. The TAP included 21 members from various surgery and anesthesiology specialty societies, each having technical expertise in the topic of surgical care. The AIR research analyst then classified each abstract as relevant, possibly relevant, and not relevant.

From that point on, only the research analyst's "relevant" and "possibly relevant" classifications were reviewed by the project director. From this abstract review method, 37 abstracts were chosen to review in full. In addition, one article was added from an SQA Technical Advisory Panel (TAP) member. The 38 articles reviewed indicated 14 domains, or primary issues, of surgical patient care experience from a patient's perspective: information/education, interpersonal manner, pain, emotional support, accessibility/convenience, technical quality of care, efficacy/outcomes of care, availability, environment, customization/personalized care, patient involvement in care, continuity of care, overall satisfaction, and finances. These results guided the development of the surgical survey based on relevance to quality and to consumers and the ability of consumers to act as reliable reporters of the domain.

In addition, American Institutes for Research conducted six focus groups to identify important quality issues inherent in patients' experiences of surgical care. The results of the focus groups are included in Attachment D Main 1c5 Surgical CAHPS Focus Groups_2nd Round_2010.pdf. In total, 49 people participated in the focus groups, all of whom were 18 years of age or older and had undergone a surgical procedure billable with a 90-day global fee within the last 7 months. The groups were diverse both demographically and in the type of surgery. For recruiting the additional two groups of patients with more complex surgeries, each participant had their surgery in a hospital and stayed overnight for at least one night.

After the focus groups were conducted, the AIR project director, senior research analyst, and research analyst examined the notes and focused in on each of the above issues. They gathered recurring themes and issues into a report of results, used to develop insights about what patients feel characterizes quality surgical care. The three main domains were surgeon's interpersonal skills and behaviors, surgeon's expertise/technical competence, and surgeon's skill in communicating or providing health information and patient education. The results of the focus groups and the surgical patients' experiences of care ultimately guided development of the surgical survey. Though patients rated technical skill of the surgeon as highly relevant to quality, it was determined that patients are not the best reporters of technical surgical expertise, so this was not included as a domain in the survey instrument. Additionally, based on the psychometric analysis and discussions within the Surgical Quality Alliance (SQA) Technical Advisory Panel (TAP), ACS did not recommend the Anesthesia Care items as a reporting composite. While ACS believes these items should remain in the survey because anesthesia

care is an important factor in the surgical patient's experience of care, the Anesthesiologist works within the hospital system, and thus, the surgeon cannot adequately control the experiences between patient and anesthesiologist, and should not be measured on such.

Numerator Statement: We recommend that S-CAHPS Survey items and composites be calculated using a top-box scoring method. The top box score refers to the percentage of patients whose responses indicated excellent performance for a given measure. This approach is a kind of categorical scoring because the emphasis is on the score for a specific category of responses.

The top box numerator for the Overall Rating of Surgeon is the number of respondents who answered 9 or 10 for the item, with 10 indicating "Best provider possible".

For more information on the calculation of reporting measures, see What´s Available for the CAHPS Surgical Care Survey: https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/surgical/about/whats-available-surgical-care-survey.pdf

Also see Patient Experience Measures from the CAHPS Surgical Care Survey Document 409 obtained by going to: https://www.ahrq.gov/cahps/surveys-guidance/surgical/instructions/get-surg-care-survey-instruct.html

Also, for more information on the calculation of reporting measures, see How to Report Results of the CAHPS Clinician & Group Survey, available at https://cahps.ahrq.gov/surveys-guidance/cg/cgkit/HowtoReportResultsofCGCAHPS080610FINAL.pdf.

Denominator Statement: The measure's denominator is the number of survey respondents. The target population for the survey is adult patients (age 18 and over) who had a major surgery as defined by Common Procedural Terminology (CPT) codes (90 day globals) within 3 to 6 months prior to the start of the survey.

Results will typically be compiled over a 12-month period.

For more information on the calculation of reporting measures, see Patient Experience Measures from the CAHPS Surgical Care Survey, available at https://www.ahrq.gov/cahps/surveys-guidance/surgical/instructions/get-surg-care-survey-instruct.html.

Denominator Exclusions: The following are excluded when constructing the sampling frame:
- Surgical patients whose procedure was greater than 6 months or less than 3 months prior to the start of the survey.
- Surgical patients younger than 18 years old.
- Surgical patients who are institutionalized (put in the care of a specialized institution) or deceased.

Measure Type: Outcome: PRO-PM
Data Source: Instrument-Based Data
Level of Analysis: Clinician : Group/Practice

IF Endorsement Maintenance – Original Endorsement Date: May 1, 2012 **Most Recent Endorsement Date:** May 1, 2012

Maintenance of Endorsement -- Preliminary Analysis

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

Criteria 1: Importance to Measure and Report
1a. Evidence **Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.**
1a. Evidence. The evidence requirements for a health outcome measure include providing empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service; if these data not available, data demonstrating wide variation in performance, assuming the data are from a robust number of providers and results are not subject to systematic bias. For measures derived from patient report, evidence also should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful. The submission contains information for 7 patient-reported outcome based performance measures (PRO-PMs) that are calculated from data aggregates from responses to the Surgical CAHPS survey. The 7 PRO-PMs include: 1. Measure 1: Information to help you prepare for surgery (2 items) 2. Measure 2: How well surgeon communicates with patients before surgery (4 items) 3. Measure 3: Surgeon's attentiveness on day of surgery (2 items) 4. Measure 4: Information to help you recover from surgery (4 items) 5. Measure 5: How well surgeon communicates with patients after surgery (4 items) 6. Measure 6: Helpful, courteous, and respectful staff at surgeon's office (2 items) 7. Measure 7: Rating of surgeon (1 item) **Summary of evidence from 2012 evaluation:** • Consumer Assessment of Healthcare Providers and Systems (CAHPS)® Surgical Care Survey Version 2.0 indicates performance on the CAHPS Surgical Care Survey, which measures key components of patient experience, such as provider communication, that are consistent with patient-centered care. • This is an patient-reported outcome performance measure. The developer provided a conceptual logic model describing the relationship between the Surgical Consumer Assessment of Healthcare Providers and Systems (S-CAHPS) measure and clinical outcomes. • The developer provides a literature review of 18 studies and articles related to the measure including: a positive association between patient experience and clinical outcomes; a positive effect on key patient survey responses as a result of improved provider/staff communication and increasing patient partnership; and a positive effect of improved quality of medical consultations and patient education on the surgical patients' length of stay, anxiety levels, recovery time and compliance with treatment regimens. **Changes to evidence from last review** ☐ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.** ☒ **The developer provided updated evidence for this measure:** **Updates:** • The developer provides updated evidence of three systematic reviews that have examined the relationship between patient experience, clinical process and patient outcomes. o Patient experience is favorably associated with adherence to recommended medications and treatments, preventative care such as screenings and immunizations, patient-reported health

outcomes, clinical outcomes, reduced healthcare utilization, and reduced adverse events. (Doyle et al., 2013)
    - Better patient care experiences are associated with higher levels of adherence to recommended prevention and treatment processes, better clinical outcomes, and less health care utilization. (Anhang Price, 2014)
    - Beattie et al. critiqued the utility of published patient-reported experience measures (PREMs) aiming to measure the adult inpatient experience of hospital quality of care based on the PREMs' validity, reliability, cost efficiency, acceptability and educational impact (Beattie et al., 2015). They identified eleven international PREMs and concluded that patient experience data could be used to drive improvements in hospital care at national, local, and healthcare team levels.

- The developer provided six additional sources that were not included in the previous measure submission.
    - Hospitals in the highest quartile of performance on patent satisfaction had length of stay that was on average 0.6 days shorter than those with the lowest patient satisfaction. (Tsai, Orav, Jah, 2015)
    - There is a significant association between patient satisfaction and bot 30-day readmissions and the occurrence of postoperative surgical complications. (Lobo Prabhu, 2017)

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*
- *If derived from patient report, does the target population value the measured outcome and finds it meaningful?*

Guidance from the Evidence Algorithm
Assess performance on a health outcome or PRO (box 1) → Relationship between PRO and healthcare action (box 2) → Pass

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

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

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

- Current performance data are calculated from a pilot study of the S-CAHPS in 2010-2011, which included data from 2,719 survey results from 32 practices across nine specialty types.

Measures	Top Box Mean	Standard Deviation	Median	Min	Max
Information to Help you Prepare for Surgery (2 items)	90%	0.05	90%	79%	98%
How Well Surgeon Communicates with Patients Before Surgery (4 items)	85%	0.07	84%	67%	98%

Surgeon's Attentiveness on Day of Surgery (2 items)	81%	0.12	84%	42%	97%
Information to Help You Recover From Surgery - from Surgeon or another Health Provider from the office (4 items)	82%	0.07	83%	64%	100%
How Well Surgeon Communicates With Patients After Surgery (4 items)	84%	0.06	84%	73%	97%
Helpful, Courteous, and Respectful Staff at Surgeon's Office (2 items)	87%	0.08	88%	58%	100%
Global rating of Surgeon	86%	0.07	88%	70%	98%

- In addition to performance data, the developer cited 13 studies and articles, including eight additional citations that were not included in the broader evidence overview, that indicate an opportunity for improvement or less than optimal performance rates in this healthcare area.

Disparities

- The developer reported disparities data on the factors gender, age, and ethnicity.
  - The developer reported slight differences in the rates by gender. On the item *Surgeon's Attentiveness on Day of Surgery*, male respondents report 82% while female reported 80%. On the item *How Well Surgeon Communicates with Patients After Surgery*, male respondents reported 85% while female respondents reported only 83%.
  - No difference in scores due to ethnicity were reported.
  - Older patients generally reported more positive patient experiences.
  - Patients age 18-24 (the youngest group) reported the highest top box scores for global rating of surgeons.

Questions for the Committee:

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

Preliminary ratings for opportunity for improvement:

1) Information to help you prepare for surgery:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficien**
2) How well surgeon communicates with patients before surgery:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficien**
3) Surgeon's attentiveness on day of surgery:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficien**
4) Information to help you recover from surgery:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**
5) How well surgeon communicates with patients after surgery:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**
6) Helpful, courteous, and respectful staff at surgeon's office:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**
7) Rating of Surgeon:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**

Committee pre-evaluation comments

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

1a. Evidence

Comments:

**The evidence for support of this measure is solid with three new systematic reviews and six additional studies being cited since this measure was initially endorsed. I am unaware of new studies that change the base of evidence for this measure.

**This measure does not depart substantially from the constructs already represented in H-CAPHS.

**Developer provided data form 3 systematic reviews supporting that patient-reported experience measures (PREMs) were associated with multiple positive outcomes (reduced complications ,better adherence, improved clinical outcomes, reduced utilization)

1b. Performance Gap

Comments:

**Data are provided that demonstrate a gap in performance and thus a performance measure is warranted. Disparities data was provided demonstrating differences between gender and age, but not ethnicity (however, the respondent sample was overwhelmingly white - 86%).

**I have concerns about: 1) at which level scores (i.e. who is the 'target population'?) are to be reported (site, surgeon); 2) in the data provided, for 11 of the 32 sites, practice and state were completely confounded; and 2) there appear to b ceiling effects in the top box scores 4 of the 7 S-CAHPS constructs based on data from the pilot study (p. 33).

**While all measures are favorably skewed, there is variation. Also, my own experience working with ABIM on physician peer review data indicated that even in assessments with ceiling effects, marked negative performance outliers can still be identified.

The SDS analysis seemed thin, with little ethnic gap demonstrated and no racial analysis (but racial data provided for test sample but not ethnicity data)? This strikes me as a significant gap and I would like to see developer evaluate disparities more comprehensively (particularly as I would presume that ethnic and racial minorities respond less frequently to PREMs than whites/non-Hispanics and they recommend risk adjustment by these factors, which might obscure relevant disparities compared with stratification).

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: NQF Staff

Evaluation of Reliability and Validity:
Link A

Additional Information regarding Scientific Acceptability Evaluation:

While this is an outcome measure and therefore considered complex, the measure was not reviewed by the Scientific Acceptability Methods Panel because no new testing was submitted from previous endorsement review.

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

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

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

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

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

2a1: Reliability- Specifications
Comments
** Reliability specifications appear to be defined and clearly described. Consistent implementation may be challenged by resources to deploy the survey (mailing and emailing patients), data entry with accuracy checks (paper surveys), and conducting survey analyses
** I am most concerned about the sampling strategy given the intended attribution to apparently both the surgeon and the hospital. It is not clear from the data provided from the pilot exactly how many patients per surgeon and surgeons per practice were sampled. Score estimates that do not account for the nested nature of the data are problematic.
** The greatest concern is responder bias which seems to be mostly addressed through possible risk adjustment, which seems an inadequate response to this concern.

2a2: Reliability- Testing
Comments
**No.
** Yes. I am concerned about the "site" level analysis performed. Because of the hierarchical nature of the data, in some cases it appears that there may be 1 surgeon for some sites (i.e. complete confound) vs. up to 19 providers at each site. Variation between providers across sites vs within providers at the same site (where the site and surgeon are not completely confounded) did not appear to be provided. Also, the "site" level ICC's are very high. I am concerned that

the reliability testing was at the item level within patients averaged for the site, not a true between provider/site vs. within provider site variation. Site level ICC's calculated in that way tend to produce coefficients in the range of .03-.05.
**No

2b1: Validity—Testing
2b4-7: Threats to Validity
2b4: Meaningful Differences
Comments
** The validity appears to be acceptable. In the initial survey of 2719 respondents, only 3.7% of data were missing whicl should not threaten validity.
** To assess whether differences observed in the quality improvement studies cited are meaningful (vs. within the standard error of measurement), estimates of the error variance at the units being compared are needed.

I am concerned about the inclusion of patients who completed at least one item of the composite (see p. 40). For the "how well surgeon communicates with patients after surgery", missing data was observed to be 16% at the individual level. It is not clear how/what imputation was done.
** I think the validity is acceptability strong. I particular like the developers use of patient focus groups to ensure domains/instrument content were truly patient-centered.

2b2-3: Other Threats to Validity
2b2: Exclusions
2b3: Risk Adjustment
Comments
** Appropriate patient groups are included. Risk adjustment - optional case-mix risk adjustment is available and is supported by available analysis instructions.
** It appears that the majority of patients (48% of survey respondents) from the pilot were 65+; this population typicall reports higher scores on satisfaction measures, independent of the quality of care provided.
** The risk adjustment appears to be guidance and detailed information about predictive ability of the recommended risk variables was not found. Also, they do not provide any performance data stratified by SDS factors, despite flagging that these might be reasonable risk adjusters. In all, I found the information about risk adjustment less clear than the other components and I have concerns that stratification or other opportunities to illuminate disparities were not detailed or perhaps even considered.

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

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

- This measure is collected via a survey of surgeon's patients.
- CAHPS surveys are primarily delivered via mail. Electronic databases are then created after mailed surveys are returned.
- Email has been added as a mixed mode strategy for surgeon groups with reliable email addresses for all of their population.
- Because the survey instrument, protocol, analysis, and reporting are standardized, surgeons can benchmark and compare their performance with that of their peers within the same practice or outside of their practice.
- There is no fee associated with the CAHPS measure.
- In addition to the survey instrument, users can access comprehensive fielding, analysis, and reporting guide as well as SAS programming code that performs analysis and significance testing.

Questions for the Committee:

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

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3. Feasibility
Comments
** All data are derived from patient responses to a lengthy survey (47 questions). Having an electronic option may reduce the survey burden for patients with access to a computer, may increase data accuracy. and may enhance response rate.
** The response rate on S-CAHPS, as on H-CAHPS is low, raising issues about representativeness of the population of patients seen at sites/by providers.
** Feasibility remains a concern, as with all PROMs/PREMs, but the use of multiple modalities for data collection and I hope movement in the future towards lower burden electronic options will continue to minimize this issue.

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact/improvement and unintended consequences

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

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

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

Current uses of the measure
Publicly reported? ☒ **Yes** ☐ **No**

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

Accountability program details

- The measure is currently used in CMS Quality Payment Program (QPP) Merit-based Incentive Payment System (MIPS) and as part of an Advanced Alternative Payment Model (A-APMs). The measure is included in the CMS Core Quality Measures Orthopedics Set.

- The measure is also used quality improvement programs including the American College of Surgeon's National Surgical Quality Improvement Program.
- In the 2015 Medicare Physician Fee Schedule final rule, CMS agreed that the S-CAHPS survey would be more relevant to a surgical group practice compared to the CG-CAHPS and noted that the majority of commenters supported the use of S-CAHPS in the PQRS program. However, CMS did not accept and finalize the measure. CMS explained "due to the cost and time it would take to find vendors to collect S-CAHPS data, it is not technically feasible to implement the reporting of the S-CAHPS survey measures for the 2017 or 2018 PQRS payment adjustments.
- Pending CMS approval for the 2018 MIPS program, participants in the American College of Surgeons (ACS) Surgeon Specific Registry (SSR) will have the option to report on components of the S-CAHPS as part of the Phases of Surgical Care Measures for the MIPS Qualified Clinical Data Registry (QCDR) reporting mechanism.

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

Feedback on the measure by those being measured or others

- The developer noted that it was "too early to broadly determine" how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

Additional Feedback: N/A

***Questions for the Committee**:*

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

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

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

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

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

Improvement results

- There is no repository for S-CAHPS data, therefore trend data is not available.

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

Unexpected findings (positive or negative) during implementation

- The developer reported no unexpected findings have been uncovered.

Potential harms

- The developer reported no unexpected findings have been uncovered.

Questions for the Committee:

o *How can the performance results be used to further the goal of high-quality, efficient healthcare?*

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

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a1. Use- Accountability and Transparency

Comments

** The measure is currently publicly reported and used in accountability programs.
Developers report it is too early to determine the impact of the performance data being reported.
** Uses in MIPS programs raises concerns about reliability and validity, particularly at the surgeon level.
** Currently in QPP and ACS is integrating into APMs and their registry. I think goals for transparency are laudable. I would like to see efforts to make this data more available to patients as well.

Criterion 5: Related and Competing Measures

Related or competing measures

- 0005 : CAHPS Clinician & Group Surveys (CG-CAHPS)-Adult, Child
- 0006: Consumer Assessment of Healthcare Providers and Systems (CAHPS) Health Plan Survey, Version 5.0 (Medicaid and Commercial)
- 0166: HCAHPS
- 0258: CAHPS In-Center Hemodialysis Survey
- 0517: CAHPS Home Health Care Survey (experience with care)
- 2651: CAHPS Hospice Survey (experience with care)
- 2548: Child Hospital CAHPS (HCAHPS)
- 2967: CAHPS Home- and Community-Based Services Measures

Harmonization

- The Surgical Care Survey was updated in 2011 to remain consistent with the Clinician & Group Survey, which was also updated in 2011. The updates from do not affect the ability of survey users to assess trends in performance.

Committee pre-evaluation comments

Criterion 5: Related and Competing Measures

N/A

Public and member comments

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

- No comments received.

- **Zero NQF members who have submitted a support/non-support choice.**

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: 1741
Measure Title: CAHPS Surgical Care Survey 2.0

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)

   ☐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)

For estimated reliability based on ICCs, .70 often is regarded as a minimum acceptable value

Score Level Reliability Rating	Measures	Average # of Respondents per site	Site-Level Reliability
Low	Information to Help You Prepare for Surgery (2 items)	85	0.52
Low	How Well Surgeon Communicates with Patients Before Surgery	76	0.68
Low	Surgeon Attentiveness on Day of Surgery (2 items)	83	0.50
Moderate	Information to Help You Recover from Surgery - from Surgeon or another Health Provider from the office.	85	0.71
Low	How Well Surgeon Communicates With Patients After Surgery	72	0.48
Moderate	Helpful, Courteous, and Respectful Staff at Surgeon's Office (2 items)	84	0.71
Low	One-item Global Rating of surgeon	82	0.60

In the original measure submission, developer notes that the site-level reliability analysis was conducted on field test data consisting of a relatively small selection of surgeon practices which may have led to reduced variability between sites.

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)

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

Internal Consistency Reliability (Chronbach's alpha) was used to assess reliability for each of the survey questions.

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)

Data Element Reliability Rating	Measure and Items	Standardized Cronbach's Alpha
Moderate	1. Information to Help You Prepare for Surgery (2 items)	0.74
Moderate	2. How Well Surgeon Communicates with Patients Before Surgery	0.82
Moderate	3. Surgeon Attentiveness on Day of Surgery (2 items)	0.66
Moderate	4. Information to Help You Recover from Surgery - from Surgeon or another Health Provider from the office.	0.84
Moderate	5. How Well Surgeon Communicates With Patients After Surgery	0.86
Moderate	6. Helpful, Courteous, and Respectful Staff at Surgeon's Office (2 items)	0.85

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)

   S-CAHPS Analysis program provides optional risk adjustment through case-mix adjusted scores. Case-mix adjustment is determined by users based on decision of what is most appropriate to adjust for to account for case-mix differences.

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)

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)

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 <u>empirical</u> 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)

8. Was <u>face validity</u> 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 <u>performance measure score</u> 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)

10. Was validity testing conducted with <u>computed performance measure scores</u> 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)

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

The developer examined the relationships between each individual item's top box score and the top box score for the global PRO-PM of "How would you rate your surgeon?" using Spearman rank-order correlations at the site level to determine the validity of each of the PRO-PMs.

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

The survey results show that each of the PRO-PMs are related to the global PRO-PM (rating of surgeon) at the individual level and five of the six PRO-PMs are related to the global rating at the practice level. The two Communication PRO-PMs and the Recovery Information PRO-PM have the strongest relationship with the global rating of surgeon.

Validity Measure Score Rating	Measures
High	1. Information to Help You Prepare for Surgery
High	2. How Well Surgeon Communicates with Patients Before Surgery
Moderate	3. Surgeon Attentiveness on Day of Surgery
High	4. Information to Help You Recover From Surgery - from Surgeon or another Health Provider from the office.
High	5. How Well Surgeon Communicates With Patients After Surgery
High	6. Helpful, Courteous, and Respectful Staff at Surgeon's Office
High	7. One-item Global Rating of Surgeon

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)

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)

The developer examined Spearman rank-order correlations among the PRO-PMs to assess the extent to which they measure different constructs.

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)

The developer examined Spearman rank-order correlations among the composites to assess the extent to which they measure different constructs. While the composites are correlated with each other, intercorrelations greater than 0.8 may indicate that the composites are not unique enough to be considered separate measures. One intercorrelation, "Communicate post surgery," and "communicate pre surgery" received a score at or above that level, with a score of 0.8.

Developer states that the relationships met their expectations.

Validity Rating	Measures
Moderate	1. Information to Help You Prepare for Surgery
Moderate	2. How Well Surgeon Communicates with Patients Before Surgery
Moderate	3. Surgeon Attentiveness on Day of Surgery
Moderate	4. Information to Help You Recover From Surgery - from Surgeon or another Health Provider from the office.
Moderate	5. How Well Surgeon Communicates With Patients After Surgery
Moderate	6. Helpful, Courteous, and Respectful Staff at Surgeon's Office
Moderate	7. One-item Global Rating of Surgeon

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.]

Overall Validity Rating	Measures
High	1. Information to Help You Prepare for Surgery
High	2. How Well Surgeon Communicates with Patients Before Surgery
High	3. Surgeon Attentiveness on Day of Surgery

High	4. Information to Help You Recover From Surgery - from Surgeon or another Health Provider from the office.
High	5. How Well Surgeon Communicates With Patients After Surgery
High	6. Helpful, Courteous, and Respectful Staff at Surgeon's Office
High	7. One-item Global Rating of Surgeon

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 1741

Measure Title: CAHPS Surgical Care Survey Version 2.0

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 11/8/2017

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: PRO-based Performance Measure (PRO-PM)

☒Patient-reported outcome (PRO): Experience with Surgical Care

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*): Click here to name the intermediate outcome

☐ Process: Click here to name what is being measured

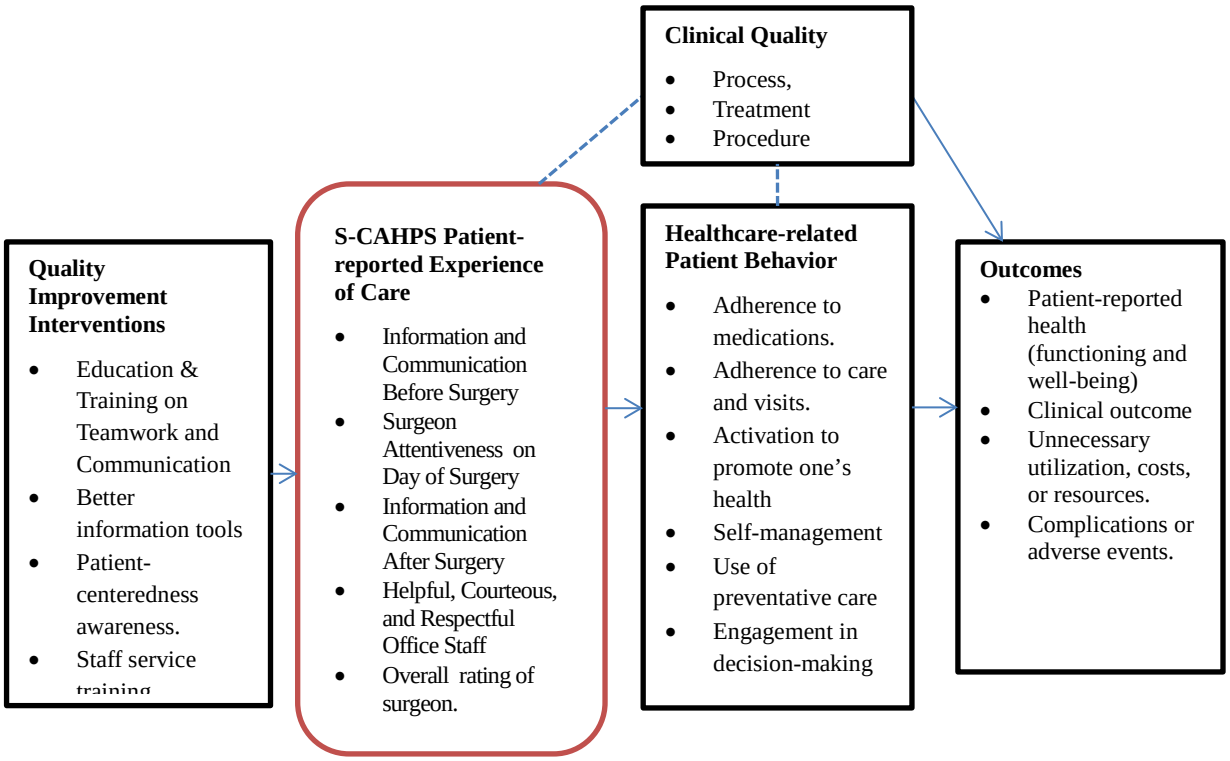
☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

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.

The following framework will be used to describe the patient-reported Surgical Consumer Assessment of Healthcare Providers and Systems (S-CAHPS) measure, the factors that influence it, and the relationship between S-CAHPS and clinical outcomes.



----- Hypothesized Association

___ Hypothesized Casual Association

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.)

****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.

INTRODUCTION

The CAHPS Surgical Care Survey measures key components of patient experience, such as provider communication, that are consistent with patient-centered care. All CAHPS surveys focus on aspects of care that consumers have identified as important and for which patients are

the best or only source of information. We reviewed the literature on the determinants of patient care experiences measured by CAHPS and their associations with other indicators of health care quality. Patient care experiences are influenced by the quality of patient care and interventions to improve quality of care. In turn, patient perceptions of high quality care improve patient adherence and activation, leading to improved clinical outcomes (See Figure 1.) Surgical CAHPS is an actionable measure that helps surgeons and their practices target interventions that will improve the quality and patient-centeredness of care.

REVIEW OF THE EVIDENCE

Three major systematic reviews have examined the relationships among patient experience, clinical processes, and patient outcomes. A systematic review performed by researchers in the U.K. found that patient experience is favorably associated with adherence to recommended medications and treatments, preventive care such as screenings and immunizations, patient-reported health outcomes, clinical outcomes, reduced healthcare utilization, and reduced adverse events (Doyle et al., 2013). More recently, in the U.S., Anhang Price et al. reviewed evidence on the association between patient experiences and other measures of health care quality (Anhang Price, 2014). They similarly found that better patient care experiences are associated with higher levels of adherence to recommended prevention and treatment processes, better clinical outcomes, and less health care utilization. Beattie et al. critiqued the utility of published patient-reported experience measures (PREMs) aiming to measure the adult inpatient experience of hospital quality of care based on the PREMs' validity, reliability, cost efficiency, acceptability and educational impact (Beattie et al., 2015). They identified eleven international PREMs and concluded that patient experience data could be used to drive improvements in hospital care at national, local, and healthcare team levels. Most importantly, the authors cautioned that clinicians, managers, policymakers and researchers need to select PREMs that are fit for purpose because only selecting the right PREM for the right purpose can the data aid in quality improvement.

RATIONALE OF INFLUENTIAL RELATIONSHIPS

Quality Improvement Initiatives/Interventions

The Agency for Healthcare Research and Quality (AHRQ) maintains a Quality Improvement (QI) Guide that assists providers and health plans in using CAHPS scores to identify problems, prepare staff for QI initiatives, select and implement interventions, and evaluate intervention effectiveness through changes in CAHPS scores. The guide presents interventions that have been successfully used to improve patient experience of care and other measures of quality relevant to CAHPS dimensions (AHRQ CAHPS QI Guide, 2012).

Relevant to surgical care, the AHRQ QI Guide provides information that can help surgeons to improve shared decision-making. Shared decision-making is a model of patient-centered care that enables and encourages people to play a role in the management of their own health. It operates under the premise that armed with good information, consumers can and will participate

in the medical decision-making process by asking informed questions and expressing personal values and opinions about their conditions and treatment options. Devine et al. (1983 and 1988) found that improved quality of medical consultations and patient education had positive effect on the surgical patients' length of stay, anxiety levels, recovery time, and compliance with treatment regimens. Interactive technology and instructional applications are employed to help prepare patients for various procedures, to inform them about their surgery, and to explain what they need to know after surgery. The QI Guide notes the importance of training physicians to help them understand how to facilitate the shared decision-making process and to ensure that they appreciate the importance of respecting patient's values, preferences, and expressed needs. The shared decision-making model should be a practice-wide, team approach so that the surgeon's time is used appropriately.

A common intervention is communication skills training, with the purpose of improving provider communication (Hardee & Kasper, 2008). Five clinics in San Francisco took part in a two-year learning collaborative aimed at improving provide/staff-patient communication and increasing patient partnership. The intervention's focus was to ensure that patients' most important concerns were addressed during their visits. After the intervention, all five clinics showed significant improvement in survey items "doctor spends enough time", "doctor's explanations are understandable", "doctor provides easy-to-understand instructions", and "clerks and receptionists are helpful". Ten months post-intervention, these clinics had sustained statistically significant improvements in eight of 12 measures (Fisher, 2011). Improving staff service is also a common intervention. To address service issues, a surgical practice associated with an academic medical center offered a series of training courses to help staff interact with patients in a more positive manner. The practice's faculty group developed efforts to improve communication with patients. This emphasis helped to support the practice manager by signaling a wider commitment to the goal of improving patient experience. Their CAHPS scores for the helpful office staff composite increased to 87 (based on a 0-100 mean score) from a baseline score of 84, meeting the target set of >85. The overall rating of the doctor score increased from 89 to 95, exceeding the target of >90 (Shaller, 2011).

The CAHPS domain measures are available to assess all of the organizational processes that are addressed by health care organizations' innovations and for which patients are the best source of information. A recent study of AHRQ's Innovations Exchange website was conducted to examine the use of patient experience surveys in assessing the impact of innovations implemented in health care settings. Researchers found that fewer than half of the innovations used a patient experience measure. The authors conclude that there is considerable untapped potential for using CAHPS measures or surveys to assess QI initiatives' effectiveness. Organizations committed to patient-centeredness will benefit by monitoring patient survey data, along with clinical and operational data, to implement and measure quality improvement interventions and/or to evaluate patient reported outcomes associated with new models of care (Weinick, 2014; McWilliams, 2014).

Health-related Patient Behavior

Interventions targeting CAHPS dimensions indirectly improve clinical outcomes by positively influencing patient behavior (Fuertes, Boylan et al. 2009). For example, in a 2009 meta-analysis, Zolnierik and Dimatteo (2009) found evidence that patients' treatment adherence improved significantly more among patients whose physicians participated in communication skills training. A 2009 meta-analysis of 127 studies assessing the link between patient treatment adherence and physician-patient communication found a 19% higher risk of non-adherence among patients whose physician communicated poorly (Zolnierek and Dimatteo 2009). Doyle's (2013) meta-analysis showed positive associations between the quality of clinician-patient communications and adherence to medical treatment in 125 of 127 studies analyzed. Studies using the CAHPS measure have found that better provider communication is positively associated with adherence to hypoglycemic medications among diabetics (Ratanawongsa, Karter et al. 2013), adherence to tamoxifen among breast cancer patients (Liu, Malin et al. 2013), and higher rates of colorectal cancer screening among adults in the US (Carcaise-Edinboro and Bradley 2008).

Clinical Quality

Sequist and colleagues (2008) found that measures of patient experience, including doctor-patient communication, clinical team interactions, and health promotion support, were positively associated with some prevention and disease management clinical process measures in clinical practices and among individual clinicians. Patients' overall ratings of their hospitals have been positively associated with hospitals' performance on CMS's process measures for pneumonia, congestive heart failure, AMI and surgical care in the US (Isaac, Zaslavsky, Cleary, & Landon, 2010), and to process indicators relating to 19 different conditions in the UK (Llanwarne, et al., 2013).

Outcomes

Out of 40 evidence papers with outcome measures, Doyle's (2013) meta- analysis found 29 studies with positive associations between patient experience and clinical outcomes, 11 with no associations, and none with negative associations. The lack of more evidence may be due to complexity between a patient's illness level, their level of care, and their likelihood for a poor outcome such as mortality, morbidity or a readmission. Often, such associations have more than one plausible direction of causality. For example, clinicians may be especially attentive to the needs of sicker patients (Kahn et al., 2007) and patients near the end of life (Elliott, Haviland et al., 2013).

Research suggests an association between better patient experiences and lower healthcare utilization. Among African Americans with Type 2 diabetes, those who reported that doctors or nurses usually listened carefully or spent enough time with them were significantly less likely to visit the emergency department in the 12 months following completion of a patient experience survey (Gary, Maiese et al. 2005).

In surgery, the association between patient experience and quality of care has been investigated. Tsai and colleagues studied more than 2,900 hospitals between 2010 and 2011 by merging 100% Medicare inpatient claims data with the H-CAHPS survey data. Patients were included if they underwent coronary artery bypass grafting, pulmonary lobectomy, endovascular aortic aneurysm repair, open abdominal aortic aneurysm repair, colectomy, and hip replacement. They found that hospitals in the highest quartile of performance on patent satisfaction had length of stay that was on average 0.6 days shorter than those with the lowest patient satisfaction. This was also true for readmission rates and risk-adjusted perioperative mortality rates. More recently, Lobo Prabhu and colleagues examined the association of patient satisfaction and clinical outcomes using the American College of Surgeons' National Surgical Quality Improvement Program (ACS NSQIP) framework. Using standard ACS NSQIP follow-up procedures, patients were provided a survey asking about their experiences. The investigators found a significant association between patient satisfaction and both 30-day readmission and the occurrence of postoperative surgical complications. They suggested further study is warranted to evaluate patient satisfaction as a healthcare quality indicator.

Since NQF-endorsement, the S-CAHPS Survey has demonstrated uptake on several fronts. First, three peer-reviewed publications have been published. Schmocker et al. used the S-CAHPS survey to determine which aspects of perioperative care were predictive of satisfaction with the surgeon. Their response rate was 45.3%. They found that on multivariable analysis, preoperative communication and attentiveness on the day of surgery were the most important determinants of overall surgeon rating. Lenherr et al. published the first experience and results of using S-CAHPS in urology. With a 33.8% response rate, their data suggested patient satisfaction with the surgeon is more influenced by postoperative communication and information. Jiang and Malkin used Lean A3 thinking to analyze S-CAHPS survey data to identify quality improvement opportunities. Care processes in their postoperative clinic were modified and they found improvement in their S-CAHPS survey scores on the domains that they targeted.

Second, the S-CAHPS has also been utilized by an institution for internal quality improvement purposes when structuring its new ambulatory surgery center. At this institution, 25% (n=238/951) of ambulatory surgical patients reported using a web-based portal on the S-CAHPS that they were dissatisfied with their surgeon's attentiveness on the day of surgery (Figure A; publication in preparation). Patient throughput and flow management processes were reviewed, which identified an unexpected consequence of a highly efficient ambulatory surgery environment — surgeons were caring for the next patient in the operating

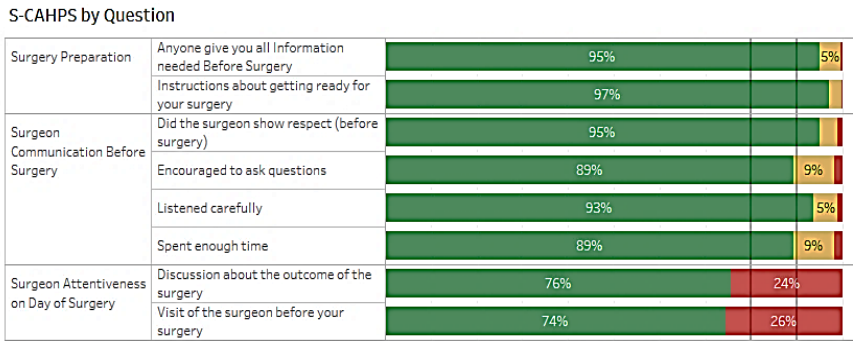


Figure A. Results of S-CAHPS survey administered to patients undergoing ambulatory surgery (n=951). Green = "Yes, definitely." Yellow = "Yes, somewhat." Red = "No."

room and unable to speak with their previous patient before discharge. Leveraging health information technology, a secure online video streaming solution was developed to facilitate surgeons in the operating room to interact face-to-face with their postoperative patient prior to discharge. The effect of this intervention on S-CAHPS scores is under active investigation.

Third, the American College of Surgeons has begun incorporating the S-CAHPS survey questions into the American College of Surgeons' National Surgical Quality Improvement Program (ACS NSQIP) to improve quality from both surgeons' and patients' perspectives (i.e., tracking clinical outcomes and patient-reported outcomes). A pilot is currently underway, launched October 16, 2017, to evaluate the feasibility of collecting PROs into the registry – the S-CAHPS is being used for this purpose. The ACS NSQIP leadership intends to continue using S-CAHPS in the future and make it available to all hospitals participating in ACS NSQIP (more than 700 as of 2017) in the first quarter of 2018.

Fourth, pending CMS approval for the 2018 Merit-based Incentive Payment System (MIPS) program, participants in the ACS Surgeon Specific Registry (SSR) will have the option to report on components of the S-CAHPS as part of the Phases of Surgical Care Measures set for the MIPS Qualified Clinical Data Registry (QCDR) reporting mechanism. The Phases of Surgical Care Measure set is an episode-based measure framework which follows the five phases of surgical care: preoperative, perioperative, intraoperative, postoperative, and postdischarge. The measure set is composed of surgical outcome measures, high-value process measures, and the appropriate S-CAHPS measures which follow up on key processes within the measure set.

Fifth, the ACS is also currently working with CMS on an Advanced Alternative Payment Model that we anticipate may soon be tested by Center for Medicare and Medicaid Innovation (CMMI). This model incorporates a novel quality measurement framework which measures care around the patient for a given episode and incorporates patient reported experience and patient reported outcomes. As currently proposed, the surgical episodes include measures included in the S-CAHPS survey.

Citations

Jiang N. and B. Malkin. Use of Lean and CAHPS Surgical Care Survey to Improve Patients' Experiences with Surgical Care. Otolaryngol Head Neck Surg. 2016. 155(5): 743-7.

Lenherr SM, DeCicco B, Cameron AP, et al. The S-CAHPS Survey in Urology. Urol Prac. 2015. 2(1):12-6.

Liu JB, Pusic AL, Temple LK, Ko CY. Patient-reported outcomes in surgery: Listening to patients improves quality of care. Bull Am Coll Surg. 2017; 102(3): 19-23.

Lobo Prabhu K, Cleghorn MC, Elnahas A, et al. BMJ Qual Saf Published Online First. doi:10.1136/bmjqs-2017-007071

Schmocker RK, Cherney Stafford LM, Siy AB, et al. Understanding the Determinants of Patient Satisfaction with Surgical Care Using the CAHPS Surgical Care Survey (S-CAHPS). Surgery. 2015. 158(6):1724-33.

Tsai TC, Orav EJ, and AK Jha. Patient satisfaction and quality of surgical care in US hospitals. Ann Surg. 2015; 261(1):2-8

AHRQ CAHPS Quality Improvement Guide (2013). Accessible at https://cahps.ahrq.gov/quality-improvement/improvement-guide/improvement-guide.html.

Anhang Price R, Elliott MN, Zaslavsky AM, Hays RD, Lehrman WG, Rybowski L, Edgman-Levitan S, Cleary PD. Examining the role of patient experience surveys in measuring health care quality. *Med Care Res Rev.* 2014 Oct;71(5):522-54.

Beattie M, Murphy DJ, Atherton I, et al. Instruments to measure patient experience of healthcare quality in hospitals: a systematic review. *Systematic Reviews*. 2015 4; 97.

Carcaise-Edinboro, P. and C. J. Bradley (2008). Influence of patient-provider communication on colorectal cancer screening. *Medical Care*. 46(7): 738-745.

Devine EC and Cook TD. A meta-analytic analysis of effects of psycho-educational interventions on length of postsurgical hospital stay. *Nurs Res* 1983;32(5): 267-74.

Devine EC, O'Connor FW, Cook TD, Wenk VA, Curtin TR. Clinical and financial effects of psycho-educational care provided by staff nurses to adult surgical patients in the post-DRG environment. *Am J Public Health.* 1988 Oct;78(10):1293-7.

Doyle, C., L. Lennox, et al. (2013). A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. *BMJ open* 3(1). http://bmjopen.bmj.com/content/3/1/e001570.full

Elliott, M. N., A. M. Haviland, et al. (2013). Care experiences of managed care Medicare enrollees near the end of life. *Journal of the American Geriatrics Society* 61(3): 407-412.

Fisher T and Gatewood H. (2011) Improving patient experience: A hands on guide for safety net clinics. California Healthcare Foundation. Accessible at http://www.chcf.org/publications/2011/10/patient-experience-safety-net-clinics

Fuertes, J. N., L. S. Boylan, et al. (2009). Behavioral Indices in Medical Care Outcome: The Working Alliance, Adherence, and Related Factors. *Journal of General Internal Medicine*. 24(1): 80-85.

Gary, T. L., E. M. Maiese, et al. (2005). Patient satisfaction, preventive services, and emergency room use among African-Americans with type 2 diabetes. *Disease Management*. DM 8(6): 361-371.

Isaac, T., Zaslavsky, A. M., Cleary, P. D., & Landon, B. E. (2010). The relationship between patients' perception of care and measures of hospital quality and safety. *Health services research*, 45(4), 1024-1040.

Kahn, K. L., Tisnado, D. M., et al. (2007). Does ambulatory process of care predict health-related quality of life outcomes? *Health Services Research*. 42: 63-83.

Liu, Y., J. L. Malin, et al. (2013). Adherence to adjuvant hormone therapy in low-income women with breast cancer: the role of provider-patient communication. *Breast Cancer Research and Treatment.* 137(3): 829-836.

Llanwarne, N. R., Abel, G. A., Elliott, M. N., Paddison, C. A., Lyratzopoulos, G., Campbell, J. L., et al. (2013). Relationship between clinical quality and patient experience: analysis of data from the english quality and outcomes framework and the National GP Patient Survey. *Annals of family medicine*, 11(5), 467-472.

McWilliams JM, Landon BE, Chernew ME, Zaslavsky AM. (2014) Changes in Patients' Experiences in Medicare Accountable Care Organizations. *N Engl J Med* 371:1715-24.

Ratanawongsa, N., A. J. Karter, et al. (2013). Communication and medication refill adherence: the Diabetes Study of Northern California. *JAMA Internal Medicine.* 173(3): 210-218.

Sequist, T. D., E. C. Schneider, et al. (2008). Quality monitoring of physicians: linking patients' experiences of care to clinical quality and outcomes. *Journal of general internal medicine*. 23(11): 1784-1790.

Shaller, D. (2011) A Tale of Three Practices: How Medical Groups are Improving the Patient Experience. RWJF Aligning Forces for Quality Case Study. Accessible at https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/quality-improvement/reports-and-case-studies/CaseStudyPatientExperience_July2011FINAL_revised20110728.pdf. .

Weinick RM, Quigley DD, et al. Use of CAHPS Patient Experience Surveys to Assess the Impact of

Health Care Innovations. *The Joint Commission Journal on Quality and Patient Safety*. September 2014. Volume 40(9): 418-427.

Zolnierek, K. B. and M. R. Dimatteo (2009). Physician communication and patient adherence to treatment: a meta-analysis. *Medical care*. 47(8): 826-834.

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

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

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Source of Systematic Review: • **Title** • **Author**	

• **Date** • **Citation, including page number** • **URL**	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

1a.4 OTHER SOURCE OF EVIDENCE

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

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

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

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

1. Evidence, Performance Gap, Priority – 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
nqf_evidence_attachment_7.1_SCAHPS_FINAL.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.
Yes

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.
All of the measures submitted to NQF for endorsement share the main objective of the S-CAHPS survey, which is to measure aspects of surgical quality that are important to consumers and for which consumers are the best source of information. At the same time, consistent with the survey, the measures aim to directly benefit a

variety of users, including: patients, practice groups, health plans, insurers, and specialty boards. Each group has a need for information regarding the quality of surgeons and surgical care.

The S-CAHPS survey measures patients' perceptions of their experiences with pre-, during and post-surgical care. The survey addresses issues relevant to surgical care, such as informed consent, communication, clear and helpful information, anesthesia care, office staff helpfulness, and post-operative follow-up.

To offer surgical patients and surgeons valid and reliable information on patient experience of care, the American College of Surgeons (ACS), in partnership with other surgical and anesthesia organizations, sponsored the development of the S-CAHPS survey. The S-CAHPS survey is a patient experience-of-care survey measure specifically tailored for surgical patients. The S-CAHPS survey was developed by working with patients to report on the full experience of surgical care, including their experience with the surgeon, the anesthesiologist, and the facility. Qualitative research (i.e., focus groups, cognitive testing, literature review) prior to field testing and focus groups following the main field test contributed to the development of the final composite measures. The data gathered through S-CAHPS survey data can assist consumers in identifying a high-quality surgeon and help surgeons to better understand and ultimately improve patient care.

To capture and ascertain the major domains surrounding the consumers' view of quality surgical care, the American Institutes for Research performed a literature review prior to developing the surgical CAHPS instrument. Using four literature databases: Medline, PsychInfo, CINAHL, and Evidence-Based Medicine Review, AIR reviewed 930 abstracts. These abstracts were narrowed from certain search terms and limitations. AIR asked the Surgical Quality Alliance (SQA) Technical Advisory Panel (TAP) members to provide input and guidance in these terms and limitations. The TAP included 21 members from various surgery and anesthesiology specialty societies, each having technical expertise in the topic of surgical care. The AIR research analyst then classified each abstract as relevant, possibly relevant, and not relevant.

From that point on, only the research analyst's "relevant" and "possibly relevant" classifications were reviewed by the project director. From this abstract review method, 37 abstracts were chosen to review in full. In addition, one article was added from an SQA Technical Advisory Panel (TAP) member. The 38 articles reviewed indicated 14 domains, or primary issues, of surgical patient care experience from a patient's perspective: information/education, interpersonal manner, pain, emotional support, accessibility/convenience, technical quality of care, efficacy/outcomes of care, availability, environment, customization/personalized care, patient involvement in care, continuity of care, overall satisfaction, and finances. These results guided the development of the surgical survey based on relevance to quality and to consumers and the ability of consumers to act as reliable reporters of the domain.

In addition, American Institutes for Research conducted six focus groups to identify important quality issues inherent in patients' experiences of surgical care. The results of the focus groups are included in Attachment D Main 1c5 Surgical CAHPS Focus Groups_2nd Round_2010.pdf. In total, 49 people participated in the focus groups, all of whom were 18 years of age or older and had undergone a surgical procedure billable with a 90-day global fee within the last 7 months. The groups were diverse both demographically and in the type of surgery. For recruiting the additional two groups of patients with more complex surgeries, each participant had their surgery in a hospital and stayed overnight for at least one night.

After the focus groups were conducted, the AIR project director, senior research analyst, and research analyst examined the notes and focused in on each of the above issues. They gathered recurring themes and issues into a report of results, used to develop insights about what patients feel characterizes quality surgical care. The three main domains were surgeon's interpersonal skills and behaviors, surgeon's expertise/technical competence, and surgeon's skill in communicating or providing health information and patient education. The results of the focus groups and the surgical patients' experiences of care ultimately guided development of the surgical survey. Though patients rated technical skill of the surgeon as highly relevant to quality, it was determined that patients are not the best reporters of technical surgical expertise, so this was not included as a

domain in the survey instrument. Additionally, based on the psychometric analysis and discussions within the Surgical Quality Alliance (SQA) Technical Advisory Panel (TAP), ACS did not recommend the Anesthesia Care items as a reporting composite. While ACS believes these items should remain in the survey because anesthesia care is an important factor in the surgical patient's experience of care, the Anesthesiologist works within the hospital system, and thus, the surgeon cannot adequately control the experiences between patient and anesthesiologist, and should not be measured on such.

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.*

Performance data with the statistics requested are provided in the attached Excel file named "Attachment B Main 1b2 S-CAHPS_score_Tables.xlsx." The summary statistics at the practice site level (mean, SD, min, max, deciles) can be found in the worksheet tab titled "Summary Stats".

Performance data from a pilot study of the S-CAHPS in two surgical subspecialties is shown below. The pilot study was an IRB-approved secondary analysis conducted of S-CAHPS data collected by Duke Otolaryngology Head and Neck Surgery (OHNS, via web-based electronic data capture system) and University of Michigan Urology (GU, via postal mail) from 2011-2013. A total of 2695 adult patients were administered S-CAHPS within 4 weeks of surgery (1424 OHNS, 1271 GU). Survey content was separated into 6 composite scores and analyzed by %-top box scoring for separate and pooled data. A total of 727 patients completed the survey (n=303, 21.3% OHNS and n=424, 33.8% GU). Full survey completion rate for all rated questions was 62% OHNS and 72% GU. Composite top-box scores were similar between OHNS and GU except for Communication Pre-Operatively. As shown in the Table below, pooled overall surgeon rating was high (88% top-box scores, ranked 9-10 out of 10). The overall surgeon rating was most correlated with surgeon communication pre- and post-operatively, followed by information to recover from surgery. Differences in both surgical subspecialty and mode of administration yielded similar responses to S-CAHPS, but mailed GU survey yielded a higher response rate.

S-CAHPS Composite Top-Box Results from Lenherr et al. study

Composite	% Top Box Pooled OHNS and GU (n=727)
Information to prepare for surgery	92%
Communication pre-operatively	92%
Surgeon's attentiveness day of surgery	84%
Information to recover from surgery	83%
Communication post-operatively	91%
Helpful, courteous and respectful staff	92%
Overall surgeon rating	88%

Citations:

Lenherr SM, DeCicco B., Cameron AP, Malaeb BS., Oldendorf AL, Stoffel JT, Karls EM, and Clemens JQ. The S-CAHPS Survey in Urology. (in press) Urology Practice. Vol 2. Available online at http://www.urologypracticejournal.com/article/S2352-0779(14)00109-5/pdf.

Schulz KA, Rhee JS, et al. (2012) Consumer Assessment of Healthcare Providers and Systems Surgical Care Surveys: Benefits and Challenges. Otolaryngology - Head and Neck Surgery. 147(4):671-7.

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.**

The CAHPS Surgical Care Survey was implemented by the American Academy of Otolaryngology (AAO) – Head and Neck Surgery from August 2010 through March 2011. The S-CAHPS was administered in four sites with a

total of 14 surgeons. Two sites were academic medical centers, and the other two sites were community practices. A total of 354 surveys were collected across the three sites. An overall response rate of 39.9% was achieved with individual site response rates of 37.7%, 35.7% and 47.9% respectively. As can be seen in the table from Schulz et al., minimum scores at the surgeon level were quiet low. The overall scores for Recovery Information and Surgeon Communication have much room for improvement.

Schmocker et al. used the S-CAHPS survey to determine which aspects of perioperative care were predictive of satisfaction with the surgeon. Their response rate was 45.3%. They found that on multivariable analysis, preoperative communication and attentiveness on the day of surgery were the most important determinants of overall surgeon rating. Lenherr et al. published the first experience and results of using S-CAHPS in urology. With a 33.8% response rate, their data suggested patient satisfaction with the surgeon is more influenced by postoperative communication and information. Jiang and Malkin used Lean A3 thinking to analyze S-CAHPS survey data to identify quality improvement opportunities. Care processes in their postoperative clinic were modified and they found improvement in their S-CAHPS survey scores on the domains that they targeted.

Below are several case studies of physician practices that have used the CAHPS survey to focus on and implement successful quality improvement interventions.

Five clinics in San Francisco improved the scores in communication after focusing on patients' most important concerns (Fisher and Gatewood, 2011). In Massachusetts, the Massachusetts General Physicians Organization implemented procedures for informing patients of waits, service recovery, physician communication and coaching, and staff huddles. These approaches coupled with education, training and recognition programs led to improved CG-CAHPS scores.

At the Dean Clinic in Wisconsin (over 800 medical staff in 60 locations), the service department shadowed staff and provided feedback. To improve consistency in service across all sites, the Clinic developed an orientation for all new employees on customer service expectations. They also offered ongoing training in the form of service workshops, videos, and Webinars, as well as targeted interventions for the lowest scoring offices. As a result, Dean Clinic's overall performance on the "Helpful, Courteous, and Respectful Office Staff" composite measure increased from 79 percent in 2011 to 83 percent in 2013 (AHRQ CAHPS Website).

In July 2014, The Robert Wood Johnson Foundation updated their inventory of CAHPS quality improvement resources. This document offers tools to support health care organizations in determining what they need to do to improve patient experience and how to implement those improvements. These resources are available for both ambulatory care settings and hospitals and can be found here: http://forces4quality.org/af4q/download-document/6540/Resource-12-125_inventory_of_pat_exp_improvement_resources_-_designed_-_revised_11.3.pdf .

Providers routinely use patient experience measures such as CAHPS to guide quality improvement (QI) efforts (Friedberg et al, 2011; Davies, Shaller et al., 2013). Friedberg et al (2011) found that physician groups commonly targeted improvement at access, communication with patients, and customer service by addressing office workflow, providing additional training for nonclinical staff, and adopting or enhancing an electronic health record.

Citations:

AHRQ CAHPS Website: Quality Improvement Reports and Case Studies accessible at https://www.ahrq.gov/cahps/quality-improvement/reports-and-case-studies/Case-Study_QI-Initiatives.html

Fisher T and Gatewood H. Improving patient experience: A hands on guide for safety net clinics. California Healthcare Foundation. 2011. Accessible at http://www.chcf.org/~/media/MEDIA%20LIBRARY%20Files/PDF/PDF%20I/PDF%20ImprovingPatientExperienceHandsOnGuide.pdf

Davies E, Shaller D, Edgman-Levitan S, Safran DG, Oftedahl G, Sakowski J, Cleary PD. (2008) Evaluating the use of a modified CAHPS survey to support improvements in patient-centred care: lessons from a quality improvement collaborative. Health Expect. Jun;11(2):160-76.

Friedberg MV, SteelFisher GK, Karp M, and Schneider EC. (2011) Physician groups' use of data from patient experience surveys. J Gen Intern Med. 26(5): 498-504.

Schulz KA, Rhee JS, et al. (2012) Consumer Assessement of Healthcare Providers and Systems Surgical Care Surveys: Benefits and Challenges. Otolaryngology – Head and Neck Surgery. 147(4):671-677.

Henderson A, Caplan G, Daniel A. Patient satisfaction: the Australian patient perspective. Aust Health Rev. 2004;27(1):73-83.

Hepner DL, Bader AM, Hurwitz S, Gustafson M, Tsen LC. Patient satisfaction with preoperative assessment in a preoperative assessment testing clinic. Anesth Analg. 2004 Apr;98(4):1099-105, table of contents.

Lledó R, Rodriguez-Ros T, Targarona EM, Trias M, Trilla A, Asenjo MA. Perceived quality of care of inguinal hernia repair: assessment before and after the procedure. Int Surg. 2000 Jan-Mar; 85(1):82-7.

Tabolli S, Molino N, Abeni D, Sampogna F. Satisfaction with care in vascular surgery inpatient units. Eur J Vasc Endovasc Surg. 2003 Jan;25(1):48-52.

Wolosin RJ. Patient satisfaction in gastroenterology clinics. Gastroenterol Nurs. 2003 Sep-Oct; 26(5):203-8.

Jiang N. and B. Malkin. Use of Lean and CAHPS Surgical Care Survey to Improve Patients' Experiences with Surgical Care. Otolaryngol Head Neck Surg. 2016. 155(5): 743-7.

Lenherr SM, DeCicco B, Cameron AP, et al. The S-CAHPS Survey in Urology. Urol Prac. 2015. 2(1):12-6.

Schmocker RK, Cherney Stafford LM, Siy AB, et al. Understanding the Determinants of Patient Satisfaction with Surgical Care Using the CAHPS Surgical Care Survey (S-CAHPS). Surgery. 2015. 158(6):1724-33.

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.*

Top box scores from the adult survey by gender, race and by age category are provided in worksheets "Sum Stats Ethnicity", "Sum Stats Age" and "Sum Stats Gender" in the attached Excel file named Attachment B Main 1b2 S-CAHPS_score_Tables.xlsx.

Differences in the top box scores by gender are small. There is also not an apparent difference in scores due to ethnicity.

Older patients generally reported more positive patient experiences. The group that gave the highest top box scores for global rating of surgeon were the youngest group of patients (18-24 years) and patients age 65-74 (93% and 89%, respectively). The scores from the older group is consistent with a study of Hospital CAHPS. In that study, O'Malley et al (2005) found that younger age patients (18-24) scored significantly lower than patients 25-34 and patients in the eight age categories above 34 all scored higher than those 25-34.

Citations:
O´Malley AJ, Zaslavsky AM, Elliott MN, Zaborski L and Cleary PD (2005) Case-Mix Adjustment of the CAHPS® Hospital Survey. Health Services Research. Dec. Volume 40, Issue 6p2, pages 2162–2181.

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**
N/A

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):*
Surgery

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):*
Elderly

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
https://www.ahrq.gov/cahps/surveys-guidance/surgical/index.html

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.
Attachment **Attachment:** surgical_eng-636461813392404132.pdf

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.
Patient

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.
Yes

S.3.2. **For maintenance of endorsement,** please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.
Recommendation to include E-mail as part of Mixed-mode Survey Administration

Based on field test results, the CAHPS team recommends the following modes:
- Mail only.
- Telephone only.
- Mixed mode (mail and telephone, email and mail, or email and telephone).

The addition of e-mail administration (i.e., notification for web-based surveys) as a type of mixed-mode data collection (Drake et al., 2014; McInnes et al., 2012) is a recommendation since last endorsement. The CAHPS Consortium recommends including an option to conduct a mixed mode survey that would have two e-mail reminders and a follow-up by mail or telephone to all who are surveyed. The follow-up to the entire sampling frame is necessary to get a representative sample from a practice that is not based just on e-mail alone. The Consortium does not recommend a mailed hard copy letter with a link to a web survey.

While there are no explicit data on the use of electronic capture for the CAHPS Surgical Care Survey, the psychometric properties of surveys does not appear to change when transition from paper to electronic data capture (Bennett AV, et al. 2014).

Citation:
Bennett AV, Keenoy K, Shouery M, Basch E, Temple LK. Evaluation of mode equivalence of the MSKCC Bowel Function Instrument, LASA Quality of Life, and Subjective Significance Questionnaire items administered by Web, interactive voice response system (IVRS), and paper. Qual Life Res. 2016; 25:1123-30.

Drake KM, Hargraves JL, Lloyd S, Gallagher PM, Cleary PD. The Effect of Response Scale, Administration Mode, and Format on Responses to the CAHPS Clinician and Group Survey. Health Serv Res. 2014 Aug;49(4):1387-99. doi: 10.1111/1475-6773.12160

McInnes DK, Brown JA, Hays RD, Gallagher P, Ralston JD, Hugh M, Kanter M, Serrato CA, Cosenza C, Halamka J, Ding L, Cleary PD. (2012) Development and evaluation of CAHPS questions to assess the impact of health information technology on patient experiences with ambulatory care. Med Care. 2012 Nov;50 Suppl:S11-9.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
We recommend that S-CAHPS Survey items and composites be calculated using a top-box scoring method. The top box score refers to the percentage of patients whose responses indicated excellent performance for a given measure. This approach is a kind of categorical scoring because the emphasis is on the score for a specific category of responses.

The top box numerator for the Overall Rating of Surgeon is the number of respondents who answered 9 or 10 for the item, with 10 indicating “Best provider possible”.

For more information on the calculation of reporting measures, see What´s Available for the CAHPS Surgical Care Survey: https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/surgical/about/whats-available-surgical-care-survey.pdf

Also see Patient Experience Measures from the CAHPS Surgical Care Survey Document 409 obtained by going to: https://www.ahrq.gov/cahps/surveys-guidance/surgical/instructions/get-surg-care-survey-instruct.html

Also, for more information on the calculation of reporting measures, see How to Report Results of the CAHPS Clinician & Group Survey, available at https://cahps.ahrq.gov/surveys-guidance/cg/cgkit/HowtoReportResultsofCGCAHPS080610FINAL.pdf.

S.5. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
This section is used to describe the composite top box score. The composite top box score is the average proportion of respondents who answered the most positive response category across the questions in the composite.

The top box numerators for items within Composite measures 1, 2, 4, 5, and 6 is the number of respondents who answered “Yes, definitely" across the items in each composite. The top box composite score is the average proportion of respondents who answered “Yes, definitely" across the items in the composite.

The top box numerator for items within Composite measure 3 is the number of respondents who answered "Yes" across the items in this composite. The top box composite score is the average proportion of respondents who answered "Yes" across the items in this composite.

The top box numerator for the Measure 7, the Global Rating Item, is the number of respondents who answered 9 or 10 to the Global Rating Item.

EXAMPLE:
Given a composite with four items, where each item has three response options, a practice’s score for that composite is the proportion of responses (excluding missing data) in each response category.

The following steps show how those proportions are calculated:
Step 1 – Calculate the proportion of cases in each response category for the first question:
P11 = Proportion of respondents who answered “yes, definitely”
P12 = Proportion of respondents who answered “yes, somewhat”
P13 = Proportion of respondents who answered “no”

Follow the same steps for the second question:
P21 = Proportion of respondents who answered “yes, definitely”
P22 = Proportion of respondents who answered “yes, somewhat”
P23 = Proportion of respondents who answered “no”

Repeat the same procedure for each of the questions in the composite.

Step 2 – Combine responses from the questions to form the composite.
Calculate the average proportion responding to each category across the questions in the composite. For example, in the "How Well Surgeon Communicates With Patients Before Surgery" composite (four items), the calculations would be as follows:

Measure top box score = proportion who said "yes, definitely" = (P11 + P21 + P31 + P41) / 4

Example results: If P11 = 81% and P21=92% and P31 = 84% and P41 = 95% then the top box score = (81% + 92% + 84% = 95%) / 4 = 88%.

Also see Patient Experience Measures from the CAHPS Surgical Care Survey Document 409 obtained by going to: https://www.ahrq.gov/cahps/surveys-guidance/surgical/instructions/get-surg-care-survey-instruct.html

S.6. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The measure's denominator is the number of survey respondents. The target population for the survey is adult patients (age 18 and over) who had a major surgery as defined by Common Procedural Terminology (CPT) codes (90 day globals) within 3 to 6 months prior to the start of the survey.

Results will typically be compiled over a 12-month period.

For more information on the calculation of reporting measures, see Patient Experience Measures from the CAHPS Surgical Care Survey, available at https://www.ahrq.gov/cahps/surveys-guidance/surgical/instructions/get-surg-care-survey-instruct.html.

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).
For each item in a composite and the provider rating item, the top box denominator is the number of respondents who answered the item per aggregate-level entity (e.g., a surgeon or practice site). For each composite score, the denominator is the number of respondents who answer at least one item within the composite. Composite scores are the average proportion of respondents who gave the highest rating across the items in the composite (as discussed in S.5).

The survey is sampled at the ambulatory care level. However, there are questions that ask about care received at the hospital or surgical care facility.

The major criterion for selecting patients is having surgery, as defined by Medicare 90-day global surgery codes within 3 to 6 months prior to the start of the survey. Since post-surgical care was an important component of the survey, surveys could not be appropriately administered until an adequate time for experiencing post-surgical care (3 months) had passed. The time frame for the surgery was selected to (1) minimize recall bias and (2) ensure ample time was allowed for follow-up care after surgery. The survey is not administered more than 6 months post-surgery because of concerns about recall bias.

Patients have to be adults and non-institutionalized. Included surgeries should be scheduled and not an emergency procedure. This is because an important component of the survey deals with pre-surgical office visits – a topic which would not be relevant for most emergency surgeries.

The Survey's denominator code table lists 90-day global CPT codes for major surgery, representing over 10,000 possible codes across multiple surgical specialties. The Surgical Quality Alliance felt that specifying only Medicare's 90-day global procedure codes would include appropriate procedures while excluding minor procedures that were not intended to be included.

The attached excel file named "Attachment A Main S7 CY2015-90-day-global codes.xlsx" includes the CPT codes that are currently used to identify the S-CAHPS survey's target population of patient with major surgery (i.e., measure denominator).

S.8. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
The following are excluded when constructing the sampling frame:
- Surgical patients whose procedure was greater than 6 months or less than 3 months prior to the start of the survey.
- Surgical patients younger than 18 years old.
- Surgical patients who are institutionalized (put in the care of a specialized institution) or deceased.

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.)*
The following patients would be excluded from the measure's denominator:
- Survey users and vendors should exclude surveys where the respondent reports he or she has not had surgery performed on the date listed by the surgeon named. (First question of survey.)
- Surgical patients that had an emergency surgical procedure since emergency procedures are unlikely to have visits with the surgeon before the surgery.
- Individuals from a household that has already been sampled.
- Respondents who did NOT answer at least one item of the measure are NOT included in the denominator.

Instructions on how to transform raw data from a CAHPS survey into data that the CAHPS Analysis Program can use can be found in Preparing and Analyzing Data from the CAHPS Clinician & Group Surveys available at https://www.cahps.ahrq.gov/surveys-guidance/survey4.0-docs/1035_Preparing_analyzing_data_from_cg.pdf

Survey code specifications --- including how to code an appropriately skipped item, multiple marks or blank items --- can be found in the Instructions for Analyzing Data available at https://cahps.ahrq.gov/surveys-guidance/survey4.0-docs/2015-Instructions-for-Analyzing-Data-from-CAHPS-Surveys.pdf.

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.)*
If survey users want to combine data for reporting from different sampling strata, they will need to create a text file that identifies the strata and indicates which ones are being combined and the identifier of the entity obtained by combining them.

See pages 18-19 of the Instructions for Analyzing Data available at https://cahps.ahrq.gov/surveys-guidance/survey4.0-docs/2015-Instructions-for-Analyzing-Data-from-CAHPS-Surveys.pdf.

S.11. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in measure testing attachment)
Other
If other: Case-mix adjustment

S.12. **Type of score:**
Other (specify):
If other: Top-box Score; case-mix adjusted score

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.)*
Top Box Score Calculation:
1) Target Population: Patients that had a non-emergency surgery within 3 to 6 months prior to the start of the survey.
2) Exclusions = Patients who did not answer at least one item of the composite measures or rating item.
3) Screener items. Example: Patients who answered "No" to the first item indicating that the patient had surgery performed on the date listed by the surgeon named.
4) Top-box scores (percent with highest rating) are computed for each item
5) Top-box scores are averaged across the items within each composite, weighting each item equally.
Note that for users who want to case-mix adjust their scores, case-mix adjustment can be done using the CAHPS macro and the adjustment is made prior to the calculation of the total score.

Case-mix Adjusted Scores
Case-mix adjustment is done via linear regression. The CAHPS Consortium recommends self-reported overall health, age, and education as adjusters. These items are printed in the "About You" section of the survey, questions 38-45.

The steps for user-defined calculations of risk-adjusted scores can be found in Instructions for Analyzing Data from CAHPS® Surveys: Using the CAHPS Analysis Program Version 4.1 available at https://cahps.ahrq.gov/surveys-guidance/survey4.0-docs/2015-Instructions-for-Analyzing-Data-from-CAHPS-Surveys.pdf.

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.
The following sampling guidelines are provided to users as part of the "Fielding the CAHPS® Clinician & Group Surveys: Sampling Guidelines and Protocols" document available at https://cahps.ahrq.gov/surveys-guidance/survey4.0-docs/1033_CG_Fielding_the_Survey.pdf.

Defining the Sample Frame: Eligibility Guidelines
The sample will be drawn from a list of individuals (adults age 18 and older) who had a planned surgery from the surgeon during the specified time interval (see below). The list is called a sample frame.

The source of sample information will vary by survey sponsor. The decision will depend on which organization has the most accurate and complete data. Health plans or purchasers of care may have administrative or billing data to identify individual patients. In some instances, the data to identify individual patients may be found only in the records of medical practices. It may be necessary to pull data from two or more sources in order to have both up-to-date contact information and to be able to connect the procedure to a specific surgeon.

Users should review these guidelines for determining who to include in the sample frame:

- Include only adult patients (age 18 and over) that had a major surgery as defined by CPT codes (90 day globals) within the last 3 to 6 months. This time frame is also known as the look back period.
- The sampling frame is a person-level list and not a procedure list. Therefore, patients should appear only once in the sampling frame regardless of how many surgeries they have had in the look back period. Use the patient's most recent surgery for inclusion in the sampling frame.
- Draw the sample irrespective of reason for surgery and duration of patient-provider relationship, so that the full range of patients is represented.
- Include all patients who meet the sampling criteria even if they are no longer receiving care from the practice.
- To identify the sampling frame, use the anticipated start date of data collection to determine the reference period. For example, if your anticipated start date is September 1, 2011, include all those who have had surgery during February 1 – May 31.
- Allow the sample frame to include multiple individuals from the same household, but the sample you draw should not have more than one adult per household. In other words, the sample that is selected for data collection should be de-duplicated to ensure that only one person per household receives a survey.
- All CAHPS survey items have been designed for the general population. Appropriate screening items are included for items targeted to assess a specific experience. In order to ensure that results are comparable to those produced by other sponsors and vendors, targeted sampling, such as selecting only patients with particular conditions or experiences, is not recommended. Targeted sampling should only be used to supplement the general population sample, if desired.
- To administer the survey, the name of the surgeon must be available, even if you are surveying at the practice level. If the sampling frame does not accurately identify the surgeon who performed the surgery, you may want to select a larger sample to account for errors in connecting health care received to a specific provider.

Recommended Number of Completes

In order to determine the size of the sample, you first need to determine the level of sampling and how many completed surveys are required to obtain usable information at that level. The CAHPS Clinician & Group Survey and the CAHPS Surgical Care Survey can be used to assess care at the individual provider, practice site/clinic, or medical group level. A practice site/clinic is based on a single geographic location. A medical group may contain multiple practice sites/clinics and is defined by a specific list of providers.

- Individual providers: 45 completed surveys per provider. For applications of the survey intended to report or assess performance for individual providers, the Consortium recommends at least 45 completed surveys per provider.
- Practice site: The recommended number of completed surveys is based on the number of surgeons at the site. The site-level sample size recommendations vary by the number of surgeons per location site. The CAHPS Consortium has issued the following recommended sample sizes for collecting CAHPS Clinician and Group (CG-CAHPS) data at the practice site level: 50 completed surveys for 1 provider at the site, 100 surveys for 2 providers at the site, 150 for 3 providers, 175 for 4 to 9 providers, 200 for 10-13 providers, and 250 for 14-19 providers at the site. These recommendations are set to achieve between .70 and .80 site-level reliability for all composite measures. Sample size requirements increase with the number of providers practicing at the site.

The minimum sample size recommendations are based on extensive research conducted on the CAHPS Clinician & Group Survey. These recommendations are based on data regarding the number of completed questionnaires necessary to achieve adequate physician-level reliability for a measure. That is, how many completed surveys does one need to reliably distinguish among different physicians? To answer this question, CAHPS investigators examined data from multiple field trials.

Sample Size Calculation: To have a sufficient number of responses for analysis and reporting, survey users need to select enough individuals to obtain approximately 45 completed questionnaires per physician. Assuming you achieve the recommended response rate of 40 percent, survey users would need to start with a minimum sample size of 113.

More detail and reasoning behind the recommendations can be found in the 2011 document titled "Fielding the CAHPS® Clinician & Group Surveys: Sampling Guidelines and Protocols" available at https://cahps.ahrq.gov/surveys-guidance/survey4.0-docs/1033_CG_Fielding_the_Survey.pdf.

Citation:
Dyer, N., J. S. Sorra, et al. (2012). "Psychometric properties of the Consumer Assessment of Healthcare Providers and Systems (CAHPS) Clinician and Group Adult Visit Survey." Medical care 50 Suppl: S28-34.

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.
The measures are dependent on the standardized implementation of the complete CAHPS Surgical Care Survey. Failure to administer the entire survey will compromise the validity and reliability of the measures.

Developing the Sampling Frame
The sampling frame for the CAHPS Surgical Care Survey consists of patients that have had a non-emergency 90-day global procedure in the last 3-6 months from the date that the survey will be administered. The sampling frame is a patient level file – patients with multiple procedures that meet the criteria may not be included multiple times in the sampling. The actual sample will be de-duplicated to ensure that patients are not provided more than one survey. Users of S-CAHPS should follow the recommended guidelines of the CAHPS Clinician and Group Survey found in the "Fielding the CAHPS® Clinician & Group Surveys: Sampling Guidelines and Protocols" document available at https://cahps.ahrq.gov/surveys-guidance/survey4.0-docs/1033_CG_Fielding_the_Survey.pdf.

Preparing Sample Files for Data Collection
Once the sample has been selected, the vendor assigns a unique identification (ID) number to each sampled person. This unique ID number should not be based on an existing identifier such as a Social Security number or a patient ID number. This number will be used only to track the respondents during data collection.

As previously noted, some sample frames may not include complete and accurate contact information, requiring the combination of information from two (or more) sources – such as administrative records from the plan and contact records from the medical group or clinician office. When information from two sources differs, sponsors and their survey vendors should consult with each other to decide which sources of information are most accurate and should be used. This may be a complex, multistep process that requires time and rigorous quality control. In addition, because the sponsor may be responsible for some elements of this process and the vendor for others, it is important to carefully coordinate this process. The pieces of information that are most critical to the success of data collection are accurate and complete patient [parent/guardian] and provider names and contact information appropriate for the mode of administration (i.e., addresses for mail surveys, telephone number for telephone administration, and e-mail addresses for web-based administration). When you have incomplete address information or reason to believe that this information may be inaccurate, sponsors and/or vendors may be able to use other sources, such as CD-ROM directories, Internet sources, or directory assistance, to clean the sample file.

Data Collection Modes
Each survey sponsor will need to choose the data collection mode that maximizes the response rate at an acceptable cost.

Based on field test results, the CAHPS team recommends the following modes:
- Mail only.
- Telephone only.
- Mixed mode (mail and telephone, email and mail, or email and telephone).

Survey sponsors that employ one of these modes using the recommended protocols can expect to achieve response rates of approximately 40 percent or higher.

Results from the field tests, as well as the experiences of organizations that have fielded similar surveys, indicate that the mail with telephone followup method is most effective; results from survey research literature indicate that followup by telephone often adds 10 to 15 percentage points to the response rate. A sample telephone script for Surgical CAHPS in both English and Spanish is available at https://www.cahps.ahrq.gov/surveys-guidance/survey4.0-docs/sample_tele_script_surgical_care_survey.pdf.

Note that the addition of e-mail administration (i..e, notification for web-based surveys) as a type of mixed-mode data collection is a recommendation since last endorsement. The CAHPS Consortium recommends including an option to conduct a mixed mode survey that would have two e-mail reminders and a follow up by mail or telephone to all who are surveyed. The follow up to the entire sampling frame is necessary to get a representative sample from a practice that is not based just on e-mail alone. The Consortium does not recommend a mailed hard copy letter with a link to a web survey.

Other Modes

The CAHPS team recognizes that many organizations may already be or are interested in conducting patient surveys using different modes of survey administration and has conducted preliminary testing of other modes, specifically in-office distribution and interactive voice response (IVR, also known as telephone audio computer-assisted self-interviewing, or T-ACASI). Further study is required before either of these modes can be recommended.

A study of in-office distribution found that the survey results were not comparable to those collected with recommended modes. The investigators observed incomplete distribution rates, lower response rates, and declining distribution rates. Finally, there were significant mode-physician interaction effects, which suggests that data cannot be pooled then adjusted to account for the differences. Because the implications of using these modes are not yet fully known, they should be used with caution. If a sponsor uses one of these modes to collect data, the ability to compare survey results across sponsors may be limited.

The Consortium's support of the use of multiple modes of survey administration is intended to minimize disruption to organization's current survey processes. Thus, organizations that conduct mail surveys can continue using mail, those that conduct telephone surveys can continue using telephone, and likewise for other modes. More detail on the protocols per mode, can be found in the full fielding guide. The Consortium is currently collecting and analyzing data in order to assess the need for procedures for data adjustments as a function of each survey mode and to enable the team to develop such procedures.

Tracking Returned Questionnaires

Most vendors have established methods for tracking the sample. You should also set up a system to track the returned surveys by the unique ID number that is assigned to each respondent in the sample. This ID number should be placed on every questionnaire that is mailed and/or on the call record of each telephone case.

To maintain respondent confidentiality, the tracking system should not contain any of the survey responses. The survey responses should be entered in a separate data file linked to the sample file by the unique ID number. (This system will generate the weekly progress reports that sponsors and vendors should review closely.)

Each respondent in the tracking system should be assigned a survey result code that indicates whether the respondent completed and returned the questionnaire, completed the telephone interview, was ineligible to participate in the study, could not be located, is deceased, or refused to respond. The tracking system should also include the date the survey was returned or the telephone interview completed. The interim result code

reflects the status of the case during the different rounds of data collection, and the final result code reflects the status at the end of data collection. These result codes are used to calculate response rates.

Citation:
Drake KM, Hargraves JL, Lloyd S, Gallagher PM, Cleary PD. The Effect of Response Scale, Administration Mode, and Format on Responses to the CAHPS Clinician and Group Survey. Health Serv Res. 2014 Aug;49(4):1387-99. doi: 10.1111/1475-6773.12160

McInnes DK, Brown JA, Hays RD, Gallagher P, Ralston JD, Hugh M, Kanter M, Serrato CA, Cosenza C, Halamka J, Ding L, Cleary PD. (2012) Development and evaluation of CAHPS questions to assess the impact of health information technology on patient experiences with ambulatory care. Med Care. 2012 Nov;50 Suppl:S11-9.

S.17. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.18.
Instrument-Based Data

S.18. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data are collected.)*
IF instrument-based, identify the specific instrument(s) and standard methods, modes, and languages of administration.
Consumer Assessment of Health Providers and Systems (CAHPS) Surgical Care Survey Version 2.0

Available in English at https://cahps.ahrq.gov/surveys-guidance/survey4.0-docs/surgical_eng.pdf

Available in Spanish at https://cahps.ahrq.gov/surveys-guidance/survey4.0-docs/surgical_span.pdf

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)*
Available at measure-specific web page URL identified in S.1

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

S.21. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Inpatient/Hospital, Other, Outpatient Services
If other: Hospital Outpatient Surgery Center, Ambulatory Surgical Centers

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
Attachments_1741_11062017-636461679792662086.zip,nqf_testing_attachment_7.1_SCAHPS_FINAL2.docx

2.1 For maintenance of endorsement
Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.
No

2.2 For maintenance of endorsement

Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.
No

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

Measure Title: CAHPS Surgical Care Survey 2.0
Date of Submission: 11/11/2017
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 [16] differences in performance;

OR

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

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

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

Notes

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

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores

are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions.
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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

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

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

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

*Metrics presented throughout are derived from analysis of the CAHPS Surgical Care Survey, Adult Version 2.0 core items.

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*).

The main field test was conducted in the summer of 2008 and included 33 surgical practices across 9 surgical specialties. There were four mailings sent to patients in two waves, or batches. The response rate was 49%. One site only had 5 responses and is excluded from this dataset for testing. Forty-six surveys that answered "NO" to the first question "Did you have surgery from this surgeon on specified date?" were excluded. The final dataset has 2,719 survey results from 32 practices across nine specialty types.

1.3. What are the dates of the data used in testing? February 2007 to April 2008

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

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*)

The measured entity is referred to as a "practice site." Surgeons in a single practice site should share administrative and clinical support staff. Practice site level survey results are calculated across the respondents within a specific site.

The data for CAHPS Surgical Care (S-CAHPS) Version 2.0 includes survey response data for surgeries that occurred between February 2007 – April 2008. Data are being pooled from investigators that have utilized the S-CAHPS survey since 2008, such that a meta-analysis can be conducted. We anticipate submitting results to the NQF during a future measure update period.

The data in the testing and analysis includes 32 practice sites and 2,719 respondents across nine specialty types. The average number of respondents per site is 85 (standard deviation = 75) ranging from 19 individuals per site to 298 individuals per site. The median number of respondents per site is 47.5. Three of the respondents took the survey using the Spanish version. The total number of unique surgeons in the dataset was 72 resulting in an average of 2.25 surgeons per practice. Table 1.5a below shows the distribution of states within the dataset. Table 1.5b below shows the distribution of specialty type within the dataset.

Tables 1.5a Distribution of States in CAHPS Surgical Care Version 2.0 Dataset

Count of Complete Records & Practice Sites within State		
STATE	Total Complete Records	Total Practices
ARIZONA	218	1
CALIFORNIA	49	1
FLORIDA	82	1
GEORGIA	45	1
ILLINOIS	61	2
INDIANA	117	1
KANSAS	50	1
MARYLAND	67	2
MASSACHUSETTS	122	2
MICHIGAN	683	5
MISSISSIPPI	35	1
NEW JERSEY	64	2
NEW YORK	19	1
NORTH CAROLINA	217	1
OHIO	517	4
PENNSYLVANIA	143	1
TEXAS	186	4
UTAH	44	1
Total	*2,719*	*32*

Tables 1.5b Distribution of Specialty Type in CAHPS Surgical Care Version 2.0 Dataset

Count of Complete Records & Practice Sites within Specialty Type		
Specialty Type	Total Complete Records Within Specialty	Total Practices
Colon and Rectal	326	5
General Surgery	396	6
Ophthalmology	405	4
Orthopaedic	415	2
Otolaryngology	294	4
Thoracic	235	3
Urology	534	5
Vascular	114	3
Total	*2,719*	*32*

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*)

In each practice site, surveys are completed by non-institutionalized adult patients that had a major scheduled surgery as defined by CPT codes (90 day globals) within 3 to 6 months prior to the start of that practice's survey. The attached excel file named "Table Testing 1.6 – S-CAHPS CPT Codes.xlsx" includes the CPT codes that were used during the main field test to identify the S-CAHPS survey's target population of patients with major surgery (i.e., measure denominator).

Although multiple individuals in a single clinician group may be on the sampling frame, the final sample contained only one respondent per household. Where a duplicate household was sampled, it was discarded and replaced by another random draw from the frame. The fielding guidelines provide additional advice on drawing representative samples for multiple products and simultaneous sampling of adult and child enrollees. Fielding guidelines are available at:

https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/cg/survey3.0/fielding-the-survey-cg30-2033.pdf. Table 1.6 shows descriptive characteristics of the individuals surveyed included in our analyses. Practice sites had adult patients that were predominantly white (86%) and 55 years or older (71%). Eighty-three percent of the surveys were collected through the mail.

Table 1.6. Descriptive Characteristics for S-CAHPS 2.0 Dataset (32 Practices, 2,719 Respondents)

2008 Field Test	
	N and (Percent of Total)
GENDER	
Female	1355 (49.8%)
Male	1261 (46.4%)
Missing	101 (3.8%)
RACE	
White	2346 (86.3%)
Black or African American	156 (5.7%)
Asian	25 (0.9%)
Hawaiian or Pacific Islander	2 (0.1%)

American Indian or Alaskan	10 (0.4%)
Other	50 (1.8%)
Multi-racial	29 (1.1%)
Missing	101 (3.7%)
AGE CATEGORY	
18-24 Years	38 (1.4%)
25-34 Years	100 (3.7%)
35-44 Years	207 (7.6%)
45-54 Years	362 (13.3%)
55-64 Years	600 (22.1%)
65-74 Years	701 (25.8%)
75 Years or older	623 (22.9%)
Missing	88 (3.2%)
HOW SURVEY COLLECTED	
Mail	2253 (83%)
Web	466 (17%)

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.

Not applicable; same data used for each aspect of testing below.

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

Items collected in the survey that could be considered social risk factors are Education and Race/Ethnicity. They are not analyzed.

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*)

Cronbach's alpha is the most common measure of internal consistency reliability and commonly used for surveys with scale-type questions. We calculated a Cronbach's alpha for each composite to assess the extent to which respondents consistently answered the items, with a reliability of at least 0.70 considered acceptable (Nunnally and Bernstein, 1994). For composites with more than two items, we show the impact on Cronbach's alpha of deleting one of the items from the composite.

Given that individual responses are nested within practice sites, we measure site reliability on multi-item composite top-box scores and global one-item top-box scores, which partition within- and between-site variance. For this test, we used case-mix adjusted scores generated by following the specifications from the CAHPS Analysis Program Version 4.1 for adjusting on patients' general health rating, education and age. Similar to internal consistency reliability (i.e., Cronbach's alpha), values of 0.70 and higher are considered acceptable for site reliability (Nunnally and Bernstein, 1994; CAHPS Analysis Program, 2017; Instructions for Analyzing Data from CAHPS Surveys, 2017) .

The individual site reliability was calculated using the following formula and then averaged across the practice sites to derive measure-level reliability:

$$\text{Reliability}_g = \frac{\Sigma_B}{\Sigma_B + \frac{\Sigma_W}{N_g}},$$

where $\sum_B$ refers to the between-group variance; $\sum_W$ refers to the within-group variance, and Ng is the sample size for site g (Raudenbush and Bryk, 2002).

Citations:
Nunnally JC, Bernstein IH. Psychometric Theory. New York: McGraw Hill; 1994.
Raudenbush SW, Bryk AS. Hierarchical Linear Models. 2nd ed. Thousand Oaks, CA: Sage; 2002.

The CAHPS Analysis SAS Program Version 4.1c is downloadable from.
https://www.ahrq.gov/cahps/surveys-guidance/surgical/instructions/get-surg-care-survey-instruct.html

Instructions for Analyzing Data from CAHPS® Surveys is accessible at
https://www.ahrq.gov/cahps/surveys-guidance/helpful-resources/analysis/index.html

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*)

Table 2a2.3a. Cronbach's Alpha Individual-level Reliability Coefficient for S-CAHPS Version 2.0 Dataset. 2008 (32 practice sites, 2,719 Respondents)

Measure and Items	Standardized Cronbach's Alpha	Cronbach's Alpha if Item Deleted*
7. Information to Help You Prepare for Surgery (2 items)	0.74	
8. How Well Surgeon Communicates with Patients Before Surgery	**0.82**	
Listened carefully to them (Q9).		0.74
Spent enough time with them (Q10).		0.77
Encouraged them to ask questions (Q11).		0.78
Showed respect for what they had to say (Q12).		0.81
9. Surgeon Attentiveness on Day of Surgery (2 items)	**0.66**	
10. **Information to Help You Recover from Surgery - from Surgeon or another Health Provider from the office.**	**0.84**	
Explained what to expect during recovery (Q26).		0.77
Gave easy to understand instructions about what to do during recovery (Q28).		0.79
Warned them about symptoms that need immediate attention (Q27).		0.77
Made sure they were physically comfortable or had enough pain relief after leaving the hospital or surgery facility (Q29).		0.85
11. How Well Surgeon Communicates With Patients After Surgery	**0.86**	
Listened carefully to them (Q31).		0.81
Spent enough time with them (Q32).		0.82
Encouraged them to ask questions (Q33).		0.81
Showed respect for what they had to say (Q34)		0.86
12. Helpful, Courteous, and Respectful Staff at Surgeon's Office (2 items)	**0.85**	

* Not applicable if less than three items per composite.

Table 2a2.3c. Top Box Practice Site-Level Reliability Statistics for Surgical CAHPS, Version 2.0 (32 Practices)

Measures	Average # of Respondents per site	Site-Level Reliability
Information to Help You Prepare for Surgery	85	0.52
How Well Surgeon Communicates with Patients Before Surgery	76	0.68
Surgeon Attentiveness on Day of Surgery	83	0.50
Information to Help You Recover From Surgery - from Surgeon or another Health Provider from the office.	85	0.71
How Well Surgeon Communicates With Patients After Surgery	72	0.48
Helpful, Courteous, and Respectful Staff at Surgeon's Office	84	0.71
One-item Global Rating of surgeon	82	0.60

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?*)

Table 2a2.3a shows the Cronbach's alpha for each composite. These individual-level test results show that each composite of the S-CAHPS survey has an acceptable level of reliability. For items within a composite consisting of 3 or more items, the Cronbach's alpha if the item were deleted is provided to determine if there was room for improving the alpha by dropping an item. The results do not suggest removal of any questions.

Table 2a2.3b shows the mean number of respondents per practice site and practice site-level reliability statistics for the surveys. The dataset excludes sites with less than 10 surveys. The majority of the composites and global ratings exhibit satisfactory site-level reliability with two measures above 0.70 and two between 0.6 and 0.7. Information to Prepare for Surgery and Post-Surgery Communication exhibit lower reliability at the site level than they did at the individual level. It should be noted that the site-level reliability analysis was conducted on field test data consisting of a relatively small selection of surgeon practices which may have led to reduced variability between sites.

2b1. VALIDITY TESTING

2b1.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☒ **Empirical validity testing**

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

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

At the practice site level, we examined the relationships between each composite's top box score and the top box score for the global measure of "How would you rate your surgeon?" using Spearman rank-order correlations to determine the validity of the composite measures. For example, the composite measuring how well patients' surgeon communicates with before surgery is expected to be strongly related to the patients' overall rating of their surgeon. Finding such a relationship supports interpretation of the composite as a valid measure of patient experience with a surgeon and that surgeon's office. Specifically, Table 2b1.3b shows validity testing results of the instrument tested across all responses. Tables 2b1.3a and 2b1.3c reflect practice-level scoring validity.

We also examined Spearman rank-order correlations among the composites to assess the extent to which they measure different constructs. As measures of patient experience, we expected the composites to be correlated. However, very high intercorrelations indicate that the composites may not be unique enough to be considered separate measures.

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

Table 2b1.3a. Practice Site-Level Correlation of Composites and Global Rating for S-CAHPS Version 2.0 Sample, 2008 (32 practice sites)

Rating of Surgeon	Info Pre-Surgery	Communicate Pre-Surgery	Attentiveness on Day of Surgery	Recovery Information	Communicate Post-Surgery	Office Staff Service
GLOBAL	0.48	0.59	0.02*	0.68	0.50	0.49

P < 0.01 for all values. *P>0.05.

Note: Values are Spearman rank-order correlations on top box scores.

Table 2b1.3b. Individual-Level Correlation of Composites and Global Rating for S-CAHPS Version 2.0 Sample, 2008 (2,719 respondents)

Rating of Surgeon	Info Pre-Surgery	Communicate Pre-Surgery	Attentiveness on Day of Surgery	Recovery Information	Communicate Post-Surgery	Office Staff Service
GLOBAL	0.42	0.49	0.24	0.46	0.47	0.34

P < 0.001 for all values.
Note: Values are Spearman rank-order correlations on top box scores.

Table 2b1.3c. Site-Level Top-Box Composite Intercorrelations for CAHPS Surgical Care Version 2.0 Sample, 2008 (32 Practices)

Composites	Info Pre-Surgery	Communicate Pre_Surgery	Attentiveness on Day of Surgery	Recovery Information	Communicate Post-Surgery	Office Staff Service
Info Pre-Surgery	1	0.58***	0.30	0.69***	0.64***	0.54**
Communicate Pre_Surgery		1	0.49**	0.71***	0.80***	0.31
Attentiveness on Day of Surgery			1	0.33	0.46**	-0.07
Recovery Information				1	0.64***	0.55**
Communicate Post-Surgery					1	0.36
Office Staff Service						1

***p<.001, **p<.01, *p<.05
Note: Values are Spearman rank-order correlations.

2b1.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e.*, *what do the results mean and what are the norms for the test conducted?*)

The survey results show that each of the six composites are related to the global rating scale at the individual level and five of the six composites are related to the global rating at the practice level.. The two Communication composites and the Recovery Information composite have the strongest relationship with the global rating of surgeon.

Although the composites should be correlated with each other, as they all measure aspects of patient experience, inter-correlations > 0.80 indicate that the composites may not be unique enough to be considered separate measures (O'Brien, 2007). In general, relationships among the composites met our expectations.

Citation:
O'Brien RM. A caution regarding rules of thumb for variance inflation factors. Qual Quant. 2007;41:673–690.)

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*)

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

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.

2b3.1. What method of controlling for differences in case mix is used?
☒ **No risk adjustment or stratification**
☐ **Statistical risk model with __risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☒ **Other,** Optional risk adjustment*
*The CAHPS analysis program allows users to select adjustment factors of their choosing.

The S-CAHPS measures, like the CG-CAHPS measures (NQF #0005), can be reported as either top-box scores or as case-mix adjusted scores depending on the user's purposes. Note that case-mix adjustment is possible in certain situations and is optional.

Top Box Score Calculation:
1) Target Population: Patients that had at least one visit during the past 12-months
2) Exclusions = Patients who did not answer at least one item of the composite measures or rating item.

3) Screener items. Example: Patients who answered "No" to the first item indicating that they did not receive care from the
provider entity in the last 12 months
4) Top-box scores (percent with highest rating) are computed for each item
5) Top-box scores are averaged across the items within each composite, weighting each item equally.

Case-mix Adjusted Scores:
The steps for user-defined calculations of risk-adjusted scores can be found in Instructions for Analyzing Data from CAHPS® Surveys: Using the CAHPS Analysis Program Version 4.1 available at https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/helpful-resources/analysis/2015-instructions-for-analyzing-data.pdf

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.
Not applicable.

2b3.2. If an outcome or resource use component measure is <u>not risk adjusted or stratified</u>, provide <u>rationale and analyses</u> to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.
We are not putting forth a risk-adjusted measure. Adjustment is not a necessity because CAHPS surveys are not a clinical outcome or resource use. However, the S-CAHPS measures, like the CG-CAHPS measures (NQF #0005), does give users the <u>option</u> to case-mix adjust scores depending on their specific purposes.

Note: We did use case-mix adjusted scores for reliability testing, adjusting on patients' general health rating, education and age using the CAHPS Macro.

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

This section is used to describe the rationale for case-mix adjustment that is recommended to users and the case-mix adjustment that is performed in parallel to the methods used for all CAHPS measures. Specifically, case-mix adjustment methods for the Clinician and Group CAHPS (CG-CAHPS) are applicable to the S-CAHPS.

Case-mix adjustment takes into account respondent characteristics, such as age or educational attainment, which may affect the reports and ratings of care but are unrelated to differences in care quality. In other words, it is important to account for patient characteristics that are not under the control of the group but are related to the patient's experiences and survey responses. For example, several studies have found that younger and more educated patients provide less

positive evaluations of health care (Elliott et al. 2001; Zaslavsky et al. 2001). CAHPS data can also be adjusted for other factors such as survey administration mode. Without an adjustment, differences in CAHPS scores between entities could be due to case-mix differences rather than true differences in quality. CAHPS survey results can be case-mix adjusted by users of the CAHPS analysis program.

Each user's project team must determine if it is appropriate to adjust its data to account for case-mix differences and if so, which adjusters to use. Specifically, for the S-CAHPS, available adjusters include overall health, overall mental or emotional health, age, sex, number of previous surgeries, educational attainment, Hispanic ethnicity and race. The project team must also decide whether or not to impute missing data for the adjusters at each adjuster's entity-level mean. The document "Instructions for Analyzing Data from CAHPS Surveys" dated April 2012 (available at: https://www.ahrq.gov/cahps/surveys-guidance/cg/index.html) contains instructions for coding these variables and for including them in analyses using the CAHPS Analysis Program in SAS. Notwithstanding, the CAHPS Consortium only recommends that users use the CAHPS macro to adjust their survey data for respondent age, education, and general health status if they are comparing scores across practices that may differ in the characteristics of patients. The only characteristics that may be considered social that is collected in the survey are Education level and "Race" and "Hispanic or Latino origin/descent."

One of the methodological issues associated with comparison across practice sites and/or individual clinicians is the need to adjust appropriately for differences in case-mix, particularly in situations where there are differences among sites in patient characteristics. Therefore, current CG-CAHPS guidance suggests health status, age, and education as possible case-mix adjusters for users. These patient characteristics are not under the control of the provider. Studies have found that patient health status, education and age are predictors of CAHPS' scores (Kim et al, 2005; O'Malley, Zaslavsky et al. 2005; Elliott, Zaslavsky et al. 2009). A recent study by Drake and colleagues (2014) found that telephone respondents gave more positive responses than mail respondents.

Citations:

CAHPS Clinician & Group Survey. Content last reviewed August 2017. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/cahps/surveys-guidance/cg/index.html

Drake KM, Hargraves JL, Lloyd S, Gallagher PM, Cleary PD. (2014) The Effect of Response Scale, Administration Mode, and Format on Responses to the CAHPS Clinician and Group Survey. Health Serv Res. Jan 29. doi: 10.1111/1475-6773.12160. [Epub ahead of print]

Elliott MN, Zaslavsky AM, Goldstein E, Lehrman W, Hambarsoomians K, Beckett MK, Giordano L. (2009) Effects of survey mode, patient mix, and nonresponse on CAHPS hospital survey scores. *Health Serv Res.* Apr;44(2 Pt 1):501-18. doi: 10.1111/j.1475-6773.2008.00914.x.

Elliott MN, Swartz R, Adams J, Spritzer KL, Hays RD. Case-mix adjustment of the National CAHPS benchmarking data 1.0: a violation of model assumptions? *Health Serv Res*. 2001 Jul; 36(3):555-73.

Kim M, Zaslavsky AM, Cleary PD. (2005) Adjusting Pediatric Consumer Assessment of Health Plans Study (CAHPS) Scores to Ensure Fair Comparison of Health Plan Performances. *Med Care*. Jan;43(1):44-52.

O'Malley AJ, Zaslavsky AM, Elliott MN, Zaborski L, Cleary PD. (2005) Case-mix adjustment of the CAHPS Hospital Survey. *Health Serv Res.* Dec;40(6 Pt 2):2162-81.

Zaslavsky AM, Zaborski LB, Ding L, Shaul JA, Cioffi MJ, Cleary PD. Adjusting Performance Measures to Ensure Equitable Plan Comparisons. *Health Care Financ Rev*. 2001 Spring; 22(3):109-126.

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)**

See 2b3.3a above.

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

See above.

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.**

See above.

2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

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

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**
Not applicable.

2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):
Not applicable.

2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
Not applicable.

2b3.9. Results of Risk Stratification Analysis:
Not applicable.

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)
Not applicable.

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

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

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

Statistical significance is determined based on case-mix adjusted mean scores for individual items, composites and global ratings, and is used to determine whether a practice site mean is statistically different from the mean results for all practice sites. The statistical test used is a t-test of means, with $p<.05$ used as the criterion for determining significance. When there are large sample sizes, relatively small differences between practice sites may be statistically significant.

The CAHPS analysis program allows users to perform testing for both statistical and substantive significance. Users specify the size of the difference required for substantive significance in terms of an absolute size difference or a specified fraction of the distance between the entity and the nearer of upper and lower bounds on the measure.

More information about Surgical CAHPS can be found at https://www.ahrq.gov/cahps/surveys-guidance/surgical/about/index.html.

More information about how to analyze practice level CAHPS scores can be found on this website: https://www.ahrq.gov/cahps/surveys-guidance/helpful-resources/analysis/index.html.

For guidance in adapting these analysis instructions, users can contact the CAHPS User Network by e-mail (cahps1@ahrq.gov) or telephone (1-800-492-9261).

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*)

Table 2b5.2. Variability of S-CAHPS Measures' Scores	Descriptive at the Site Level (32 Practices)				
Measure	**Mean Top Box**	**SD**	**Min-Max Range**	**Min**	**Max**
Information to Help you Prepare for Surgery (2 items)	90%	0.05	19	79%	98%
How Well Surgeon Communicates Before Surgery (4 items)	85%	0.07	31	67%	98%
Attentiveness on Day of Surgery (2 items)	81%	0.12	55	42%	97%
Information to help you Recover from Surgery (4 items)	82%	0.07	36	64%	100%
How Well Surgeon Communicates After Surgery (4 items)	84%	0.06	24	73%	97%
Helpful, Courteous, and Respectful Staff (2 items)	87%	0.08	42	58%	100%
One-item Global Rating of Surgeon	86%	0.07	28	70%	98%

2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

The field test data suggests large differences across the 32 practices in surgical performance across many of the topic areas addressed by the survey. For example, with the Recovery Information measure, there was a 36-percentage point difference between the lowest scoring practice (64%) and the highest scoring practice (100%). Each of these are meaningfully different than the mean of 82%.

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

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

Section not applicable – one set of specifications.

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*)

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*)

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*)

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*)

Survey Non-Response

For CAHPS, the recommended or target response rate is 40 percent. Survey users that follow the recommended protocols for sampling and data collection, including followup with nonrespondents, typically report response rates of 40 percent or higher. It is also important to begin with as accurate a sampling frame as possible. This section presents the guidance provided to users for calculating response rates (AHRQ, 2011).

The response rate is the total number of completed questionnaires divided by the total number of individuals selected. For CAHPS analyses and reports, this rate is adjusted as shown in the following formula:

$$\frac{\text{Number of completed returned questionnaires}}{\text{Total number of respondents selected} - (\text{deceased} + \text{ineligible})}$$

In calculating the response rate, users do not exclude respondents who refused, whom they were unable to reach because of bad addresses or phone numbers, or who were unable to complete the questionnaire because of language barriers or because they were institutionalized or incompetent.

Listed below is an explanation of the categories included and excluded in the response rate calculation:

Numerator Inclusions:

- **Completed questionnaires.** A questionnaire is considered complete if responses are available for 50 percent of key CAHPS items and at least one composite item or rating item.

Denominator Inclusions:

The total number in the denominator should include the following:

- **Refusals.** The individual (or parent or guardian of the sampled child) refused in writing or by phone to participate.
- **Nonresponse.** The respondent is presumed to be eligible but did not complete the survey for some reason (never responded, was unavailable at the time of the survey, was ill or incapable, had a language barrier, etc.).
- **Bad addresses/phone numbers.** In either case, the sampled individual is presumed to be eligible but was never located.

Denominator Exclusions:

- **Deceased**. In some cases, a household or family member may inform you of the death of the sampled individual.
- **Ineligible.** The sampled individual did not have a scheduled surgery from surgeon during the 3 to 6 months prior to the survey.

Users are provided the following advice for improving response rates:

- Improve initial contact rates by making sure that addresses and phone numbers are current and accurate (e.g., identify sources of up-to-date sample information, run a sample file through a national change-of-address database, send a sample to a phone number look-up vendor).
- Use all available tracking methods (e.g., directory assistance, CD-ROM directories, free or subscription-based Internet database services and directories).
- Improve contact rates after data collection has begun (e.g., increase maximum number of calls, ensure that calls take place at different day and evening times over a period of days, mail second reminders, use experienced and well-trained interviewers).
- Consider using a mixed-mode protocol involving both a mail and telephone data collection procedure. In field tests, the combined approach was more likely to achieve a desired response rate than did either mode alone.

- Train interviewers on how to deal with gatekeepers (someone such as a relative who stands between the interviewer and the respondent, making it difficult or impossible to complete the interview).
- Train interviewers on refusal aversion/conversion techniques.

Item Non-Response

The method used to construct CAHPS scores, discussed in sections S4-S11 of the main NQF submission form, maximizes the use of available data by averaging available individual-level responses in construction of an overall score for the practice site. For each individual item, the top box score is percentage of respondents who answered the most positive response for the item. At the site-level, the top box composite score is the average of those percentages across the items. Most often, as a result of this averaging methodology, the percent of missing items at the site level is very small (i.e., < 1%).

We provide the percentage of cases with missing values at the item level below.

Citation:
AHRQ "Fielding the CAHPS® Clinician & Group Surveys. Sampling Guidelines and Protocols. "Document No. 2033. Updated 6/12/2017. Accessible at: https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/cg/survey3.0/fielding-the-survey-cg30-2033.pdf.

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*)

Table 2b7.2a. Surgical CAHPS Version 2.0: Item-level Percent Missing (2,719 respondents)

Survey Item	% Missing at Individual Level	Item is after a Skip Question
1. Information to Help You Prepare for Surgery	**0%**	
Gave all the information needed before surgery (Q3).	1%	Yes
Gave easy to understand instructions (Q4).	1%	Yes
2. How Well Surgeon Communicates with Patients Before Surgery	**11%**	
Listened carefully to them (Q9).	2%	Yes
Spent enough time with them (Q10).	2%	Yes
Encouraged them to ask questions (Q11).	2%	Yes

Survey Item	% Missing at Individual Level	Item is after a Skip Question
Showed respect for what they had to say (Q12.)	2%	Yes
3. Attentiveness on Day of Surgeon	**1%**	
Visited them before surgery (Q15.)	3%	No
Discussed the outcome of their surgery (Q17.)	2%	No
4. Information to Help You Recover From Surgery - from Surgeon or another Health Provider from the office.	**1%**	
Explained what to expect during recovery (Q26).	2%	No
Gave easy to understand instructions about what to do during recovery (Q28).	2%	No
Warned them about symptoms that need immediate attention (Q27).	2%	No
Made sure they were physically comfortable or had enough pain relief after leaving the hospital or surgery facility (Q29).	2%	No
5. How Well Surgeon Communicates With Patients After Surgery	**16%**	
Listened carefully to them (Q36).	4%	Yes
Spent enough time with them (Q37).	4%	Yes
Encouraged them to ask questions (Q38).	4%	Yes
Showed respect for what they had to say (Q34)	3%	yes
6. Helpful, Courteous, and Respectful Staff at Surgeon's Office	**2%**	
Clerks and receptionists were helpful to patients (Q36).	2%	No
Clerks and receptionists treat patients with courtesy and respect (Q37)	3%	No
7. Single-Item Global Rating of Surgeon (Q34)	**4%**	**No**

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*)

Table 2b7.2a. shows the percentage of cases with missing data on each item and the last column denotes if the item followed a screen/skip item. Less than five percent of cases are missing response scores on the individual items, which suggests that our item-level results are likely not biased by systematic missing data due to item nonresponse.

It is important to note that ten of the items in S-CAHPS were not applicable for patients who were directed to skip the question due to their response to a screen/skip question. For example, the second question of the survey asks, "Before your surgery, how many office visits did you have with this surgeon?" If the taker checks NONE, then they are guided to skip the two questions about information to prepare for surgery (composite measure #1) and the four questions about pre-surgery communication (composite measure #2).

The post- surgical communication items also follow a screen question. The screener item asks: "After your surgery, did you talk with this surgeon by phone or visit the surgeon at his or her office?" If the response is "NO" then the survey taker skips the four questions about post-surgical communication with the surgeon.

Screening questions can result in a high percentage of missing due to appropriate skips. Survey item screeners have been found to reduce measurement error by ensuring that respondents who are not 'qualified' to answer a question are screened out instead of providing invalid responses (Rodriguez et al., 2009)

Citation:
Rodriguez HP, Glahn Tv, Li A, Rogers WH, Safran DG. The effect of item screeners on the quality of patient survey data: a randomized experiment of ambulatory care experience measures. *Patient*. 2009 Jun 1;2(2):135-41.

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.**

Other

If other: Collected by survey of surgeon´s patients

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**.
No 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).
Though multiple modes of data collection, as well as mixed mode types of administration, are possible and have been tested, CAHPS surveys are primarily delivered via mail. Electronic databases are created after mailed surveys are returned. Traditionally, the rationale for not using electronic sources more broadly is that mail and telephone are the best ways to obtain representative samples of patients based on the contact information that is available for sampling and data collection. However, email has been added as a mixed mode strategy for surgeon groups with reliable email addresses for all of their population. This is important as the uptake of electronic devices continues to grow.

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.
The CAHPS Surgical Care survey was developed in consultation with patients using the most sophisticated, valid, and reliable methodologies available in survey and measurement science. During development of the survey (both English and Spanish versions), The American Institutes for Research (AIR) conducted 30 cognitive interviews in 2008 with 20 English- and 10 Spanish-speaking patients who had undergone a non-emergency, 90-day global surgery within the past 12 months. Each cognitive interview lasted approximately 2 hours. A trained cognitive interviewer administered the draft survey and conducted each interview using a semi-structured protocol.

During development, Round 1 cognitive testing revealed a variety of problems in the initial survey drafts. Changes were implemented and tested in Round 2, resulting in an improved survey. Examples of the types of revisions made to the draft Surgical CAHPS surveys include the following:
- A chronological ordering of sections was deemed the most intuitive order for respondents.

- Some of the section headings in the survey were simplified. For example, “Your Pre-Operative Care From This Surgeon” was changed to “Before Your Surgery.”
- The introductory sentences at the beginning of sections were eliminated. In some cases, they caused confusion and, with the simplified headings, they were no longer necessary.
- A new screener question for patients with no follow-up office visits was added.
- A definition of “staff at this surgeon’s office” was added to one question to ensure that respondents include all staff in their responses (receptionists, clerks, nurses, etc.).
- A response option of “Don’t know” was added to several items.
- Wording changes were made to various items to simplify and clarify comprehension.
- Several items were eliminated because of limited response variation across rounds or significant respondent interpretation problems.
- Spanish language translation changes were made to some items.

The cognitive testing led to many improvements in the draft Surgical CAHPS survey during development and improved the quality of the field testing and, ultimately, the NQF-endorsed S-CAHPS survey.

An important distinction when comparing patient experience versus patient satisfaction is that patient experience measures aspects of care that are actionable for surgical quality improvement. And because the survey instrument, protocol, analysis, and reporting are standardized, surgeons can benchmark and compare their performance with that of their peers within the same practice or outside of their practice. Surgeons may customize the S-CAHPS survey by adding survey items that are specific to their patients and practice. However, the core survey must be used in its entirety in order to be comparable with other S-CAHPS data. The S-CAHPS survey may be used in both the inpatient and outpatient setting.

A barrier identified as a result of operational use for the measure has been patient confidentiality when attempting to administer the survey using email. Because of guidance published within the Federal Register as part of the 2013 HIPAA rules, initial emails sent to patients cannot contain any Protected Health Information unless sent via encrypted email. To comply, these initial emails must be sent in a generic manner that may have lead patients to believe they are spam, resulting in low response rates. Despite the desperate need for patient input to evaluate quality, US security laws are in effect that prevent the patient perspective from being easily incorporated into national quality improvement strategies.

More importantly, the S-CAHPS survey was designed by surgical care professionals (i.e., surgeons and anesthesiologists) for surgical patients. The survey measures aspects of patients´ surgical care that are important to them. The S-CAHPS provides more actionable data specific to surgery as compared to other available patient experience and CAHPS measures, such as the Hospital CAHPS, which focuses primarily on the facility rather than the surgeon. Studies have been published demonstrating their use for quality improvement (see Evidence attachment for more details).

Sources:
Levine, Burling, Huberman, and Hurtado from the American Institutes for Research. (2008) Surgical CAHPS: Cognitive Testing of the English and Spanish Survey Instruments. Report prepared for the American College of Surgeons. (Report was Attachment D of S-CAHPS 2011 NQF Submission).

Sage, J. (2013) What Surgeons Should Know About Using S-CAHPS. Bulletin of the American College of Surgeons. Accessible at http://bulletin.facs.org/2013/08/using-s-cahps/

Liu JB, Pusic AL, Temple LK, Ko CY. Patient-reported outcomes in surgery: Listening to patients improves quality of care. Bull Am Coll Surg. 2017; 102(3): 19-23.

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

The CAHPS Surgical Care Survey is available to users free of charge. In addition to the survey instrument, users can access comprehensive fielding, analysis, and reporting guides as well as SAS programming code that performs analysis and significance testing. All of these tools are available at: https://www.cahps.ahrq.gov/surveys-guidance/cg/instructions/index.html. Requirements for using the CAHPS name on an instrument, include:
- All core items must be present on the user's questionnaire
- No changes to core item wording are permitted
- Instruments must not omit any of the survey items related to respondent characteristics.

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 Professional Certification or Recognition Program	Payment Program https://qpp.cms.gov/ CMS Quality Payment Program Quality Improvement (external benchmarking to organizations) American College of Surgeons´ National Surgical Quality Improvement Program https://www.facs.org/quality-programs/acs-nsqip CMS Core Quality Measures Orthopedics Set https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/Core-Measures.html Quality Improvement (Internal to the specific organization) Surgical Practices N/A

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

A. Payment Program
1. Name of Program and Sponsor: CMS Quality Payment Program (QPP) Merit-based Incentive Payment System (MIPS) and as part of an Advanced Alternative Payment Model (A-APMs)
2. Purpose: National pay-for-reporting program.
3. Current and Planned Use:
3.1 Current use: QPP Improvement Activity: Participation in a CAHPS survey, such as the S-CAHPS, is a high-weighted patient safety and practice assessment activity within the Improvement Activities component of MIPS for performance years 2017 and 2018. MIPS eligible clinicians who administer the S-CAHPS survey can attest to the completion of this activity and earn points toward their MIPS Final Score for the Improvement Activity component of MIPS.
3.2 Planned use:
3.2.1 2017 MIPS Quality Component: Pending CMS approval for the 2018 MIPS program, participants in the American College of Surgeons (ACS) Surgeon Specific Registry (SSR) will have the option to report on components of the S-CAHPS as part of the Phases of Surgical Care Measures for the MIPS Qualified Clinical Data Registry (QCDR) reporting mechanism. The Phases of Surgical Care Measure set is an episode-based measure framework which follows the five phases of surgical care: preoperative, perioperative, intraoperative, postoperative, and postdischarge. The measure set is composed of surgical outcome measures, high-value process measures, and the appropriate S-CAHPS measures which follow up on key processes within the measure set.
3.2.2 Advanced Alternative Payment Model: The ACS is currently working with CMS on an Advanced Alternative Payment Model that we anticipate may soon be tested by Centers for Medicare and Medicaid Innovation (CMMI). This model incorporates a novel quality measurement framework which measures care around the patient for a given episode and incorporates patient reported experience and patient reported outcomes. As currently proposed, the surgical episodes include measures included in the S- CAHPS survey.
4. Geographic area and number and percentage of accountable entities and patients included: National Program. Percentage of entities and patients included will be determined in 2018 or 2019, depending on the CMS program.
5. Level of measurement and setting: Level of measurement will vary based on QPP program:
- Improvement Activities in MIPS: provider-level or group level;
- Quality Component of MIPS: provider-level;
- Advanced Alternative Payment Model: will be at either the provider level or at the level of the APM entity.

B. Quality Improvement
1. Name of Program and Sponsor: American College of Surgeons´ National Surgical Quality Improvement Program
2. Purpose: National Surgical Quality Improvement
3. Current Use: S-CAHPS is currently used by the ACS NSQIP to measure PROs alongside its clinical data in a pilot. Complementing the ACS NSQIP with PROs represents an opportunity to improve those outcomes that matter most to patients. None of the currently measured clinical outcomes give any insight as to whether the outcome aligned with the patients´ views and goals. Currently, more than 50 hospitals in the United States and internationally participate in this PRO pilot, which was launched in October 2017. All patients accrued into the ACS NSQIP are asked to complete the S-CAHPS survey as part of their follow-up. Data collection is ongoing, and we expect to provide annual measure updates to the NQF.
4. Geographic area and number and percentage of accountable entities and patients included: the ACS NSQIP currently includes more than 700 hospitals in the USA and internationally, representing more than 990,000 operations accrued into the registry annually.
5. Level of measurement and setting: hospitals in the ACS NSQIP

C. Quality Improvement
1. Name of Program and Sponsor: CMS Core Quality Measures Orthopedics Set
2. Purpose: The Core Quality Measure Collaborative, led by the America's Health Insurance Plans (AHIP) and its member plans' Chief Medical Officers, leaders from CMS and the National Quality Forum (NQF), as well as

national physician organizations, employers and consumers, worked hard to reach consensus on core performance measures. Through the use of a multi-stakeholder process, the Collaborative promotes alignment and harmonization of measure use and collection across payers in both the public and private sectors. For more information, visit https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/Core-Measures.html.
3.Current use: The S-CAHPS survey is included in the Orthopedics Core measure set, which is available here: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/Downloads/Orthopedic-Measures.pdf
4. Geographic area and number and percentage of accountable entities and patients included: national; number of patients and accountable entities unknown
5. Level of measurement and setting: provider-level

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?*)
There has long been support by multi-stakeholder groups that the S-CAHPS survey should have been included in the Physician Compare and PQRS programs as a stand-alone measure in addition to the CG-CAHPS because the CG-CAHPS does not accurately reflect care provided to the surgical patient. The support for the S-CAHPS has been acknowledged by CMS as well as supported by the NQF Measures Applications Partnership (MAP). The S-CAHPS survey´s initial NQF endorsement is also a testament to this.

For CY 2014-2016 proposed regulations, CMS received public comments to the PQRS program supporting the inclusion of the S-CAHPS as a stand-alone measure. Comments from The American College of Surgeons stated that the CG-CAHPS survey would not accurately reflect the care provided by single or multispecialty surgical or anesthesia groups. ACS also noted that S-CAHPS has been tested by the same standards as CG-CAHPS and follows the same collection mechanism as CG-CAHPS. In the 2014 Medicare Physician Fee Schedule final rule, several commenters opposed the publication of CG-CAHPS measures citing that the measures are not relevant to their particular specialty. They requested that CMS allow physicians the flexibility to select the survey instruments and patient satisfaction measures most appropriate for their practices, and many of the commenters recommended CMS use S-CAHPS as an optional patient experience of care measure. CMS responded that the Agency understands that CG-CAHPS is not the most applicable CAHPS survey for all specialties and service settings represented by groups on Physician Compare. Therefore, they explained that the Agency will evaluate the feasibility of including additional CAHPS surveys, such as S-CAHPS, on the site in the future.

In the 2015 Medicare Physician Fee Schedule final rule, CMS agreed that the S-CAHPS survey would be more relevant to a surgical group practice compared to the CG-CAHPS and noted that the majority of commenters supported the use of S-CAHPS in the PQRS program. However, CMS did not accept and finalize the measure. CMS explained “due to the cost and time it would take to find vendors to collect S-CAHPS data, it is not technically feasible to implement the reporting of the S-CAHPS survey measures for the 2017 or 2018 PQRS payment adjustments.” They also note that Qualified Clinical Data Registries (QCDRs) will have the option to administer the S-CAHPS as a non-PQRS measure for the 2017 or 2018 PQRS payment adjustments (Medicare Program; Revisions to Payment Policies Under the Physician Fee Schedule, Clinical Laboratory Fee Schedule, Access to Identifiable Data for the Center for Medicare and Medicaid Innovation Models & Other Revisions to Part B for CY 2015; Final Rule, 79 Federal Register 219 November 13, 2014, page 67795).

Also in the 2015 Medicare Physician Fee Schedule final rule, several commenters noted the limitations of CAHPS for PQRS measures for some health care professionals and supported adding other types of patient experience data to Physician Compare, including the Surgical CAHPS and experience data collected via other sources. CMS agreed that Surgical CAHPS data is useful to consumers and explained that the Agency is exploring how it can incorporate this information into Physician Compare (Medicare Program; Revisions to Payment Policies Under the Physician Fee Schedule, Clinical Laboratory Fee Schedule, Access to Identifiable Data for the Center for

Medicare and Medicaid Innovation Models & Other Revisions to Part B for CY 2015; Final Rule, 79 Federal Register 219 November 13, 2014, page 67777).

Similar comments were made by stakeholders in 2016 Medicare Physician Fee Schedule final rule, and similarly CMS explained that they understand that not all measures equally apply to all types of professionals included in Physician Compare, however, they believe the CAHPS for PQRS measures (ie. CG-CAHPS) apply to a large majority of professionals on the Physician Compare site (Medicare Program; Revisions to Payment Policies Under the Physician Fee Schedule and Other Revisions to Part B for CY2016, 80 Federal Register 220 November 16, 2015, page 71129).

Lastly, the NQF's MAP has recommended the inclusion of S-CAHPS for two consecutive years. In the MAP Pre-Rulemaking Report: 2013 Recommendations on Measures Under Consideration by HHS report the NQF MAP recommended the inclusion of S-CAHPS in the PQRS program. In the MAP 2014 Recommendations on Measures for More Than 20 Federal Programs, the MAP recommended the inclusion of the S-CAHPS measure in PQRS, Meaningful Use, the Value-based Payment Modifier, and Physician Compare. These reports can be found here:

MAP Pre-Rulemaking Report: 2013 Recommendations on Measures Under Consideration by HHS. Final report available at: http://www.qualityforum.org/Publications/2013/02/MAP_Pre-Rulemaking_Report_-_February_2013.aspx

MAP Pre-Rulemaking Report: 2014 Recommendations on Measures for More than 20 Federal Programs. Final report available at: http://www.qualityforum.org/Publications/2014/01/MAP_Pre-Rulemaking_Report__2014_Recommendations_on_Measures_for_More_than_20_Federal_Programs.aspx

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.*)
A. Payment Program and Public Reporting
1. Name of program and sponsor: CMS Physician Compare
2. Purpose: Physician Compare is a national public reporting website which provides consumers with quality of care information to make informed health care decisions. Physician Compare is also intended to encourage clinicians to improve the quality of care they provide to their patients and create incentives to maximize performance.
3. Planned Use: Pending CMS approval for the 2018 MIPS program, participants in the American College of Surgeons (ACS) Surgeon Specific Registry (SSR) will have the option to report on components of the S-CAHPS as part of the Phases of Surgical Care Measures for the MIPS Qualified Clinical Data Registry (QCDR) reporting mechanism. The Phases of Surgical Care Measures are an episode-based measure framework which follows the five phases of surgical care: preoperative, perioperative, intraoperative, postoperative, and postdischarge. The measure set is composed of high-value process measures, appropriate S-CAHPS measures which correlates to the process measures, and surgical outcome measures. As part of this measure set, the S-CAHPS measures will be reported on the Physician Compare downloadable database if they meet CMS statistical public reporting standards which require that measures be valid, reliable, and accurate.
4. Geographic area and number and percentage of accountable entities and patients included: National Program. Percentage of entities and patients included will be determined in 2018.

B. Payment Program. Please see above in 4a1.1 for the use of S-CAHPS in the Quality Payment Program.

C. Professional Certification or Recognition Program
1. Name of program and sponsor: American Board of Surgery (ABS) Maintenance of Certification® (MOC) Part IV (Practice Assessment Resources) URL: http://www.absurgery.org/default.jsp?exam-mocpa

2. Purpose: The American Board of Surgery Maintenance of Certification Part IV can be satisfied by ongoing participation in a local, regional or national outcomes registry or quality assessment program. Components of the S-CAHPS will be available via the American College of Surgeons (ACS) National Surgical Quality Improvement Program (NSQIP), and the ACS Surgeon Specific Registry (SSR), starting in 2018. Both the SSR and NSQIP registries can meet the MOC Part IV MOC component, and are listed on the ABS website as examples of good programs for meeting Part 4, http://www.absurgery.org/default.jsp?exam-mocpa.
3. Geographic area and number and percentage of accountable entities and patients included: National Program. Percentage of entities and patients included will be determined in 2018.

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.
Too early to broadly determine. However, as discussed in the Evidence form, several peer-reviewed studies have demonstrated the use of the S-CAHPS survey for local quality improvement purposes with good success.

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.**

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.

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

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

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.**

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.
There is no repository for S-CAHPS data. Trend data is not available. Surgeons use S-CAHPS to identify their strengths and weaknesses and to help develop strategies for improving patients' experiences with care delivered surrounding a surgery.

The objective of the S-CAHPS survey is to measure aspects of surgical quality that are important to consumers and for which consumers are the best source of information. At the same time, consistent with the survey, the measures aim to directly benefit a variety of users, including: patients, practice groups, health plans, insurers, and specialty boards. Each group has a need for information regarding the quality of surgeons and surgical care.

1. Patients will use information from the measures to help make better and more informed choices about their surgical care.
2. Practices, health plans, and insurers will use the measure results for quality improvement initiatives and incentives.
3. Specialty boards may use the measure results for maintenance of certification purposes.

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.**

No unexpected findings have been uncovered.

4b2.2. **Please explain any unexpected benefits from implementation 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.

Yes

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

0005 : CAHPS Clinician & Group Surveys (CG-CAHPS)-Adult, Child
0007 : NCQA Supplemental items for CAHPS® 4.0 Adult Questionnaire (CAHPS 4.0H)

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?
Yes

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
In December 2011, the Agency for Healthcare Research and Quality released the 2.0 version of the Surgical Care Survey. This update keeps the Surgical Care Survey consistent with the Clinician & Group Survey, which was updated in October 2011. Because the changes that led to the 2.0 designation do not represent a significant digression from the 1.0 version of this survey, the shift from 1.0 to 2.0 does not affect the ability of survey users to assess trends in performance.

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.)
The Surgical Quality Alliance (SQA; https://www.facs.org/advocacy/quality/surgical-quality-alliance) reviewed the current CAHPS® Clinician and Group Survey and identified critical gaps in content and approach related to the assessment of surgical care. For example, some critical gaps in the survey include informed consent, shared decision making, anesthesia care, and post-operative instructions and access, all of which are issues consumers find to be very important in surgery. Therefore, the SQA felt that the development of a surgical patient experience survey fit well into the mission to improve the quality of healthcare delivered to surgical patients.

Surgeries have high resource use, and poor quality can have serious consequences for patients, including death. Hospital stays that involve operating room (OR) procedures are more costly, on average, than stays that do not involve OR procedures. Therefore, improving the quality of surgical care is of paramount importance to patients and the healthcare system alike. To measure the quality of surgical care in the U.S., a survey that measures the patient-reported outcomes associated with care provided by single- or multispecialty surgical or anesthesia groups is necessary. As shown in the attached support letter from 2012, there is much support from the surgical specialty societies for NQF endorsement and use of the S-CAHPS survey. (See attached "Attachment C Main 1c3 Letter to NQF 2012_surgeryspecialty_1741.pdf".)

The Centers for Medicare and Medicaid Services (CMS) acknowledged the importance of allowing for the administration of S-CAHPS reporting in addition to the Clinician and Group CAHPS for Physician Quality Reporting System (PQRS). The ACS has emphasized to CMS that it is critical that the measures included in the quality-tiering composite are valid, reliable, and applicable to all health care professionals, to avert the unintended consequence of misclassifying a physician's care and unfairly affecting payment. Accordingly, the S-CAHPS measures have been submitted for consideration for the Merit-based Incentive Payment System of the Quality Payment Program.

Though the main emphasis of the survey is to measure aspects of surgical quality that are important to consumers and for which consumers are the best source of information, the development of measures specific to surgical care benefits many users, including: patients, practice groups, health plans, insurers, and specialty boards. Each group has a need for information regarding the quality of surgeons and surgical care. Patients will use information from the survey to help make better and more informed choices about their surgical care. Practices, health plans, and insurers will use the surgical survey results for quality improvement initiatives and incentives. Specialty boards may use the survey's measure results for maintenance of certification purposes.

CAHPS surveys were originally developed to meet the need of consumers for usable, relevant information on quality of care from the patient's perspective. But they also play an important role as a quality improvement (QI) tool for health care organizations, which can use the standardized data to identify relative strengths and weaknesses in their performance, determine where they need to improve and track their progress over time.

Cultivating the use of CAHPS surveys for QI purposes is one of the key objectives for the CAHPS grants. The CAHPS Improvement Guide is a comprehensive resource for health plans, medical groups, and other providers seeking to improve their performance in the domains of quality measured by CAHPS surveys. The guide may be used to:
- Cultivate an environment that encourages and sustains quality improvement;
- Analyze the results of CAHPS surveys to identify strengths and weaknesses; and
- Develop strategies for improving performance.

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:** Attachment_B_Main_1b2_S-CAHPS_score_Tables-636461813747558109.xlsx

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** American College of Surgeons, Division of Advocacy and Health Policy
Co.2 **Point of Contact:** Jill, Sage, jsage@facs.org, 202-672-1507-
Co.3 **Measure Developer if different from Measure Steward:** American College of Surgeons, Division of Advocacy and Health Policy
Co.4 **Point of Contact:** Jill, Sage, jsage@facs.org, 202-672-1507-

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.
Surgical Quality Alliance (SQA) Technical Advisory Panel (TAP):
1. Priscilla Arnold, MD FACS - American Society of Cataract and Refractive Surgery
2. Larissa Temple, MD FACS - American Society of Colon and Rectal Surgeons
3. Laura King, MD – American Academy of Ophthalmology
4. Robin Brody, MD – American Academy of Otolaryngology – Head and Neck Surgery
5. Lee Eisenberg, MD FACS - American Academy of Otolaryngology – Head and Neck Surgery
6. Rahul Shah, MD FACS - American Academy of Otolaryngology – Head and Neck Surgery
7. Robert Haralson, MD FACS – American Academy of Orthopaedic Surgeons
8. Frank Opelka, MD FACS – American College of Surgeons
9. Antony Sidawy, MD FACS - Society for Vascular Surgery
10. Loren Hiratzka, MD FACS – Society for Thoracic Surgery
11. James Hicks, MD – American Society of Anesthesiologists
12. Andrea Pusic, MD FACS - American Society of Plastic Surgeons

13. Sharon Merrick, Staff - American Society of Anesthesiologists
14. Chip Amoe, Staff - American Society of Anesthesiologists
15. Jason Byrd, Staff - American Society of Anesthesiologists
16. Cathy Cohen, Staff - American Academy of Ophthalmology
17. Cherie McNett, Staff - American Academy of Ophthalmology
18. Kristine Schulz, Staff - American Academy of Otolaryngology - Head and Neck Surgery
19. Stephanie Jones, Staff - American Academy of Otolaryngology - Head and Neck Surgery
20. Beth Kosiak, Staff - American Urological Association
21. Suzanne Pope, Staff - American Urological Association
22. Nancey McCann, Staff - American Society of Cataract and Refractive Surgery
23. DeLaine Schmitz, Staff - American Society of Plastic Surgeons
24. Guy Beaumont, Staff – American College of Osteopathic Surgeons
25. Cynthia Shewan, Staff – Society for Thoracic Surgery
26. Elizabeth Hoy, Staff – American College of Surgeons
27. Caitlin Burley, Staff - American College of Surgeons
28. Valerie Oster, Staff - American College of Surgeons
29. Andrea Burling – American Institutes for Research
30. Roger Levine - American Institutes for Research
31. Samantha Sheridan - Westat
32. John Rauch – Westat

The Surgical quality Alliance (SQA) of the American College of Surgeons includes a Technical Advisory Panel (TAP) which provided continuous support to the development of the Surgical care Survey from the literature review to final testing. The TAP included 32 members from various surgery and anesthesiology specialty societies, each having technical expertise in the topic of surgical care.

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2010
Ad.3 **Month and Year of most recent revision:** 06, 2012
Ad.4 **What is your frequency for review/update of this measure?** Periodic, as needed
Ad.5 **When is the next scheduled review/update for this measure?** 11, 2018

Ad.6 **Copyright statement:** "CAHPS" is a registered trademark of the Agency for Healthcare Research and Quality (AHRQ) but CAHPS surveys themselves are not copyrighted and in the public domain.
Ad.7 **Disclaimers:** None.

Ad.8 **Additional Information/Comments:** List of Attachments:
Attachments located in zip file named: Attachments 1741 11062017.zip
Attachment A Main S9 CY2015-90-day-global codes.xlsx
Attachment B Main 1b2 S-CAHPS_score_Tables.xlsx
Attachment C Main 1c3 Letter to NQF 2012_surgeryspecialty_1741.pdf
Attachment D Main 1c5 Surgical CAHPS Focus Groups_2nd Round_2010.pdf

NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 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

New Measure -- Preliminary Analysis

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?*

o *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.)?*
- *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**

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**:*

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

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

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.**

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

11. 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.

12. 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)

13. 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)

14. 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)

15. 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)

16. **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)

17. Was other reliability testing reported?

☒Yes (go to Question #8)

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

18. 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)

19. 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)

20. **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

17. 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]

18. 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)

19. 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)

20. 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.

21. 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.

22. 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

23. 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.

24. 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)

25. **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.

26. 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)

27. 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)

28. **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

29. Was other validity testing reported?

☐Yes (go to Question #14)

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

30. Was validity testing conducted with <u>patient-level data elements</u>?

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 <u>no</u>
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.

31. 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)

32. **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).

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number

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: Click here to enter composite measure #/ title

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: Click here to name the health outcome

☐Patient-reported outcome (PRO): Click here to name the 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*): Click here to name the intermediate outcome

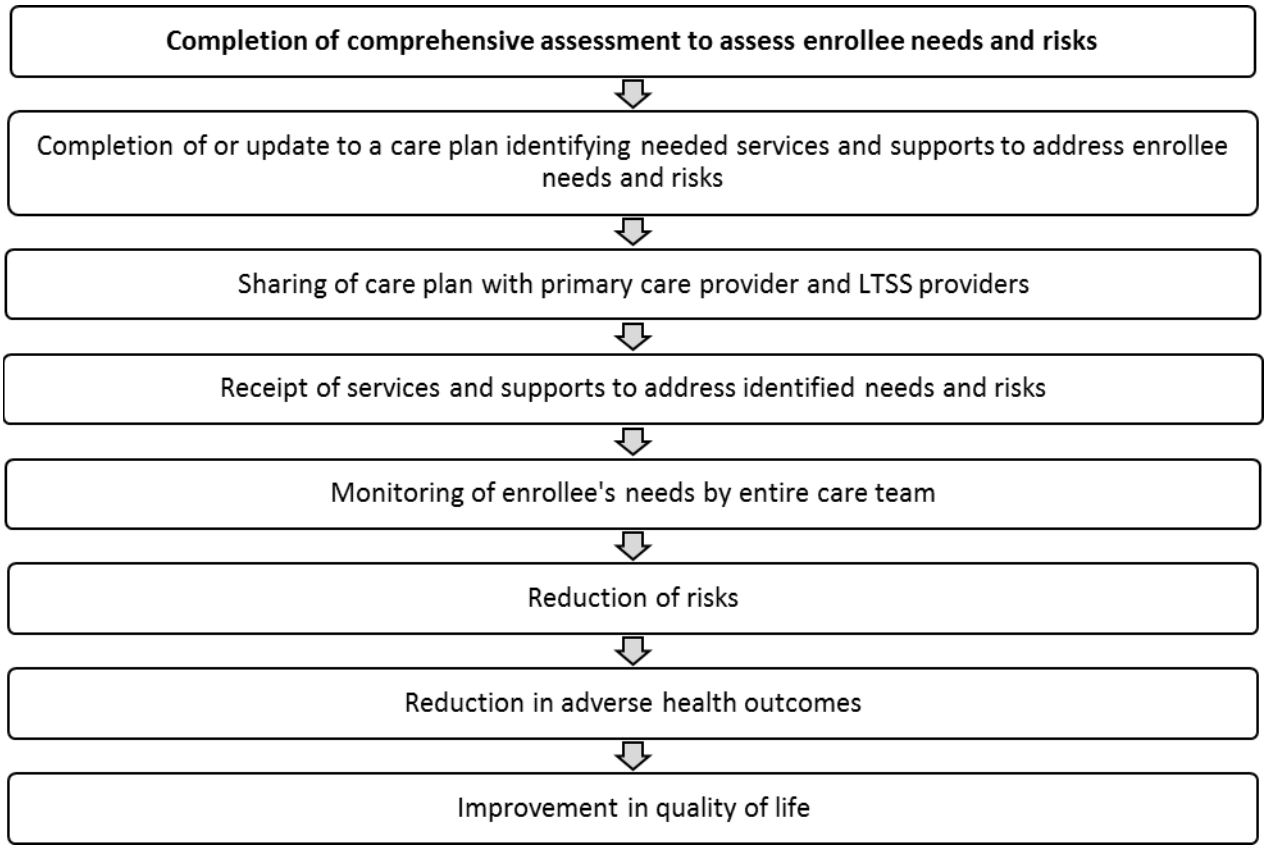
☒ Process: : This measure assesses the extent to which Managed Long Term Services and Support (MLTSS) enrollees receive a comprehensive assessment for provision of long term services and supports.

☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

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.

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, Performance Gap, Priority – 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.

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. **For maintenance of endorsement,** 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.*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The 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)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
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

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
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** [16] **differences in performance**;

OR

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

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

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

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-

item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

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

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions.
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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

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

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

Measure Specified to Use Data From: **(*must be consistent with data sources entered in S.17*)**	**Measure Tested with Data From:**
☒ abstracted from paper record	☒ abstracted from paper record
☐ claims	☐ claims
☐ registry	☐ registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒ other: 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 <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

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

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

Characteristics	Percentage of enrollees in the testing sample (n=715)
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

Characteristics	Percentage of enrollees in the testing sample (n=715)
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:

$$\hat{K} = \frac{\rho_a - \rho_e}{1 - \rho_e}$$

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

Data element	Kappa statistic	Interpretation
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** Click here to enter number of factors **risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

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

NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 3324
Corresponding Measures:
De.2. **Measure Title:** Long Term Services and Supports (LTSS) Comprehensive Care Plan 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 care plan in a specified timeframe that includes documentation of core domains. The measure has two rates:

Rate 1: Percent of MLTSS plan enrollees with a comprehensive LTSS care plan including seven (7) core elements documented within 120 days of enrollment or at least annually.

Rate 2: Percent of MLTSS plan enrollees with a comprehensive LTSS care plan including seven (7) core elements and at least four (4) supplemental elements documented within 120 days of enrollment or at least annually.
1b.1. **Developer Rationale:** Care plans based on comprehensive assessments serve as the foundation for providing high quality and well-coordinated care. While almost all MLTSS plans require care plans be developed for their members, there is little data on the scope and content of care plans among the MLTSS enrollee population. This measure would address the lack of standardization by assessing the percentage of Medicaid MLTSS enrollees who have a comprehensive care plan developed that includes specific core and supplemental elements. A standardized measure of comprehensive care planning 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 care plan 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 care plan domains. Through review and voting the TEP identified 20 elements as critical for developing LTSS care plans. However, through our field test, we determined that requiring all 20 elements was too stringent and would not result in meaningful performance rates (i.e., five out of five plans had a 0% performance rate when all 20 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 care plan 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 care plan completed within 120 days of enrollment for new enrollees, with seven (7) core elements documented, or
- A comprehensive LTSS care plan completed at least once during the measurement year for all other enrollees (established enrollees) with all seven (7) core elements documented.

Rate 2: MLTSS plan enrollees who had either of the following:
- A comprehensive LTSS care plan completed within 120 days of enrollment for new enrollees, with seven (7) core elements and at least four (4) supplemental elements documented, or
- A comprehensive LTSS care plan completed at least once during the measurement year for all other enrollees (established enrollees) with seven (7) core elements and at least four (4) supplemental elements documented.

Note: Initial care plan should be developed within 120 days of enrollment (allows for 90 days to complete assessment and 30 days to complete care plan), and updated annually thereafter.
Denominator Statement: Medicaid MLTSS 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) and 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 development of a comprehensive care plan or who refused to participate in development of a comprehensive care plan.

Measure Type: Process
Data Source: Management Data, Other, Paper Medical Records
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:
3319: LTSS Comprehensive Assessment 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

New Measure -- Preliminary Analysis

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 care plan and the outcome of improvement in quality of life.
- There is no systematic review of studies of care planning in MLTSS programs. The developer conducted a targeted literature review to gather evidence in support of the measure.
- An environmental scan conducted under a previous CMS contract (Contract No. HHSM-500-2010-00026I/HHSM-500-T0011) highlights the lack of standardization in how care plans are defined across populations using LTSS (included as Appendix 1: Environmental Scan of Assessment and Care Planning Measures).
- Developer did not provide quality of evidence specific to comprehensive care plan and update, however, they did provide evidence for goal-setting, and use of a single document for care planning:
  - Well-developed care plans are associated with positive outcomes (Rich, et al. 2012)
  - Use of structured goal-setting approaches to self-management of chronic conditions has been shown to significantly improve HbA1c levels (Naik, et al., 2012)
  - Individual goal-setting has also been linked to better outcomes and improvements in health and functioning in a variety of other populations, such as those with dementia, coronary heart disease, stroke, end stage renal disease, and rehabilitation needs (Clare, et al., 2015; Janssen, et al., 2013; Warner, et al., 2015; Kauric-Klein 2012; Muller, et al., 2011)
  - Published study found that using a care plan as a single document for sharing information across multiple settings demonstrated clinically-significant improvement in depression and improved 10-year cardiovascular risk, exercise rates, and referrals to exercise programs and mental-health clinicians (Morgan, et al., 2015)

Exception to evidence

N/A

Questions for the Committee:

- *Is the evidence provided for this measure exhaustive? Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*
- *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 based on systematic review (Box 3) → Empirical evidence presented, but not systematically reviewed (Box 7) → empirical evidence includes all studies in this body of evidence (Box 8) → Submitted evidence indicated high certainty that the benefits clearly outweigh undesirable effects (Box 9) → Moderate

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

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 provides performance gap rationale of "care plans based on comprehensive assessments serve as the foundation for providing high quality and well-coordinated care", and concludes that "a standardized measure of comprehensive care planning will allow for apples-to-apples comparisons of MLTSS plans across states".
- TEP identified 20 elements as critical for developing LTSS care plans - however field testing determined that requiring all 20 elements was too stringent and would not result in meaningful performance rates (i.e., five out of five plans had a 0% performance rate when all 20 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 care plan two rates were developed for the measure:
    - "Core" requirements that set a minimum baseline of performance; and
    - "Supplemental" requirements that demonstrate more thorough and comprehensive performance.
- Testing data were collected from five MLTSS health plans who participated in testing these measures with enrollees representing at least two or more of the major LTSS sub-populations:
    - On average, fewer than 1 percent of enrollees across MLTSS plans had documentation of a care plan developed in the specified timeframe, including the specified seven core elements (Rate 1). The range in performance from 0 percent to 2.4 percent indicates there is room for improvement.

Rate	Rate 1- Seven (7) core elements documented	Rate 2- Seven (7) core elements documented and four (4) supplemental elements documented
Mean	0.6	0.6
Standard Deviation	1.1	1.1
Minimum	0.0	0.0
Maximum	2.4	2.4

Disparities

- Developer was unable to find research on potential disparities in the use of care plans among the MLTSS enrollee population.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *If 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 cited provides moderate support for the link between the process measured (comprehensive care planning for MLTSS enrollees) and outcomes. The quantity and quality of this evidence, while not formally assessed, does seem in line with the current state of evidence around care coordination, which is still developing.
The evidence linking use of a care plan to improved depression and cardiovascular outcomes is directly applicable, while much of the remaining evidence is somewhat tangential. The strongest evidence cited links goal setting to improved outcomes, but this addresses just one of the 7 core elements required for documentation.
As shown by the logic model provided, the comprehensive care planning process is linked to patient outcomes through multiple steps. This measure focuses on one of the most up-stream steps, meaning it is more distant from the outcomes of interest.

1b. Performance Gap

Comments

** Yes, there is clear evidence of a performance gap for both rates defined under this measure. The testing sample seems adequate for assessing performance gap (e.g., 715 enrollees from 5 health plans). However, given the very low rates of performance and the lack of robust evidence linking each of the 7 core care plan elements to patient outcomes, I wonder whether the observed performance gap represents a true quality deficit, or whether it just reflects that the 7 core elements included in this measure are not the most important for performance. I am particularly struck by the developer's finding that 2/3 of enrollees had a documented care plan, but still failed this measure because the

plan did not include all 7 core elements. This certainly speaks to the wide variation in care planning, but I think its debatable whether it supports this specific set of 7 core elements (with or without supplemental elements) as the most important to measure.

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: NQF Staff

Evaluation of Reliability and Validity (and composite construction, if applicable):
Link A

Additional Information regarding Scientific Acceptability Evaluation (*if needed*): N/A

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

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

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

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

2a1. Reliability- Specifications

Comments

**I'd like to know more about who reviews care plans to determine whether all core and/or supplemental elements are present. What guidance is provided to ensure this process is done reliably from one reviewer to another? Do reviewers have any inherent conflict of interest (ie, are held accountable for performance score, so have COI in rating their own quality).

2a2. Reliability- Testing

Comments

**Two of the 7 core elements stand out to me as having poor reliability (functional needs and cognitive needs). This raises questions for me about overall measure reliability. Overall ICC is very strong which somewhat addresses this concern, but I still wonder about how consistently data will be collected across sites.

2b1. Validity—Testing
2b4-7. Threats to Validity
2b4. Meaningful Differences

Comments

** The finding that 2 of the 5 participating health plans had zero rates, for me, calls into question the validity of the measure. As previously noted, this could be interpreted as evidence of an enormous performance gap, or as evidence that the measure is not focused on important elements of care planning (ie, is not reflective of quality). In the absence of strong evidence linking the care planning process to outcomes, I am skeptical of the argument that these are valid measures without further testing. Findings from the TEP show only tepid support, further calling into question validity. Overall, I would rate validity as low based on the testing data provided.

2b2-3. Other Threats to Validity
2b2. Exclusions
2b3. Risk Adjustment

Comments

**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 provides 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: As the developer notes, many of the data elements are not in structured fields and there is a lack of supporting standards.

Committee pre-evaluation comments

Criteria 3: Feasibility

3. Feasibility

Comments

**I am not familiar with LTSS data, so can't speak to how readily available data elements required for this measure may be. However, the developers statements about reliance on review of electronic and paper records suggest at least moderate measurement burden. Did the developers obtain any information from the 5 health plans that tested this measure about the time and effort required to review data elements and calculate the measure?
I do think that measurement in this area would drive standardization of documentation practices, so over time measurement burden may decrease.
Given the very low rate of performance on Rate 1, there doesn't seem to be a strong case for using Rate 2, at least in the near term. I do agree that over time as performance improves on Rate 1, using Rate 2 could help drive further improvements.

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact/improvement and unintended consequences

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

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

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

Current uses of the measure
Publicly reported? ☐ **Yes** ☒ **No**

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

Accountability program details

- This 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
N/A
Feedback on the measure by those being measured or others N/A

Additional Feedback: N/A

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*

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

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
N/A - 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 [unexpected findings]

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

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
*What did the health plans participating in the measure testing have to say about usability? Are they planning to continue using the measure?
It is encouraging to see that this measure was included in a set of MLTSS measures recommended for use.

Criterion 5: Related and Competing Measures

Related or competing measures

Related measures include:
- 3319 LTSS Comprehensive Assessment 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

N/A

Public and member comments

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

Comments:

Morgan Buchko, Meridian Health Plan

It was acknowledged that the measure is not currently standardized and may come from free text (3b.2.). This would be difficult for health plans to report on this measure until the standardization occurs. We believe it would be helpful to standardize what is required in the plan of care across all ICOs.

- **Zero NQF members have submitted a support/non-support choice.**

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: 3324
Measure Title: Long Term Services and Supports (LTSS) Comprehensive Care Plan and Update

RELIABILITY

21. 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.

22. 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)

23. 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)

24. 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)

25. 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 assessed using ICC

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

26. **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)

27. Was other reliability testing reported?

☐Yes (go to Question #8)

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

28. 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)

29. 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)

30. **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)

There was a mix of inter-rater reliability of data elements. 15 of the 23 key data elements show a kappa statistic in the "moderate" or "high" reliability and the remaining 8 data elements indicate low reliability.

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

33. 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]

34. 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)

35. 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)

36. 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.

37. 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)

38. 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

39. 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 we inconclusive (e.g. neither validated, nor invalidated the measure). These may be viewed in the testing submission.

40. 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)

41. **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	6	7
Disagree	3	3
Strongly Disagree	3	1
No response	0	1
Total % Agree	54%	62%

Feedback from the TEP on Systematic review of face validity included:

- 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.
- 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 in care plans - including non-medical needs.

42. 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)

43. 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)

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

44. **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

45. Was other validity testing reported?

☐Yes (go to Question #14)

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

46. Was validity testing conducted with <u>patient-level data elements</u>?

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 <u>no</u>
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.

47. 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)

48. **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).

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number

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

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 11/8/2017

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: Click here to name the health outcome

☐Patient-reported outcome (PRO): Click here to name the 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*): Click here to name the intermediate outcome

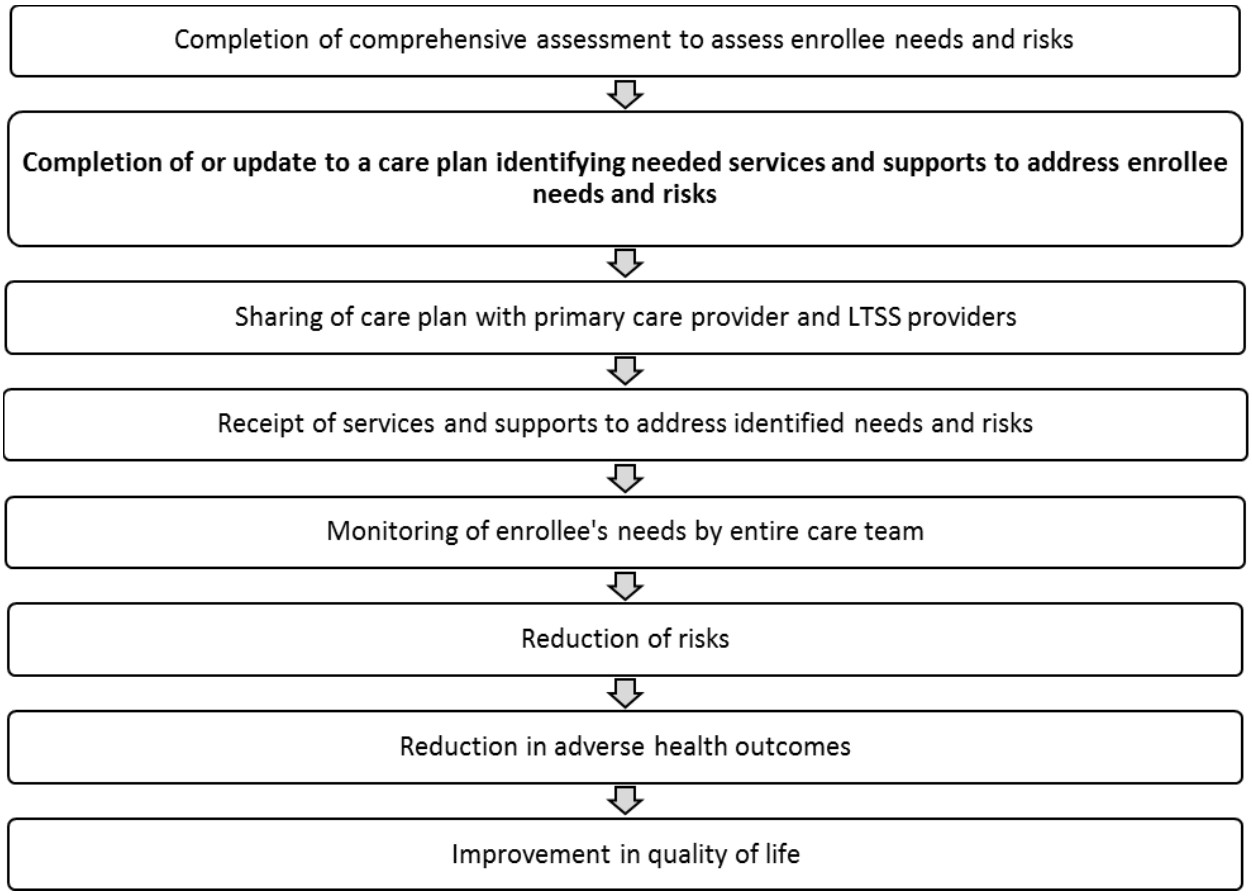
☒ Process: This measure assesses the extent to which Managed Long Term Services and Support (MLTSS) enrollees receive a comprehensive care plan for provision of long term services and supports.

☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

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 care planning 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, Medicaid 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 Medicaid MLTSS enrollees, begins with conducting and regularly updating a comprehensive assessment to identify enrollees' needs and then developing and regularly updating an individualized care plan to indicate the specific services and supports that should be provided (Rich, et al., 2012).

Variation in How Care Plans are Defined and Conducted

State Medicaid agencies have implemented numerous Medicaid MLTSS care coordination models that include care planning components (Rivard, et al., 2013). Many other programs that deliver care to individuals who are similar to (or in some cases the same as) Medicaid MLTSS enrollees require care plans and team based care, including patient-centered medical homes, Medicare managed care plans (e.g., Special Needs Plans, Financial Alignment Initiative dual eligible enrollee demonstration plans), state Medicaid home and community-based services 1915(c) waiver programs, the Program for All-Inclusive Care for the Elderly (PACE), and Medicaid Health Homes.

Despite such widespread use, uniform specifications regarding the development of care plans do not exist, and performance measures used to evaluate the quality of care plans developed are not well-established. An environmental scan conducted under a previous CMS contract (Contract No. HHSM-500-2010-00026I/HHSM-500-T0011) highlights the lack of standardization in how care plans are defined across populations using LTSS (included as Appendix 1: Environmental Scan of Assessment and Care Planning Measures). In some states MLTSS plans use a state mandated structure for care plans, other states' MLTSS plans develop an individualized service plan in addition to the structure mandated by the state. The variation in care or service plan 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 Care Planning on Outcomes

Although no uniform specifications exist, care coordination experts agree that care plans should be based on comprehensive assessments; address items related but not limited to individuals' health and functional status and their goals, preferences, and values; and clearly specify what care is to be provided and by which care team member. Care plans should also be reviewed on a frequent basis and updated as health and social support needs change (Rich, et al., 2012).

Well-developed care plans are associated with numerous positive outcomes, including improving patient-provider and provider-provider communication, encouraging care team accountability, flagging potential concerns for future evaluation, and promoting individuals' and caregivers' self-management (Rich, et al., 2012). Documenting individuals' goals of care alone has been linked to numerous positive health outcomes. The use of structured goal-setting approaches to self-management of chronic conditions has been shown to significantly improve HbA1c levels and maintain improvements for one year in primary care-based diabetes group clinics (Naik, et al., 2011). Individual goal-setting has also been linked to better outcomes and improvements in health and functioning in a variety of other populations, such as those with dementia, coronary heart disease, stroke, end stage renal disease, and rehabilitation needs (Clare, et al., 2015; Janssen, et al., 2013; Warner, et al., 2015; Kauric-Klein 2012; Muller, et al., 2011). A recently published study found that using a care plan as a single document for sharing information across multiple settings demonstrated clinically-significant improvement in depression and improved 10-year cardiovascular risk, exercise rates, and referrals to exercise programs and mental-health clinicians (Morgan, et al., 2015).

Given the large and growing body of evidence, person-centered goal-oriented care planning has become recognized as vital to improving the quality and delivery of care for Medicare-Medicaid dual eligible enrollees. However, there are few standardized measures of this critical process. Indeed, several National Quality Forum work groups have identified goal-directed, person-centered care planning and implementation as a priority measure gap for dual enrollees (NQF, 2015). Similarly, the majority of stakeholders we interviewed for this project identified person-centered care planning as a priority for measure development for the MLTSS enrollee population.

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 also 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 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.

Clare, L., Nelis, S. M., Jones, I. R., Hindle, J. V., Thom, J. M., Nixon, J. A., Whitaker, C. J. (2015). The Agewell trial: a pilot randomised controlled trial of a behaviour change intervention to promote healthy ageing and reduce risk of dementia in later life. BMC Psychiatry, 15, 25.

Janssen, V., De Gucht, V., Dusseldorp, E., & Maes, S. (2013). Lifestyle modification programmes for patients with coronary heart disease: a systematic review and meta-analysis of randomized controlled trials. European Journal of Preventive Cardiology, 20(4), 620-640.

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/.

Kauric-Klein, Z. (2012). Improving blood pressure control in end stage renal disease through a supportive educative nursing intervention. Nephrology Nursing Journal, 39(3), 217-228.

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/.

Morgan, M.A.J., Coates, M.J., & Dunbar, J.A. (2015). Using Care Plans to Better Manage Multimorbidity. AMJ, 8(6), 208–215.

Muller, M., Strobl, R., & Grill, E. (2011). Goals of patients with rehabilitation needs in acute hospitals: goal achievement is an indicator for improved functioning. Journal of Rehabilitation Medicine, 43(2), 145-150.

Naik, A. D., Palmer, N., Petersen, N. J., Street, R. L., Jr., Rao, R., Suarez-Almazor, M., & Haidet, P. (2011). Comparative effectiveness of goal setting in diabetes mellitus group clinics: randomized clinical trial. Archives of Internal Medicine, 171(5), 453-459.

NQF (2015). Advancing Person-Centered Care for Dual Eligible Beneficiaries through Performance Measurement: 2015 Recommendations from the Measure Applications Partnership. Final Report. August. Washington DC: National Quality Forum, and NQF (2012). Measuring Healthcare Quality for the Dual Eligible Beneficiary Population. Final Report. June. Washington DC: National Quality Forum.

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. Rockville, MD: 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

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. Rockville, MD: 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.

Rivard, P., B. Jackson, J. Rachel, J. Seibert, and T. Whitworth (2013). “Environment Scan of MLTSS Quality Requirements in MCO Contracts.” Assistant Secretary for Planning and Evaluation (ASPE), U.S. Department of Health and Human Services. https://aspe.hhs.gov/system/files/pdf/76871/MCOcontr.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.

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.

Warner, G., Packer, T., Villeneuve, M., Audulv, A., & Versnel, J. (2015). A systematic review of the effectiveness of stroke self-management programs for improving function and participation outcomes: self-management programs for stroke survivors. Disability and Rehabilitation, 1-23.

1. Evidence, Performance Gap, Priority – 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_CarePlan_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.

1b. Performance Gap

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

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

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

<u>If a COMPOSITE</u> (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.
Care plans based on comprehensive assessments serve as the foundation for providing high quality and well-coordinated care. While almost all MLTSS plans require care plans be developed for their members, there is little data on the scope and content of care plans among the MLTSS enrollee population. This measure would address the lack of standardization by assessing the percentage of Medicaid MLTSS enrollees who have a comprehensive care plan developed that includes specific core and supplemental elements. A standardized measure of comprehensive care planning 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 care plan 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 care plan domains. Through review and voting the TEP identified 20 elements as critical for developing LTSS care plans. However, through our field test, we determined that requiring all 20 elements was too stringent and would not result in meaningful performance rates (i.e., five out of five plans had a 0% performance rate when all 20 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 care plan 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 (<u>current and over time</u>) at the specified level of analysis. *(<u>This is required for maintenance of endorsement</u>. 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 who 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, fewer than 1 percent of enrollees across MLTSS plans had documentation of a care plan developed in the specified timeframe, including the specified seven core elements (Rate 1). The range in performance from

0 percent to 2.4 percent indicates there is room for improvement, and in general, MLTSS plans are not routinely completing care plans in the specified timeframe including documentation of standardized core elements.

It is important to note that although these data show low rates of performance, most plans in the sample were regularly developing care plans for their enrollee population (68 percent of enrollees had documentation of at least one care plan). 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. Seven (7) core elements documented:
Mean: 0.6
Standard Deviation: 1.1
Minimum: 0.0
Maximum: 2.4

Percent of enrollees with Rate 2. Seven (7) core elements documented and four (4) supplemental elements documented:
Mean: 0.6
Standard Deviation: 1.1
Minimum: 0.0
Maximum: 2.4

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.**
Care plans based on comprehensive assessments serve as the foundation for providing high quality and well-coordinated care. While almost all MLTSS plans require care plans be developed for their members, there is little data on the scope and content of care plans among the MLTSS enrollee population.
A central challenge to measuring the scope and content of care plans is the variation in the way care plans are defined across states and health plans.

In May 2016, 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" and "quality assessment and performance improvement programs for plans offering LTSS 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, (CMS, 2016; ICRC, 2016).

In January 2014, CMS also issued a final rule that established person-centered service planning requirements for beneficiaries who receive home and community based services (HCBS) through programs that operate under 1915(c) waiver authority or 1915(i) state plan amendments, which covers some states' MLTSS programs. More specifically, it requires person-centered service plans, (CMS, 2014):
- Be developed through a person-centered planning process driven by the individual that includes people chosen by the individual, provides support to the individual to ensure that the individual directs the process to the maximum extent possible, and is timely and occurs at times/locations of convenience to the individual.
- Reflect cultural considerations, use plain language, include strategies for solving disagreement, offer choices to the individual regarding services and supports the individual receives and from whom, and provide a method to request updates.
- Reflect what is important to the individual to ensure delivery of services in a manner reflecting personal preferences and ensuring health and welfare.
- Identify the strengths, preferences, needs (clinical and support), and desired outcomes of the individual.
- Include individually identified goals and preferences related to relationships, community participation, employment, income and savings, healthcare and wellness, education, and other areas.

- Include risk factors and plans to minimize them.
- Be signed by all individuals and providers responsible for its implementation. A copy of the plan must be provided to the individual and his/her representative.
- Follow specific documentation requirements.

Despite this guidance regarding the process and principles to be used in developing person-centered services plans for people who need HCBS, there are no guidelines regarding the content or components of care plans across state HCBS programs.

This measure will provide common set of standards for assessing the degree to which care plans are comprehensive by evaluating the percentage of Medicaid MLTSS enrollees who have a care plan that includes clearly defined and specified elements that are considered to be either "core" or "supplemental".

CMS. (2014). Final Rule Medicaid HCBS. Disabled and Elderly Health Programs Group. Center for Medicaid and CHIP Services. Available at https://www.medicaid.gov/medicaid-chip-program-information/by-topics/long-term-services-and-supports/home-and-community-based-services/downloads/final-rule-slides-01292014.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

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 were unable to find research on potential disparities in the use of care plans among the MLTSS enrollee population. Some studies have identified persistent racial and ethnic disparities regarding advanced care planning (Barwise, et al., 2016; Effiong, & Myrick, 2012; Garrido, et al., 2014). However, most other research focuses on the identification of disparities in the need for and use of LTSS more broadly, which highlight the need for detailed and well-documented comprehensive assessments and care plans.

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).

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.

Barwise, A., M. Wilson, R. Kashyap, O. Gajic, & B. W. Pickering. (2016). Disparities in Advanced Care Planning in The ICU and End of Life Decision Making. Available at http://www.atsjournals.org/doi/abs/10.1164/ajrccm-conference.2016.193.1_MeetingAbstracts.A7926.
Congressional Budget Office. (2013). Rising Demand for Long-Term Services and Supports for Elderly People. Washington, DC: Congressional Budget Office.
Effiong, A. & D. Myrick. (2012). H.R. 1589: addressing racial and ethnic disparities in advance care planning among Medicare enrollees. BMJ Supportive & Palliative Care, 2, 181.
Garrido, M. M., S. T. Harrington, & H. G. Prigerson. (2014). End-of-life treatment preferences: a key to reducing ethnic/racial disparities in advance care planning? Cancer, 120(24), 3981-3986. http://www.ncbi.nlm.nih.gov/pubmed/25145489

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. **For maintenance of endorsement,** 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.*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The 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 care plan completed within 120 days of enrollment for new enrollees, with seven (7) core elements documented, or
- A comprehensive LTSS care plan completed at least once during the measurement year for all other enrollees (established enrollees) with all seven (7) core elements documented.

Rate 2: MLTSS plan enrollees who had either of the following:
- A comprehensive LTSS care plan completed within 120 days of enrollment for new enrollees, with seven (7) core elements and at least four (4) supplemental elements documented, or
- A comprehensive LTSS care plan completed at least once during the measurement year for all other enrollees (established enrollees) with seven (7) core elements and at least four (4) supplemental elements documented.

Note: Initial care plan should be developed within 120 days of enrollment (allows for 90 days to complete assessment and 30 days to complete care plan), 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)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
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 care plan (created during a face-to-face encounter) with seven (7) core elements documented within 120 days of enrollment for enrollees newly enrolled in the plan between September 1 of the year prior to the measurement year to August 31 of the measurement year, or
- A comprehensive LTSS care plan (created during a face-to-face encounter) with seven (7) 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 initial care plan or care plan update must include documentation of seven (7) core elements described below. Documentation of "no change" in the care plan is not sufficient to meet numerator criteria for the care plan update. If multiple care plans are documented in the measurement year, use the last updated care plan.

Core Elements
1. Care planned to meet enrollee medical needs. Documentation must include either plan for addressing need or documentation of no need.
2. Care planned to meet enrollee functional needs. Documentation must include either plan for addressing need or documentation of no need.
3. Care planned to meet enrollee needs due to cognitive impairment or documentation of no cognitive impairment. Example of care to meet cognitive impairment needs includes support for behavioral difficulties, caregiver support or education to address cognitive impairment, or support for keeping individual cognitive engaged in activities. Documentation must include either plan for addressing need related to cognition (or cognitive impairment/dementia) or documentation of no need.
4. List of all LTSS services and supports the enrollee receives, or is expected to receive in the next month, in the home (paid or unpaid) or in other settings (e.g., adult day health center, nursing facility), including amount (e.g., hours, days) and frequency (e.g., every day, once a week). Documentation of no LTSS services is sufficient to meet the numerator criteria.
5. At least one enrollee (and family as appropriate) individualized goal (medical or non-medical goals).
6. A plan for follow-up and communication with the care manager (i.e., documentation of follow-up and communication schedule with care manager)
7. Plan for ensuring enrollee needs are met if an emergency occurs (e.g., if a personal care assistant or home health aide is unable to get to home, natural disaster). Must include at a minimum the name of an individual at the MLTSS plan or contracted provider to contact in case of an emergency.

Rate 2: MLTSS plan enrollees who had either of the following:
- A comprehensive care plan (created during a face-to-face encounter) with seven (7) core elements and at least four (4) supplemental elements documented within 120 days of enrollment for enrollees newly enrolled in the plan between September 1 of the year prior to the measurement year to August 31 of the measurement year, or
- A comprehensive care plan (created during a face-to-face encounter) with seven (7) core elements and at least four (4) supplemental 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 eight possible supplemental elements are described below. The care plan must include documentation the seven (7) core elements and at least four (4) of the eight (8) supplemental elements to count towards the second rate. If multiple care plans are documented in the measurement year, use the last updated care plan.

Supplemental Items
1. Care planned to meet enrollee emotional needs. Documentation must include either plan for addressing need or documentation of no need.
2. Care planned to meet enrollee social or community integration needs. Documentation must include either plan for addressing need or documentation of no need. Examples of care to meet social/community integration needs includes planned social activities with friends and family, participation in community based activities, or participation in work or volunteer activities.
3. Duration of all LTSS services and supports the enrollee receives, or is expected to receive in the next month, in the home (paid or unpaid) or in other settings (e.g., adult day health center, nursing facility), or time at which services will be reassessed. Documentation of no LTSS services is sufficient to meet the numerator criteria.
4. Contact information for enrollee's key LTSS providers or documentation of no LTSS services. Documentation of no LTSS services is sufficient to meet the numerator criteria.

5. Documentation of plan for assessing progress towards enrollee goal(s). Examples of plan for assessing process towards goal include plan for when provider and individual will follow-up on goal.
6. Documentation of barriers to meeting enrollee goal(s). Examples of barriers to meeting goals include factors in the enrollee´s life, community or health factors that may make it difficult to achieve enrollee defined goal.
7. First point of contact for enrollee. The name and contact information for the care manager is sufficient to meet this item if it is provided to the individual.
8. Contact information for enrollee's primary care practitioner (PCP).

Rates 1 & 2: Additional Notes
The comprehensive LTSS care plan is a document or record that identifies enrollee needs, preferences, and risks, and contains a list of the services and supports planned to meet those needs while reducing risks. There must be documentation that the care plan was created with input from the enrollee during a face-to-face encounter between the individual responsible for creating the care plan (care manager) and enrollee. The assessment and development of the care plan may be done during the same face-to-face encounter or during different encounters.

A care plan may be called a service plan in certain Medicaid MLTSS plans. Per its definition, the care plan must include:
- Documentation on whether family or friend caregiver(s) were involved in the development of the care plan, and the contact information for said caregiver(s). If there is no friend or family caregiver involved in care-planning, the lack of informal caregiver availability must be documented to meet this element.
- Documentation of enrollee (or power of attorney) agreement to comprehensive care plan, or appeal of care plan. Documentation of agreement includes: verbal agreement from the enrollee, or power of attorney (POA), received by phone or in person OR written agreement from the enrollee, or POA, received by mail (e.g., a signature). Documentation that a care plan was discussed or reviewed is not sufficient to meet this measure. The documentation must indicate that the enrollee (or POA) agreed to the care plan or the care plan is being appealed.
- Development of the initial comprehensive care plan or care plan update is not required to be done in-home.

S.6. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Medicaid MLTSS 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.
- 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 care plan 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 Assessment 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) and 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 development of a comprehensive care plan or who refused to participate in development of a comprehensive care plan.

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) and 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 care plan update was conducted.

Exclude enrollees who could not be reached for development of a comprehensive care plan or who refused to participate in development of a comprehensive care plan. Enrollees who refuse care planning are excluded from the requirement of having goals and preferences documented and enrollee signature.

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 care plan within 120 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 care plan update 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 rate of enrollees who could not be reached for a comprehensive care plan.

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 care plan within 120 days of enrollment.
Step 2c. 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 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 care plan update 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 rate of enrollees who refused comprehensive care plan.

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 care plan (Step 2b + Step 2e).
Step 3c. Identify enrollees who have documentation of a comprehensive care plan with core elements within 120 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 care plan (Step 2b + Step 2e).
Step 3g. Identify enrollees who have documentation of a comprehensive care plan with 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 care plan documented in the appropriate time frame with seven (7) core elements.

Numerator Rate 2

Step 4a. From enrollees identified in Step 3c (new enrollees with a completed care plan of the core elements), identify enrollees who have documentation of at least four (4) 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 four (4) 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 care plan documented in the appropriate time frame with nine (9) core elements and at least four (4) 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
- Comprehensive LTSS Care Plan
- 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 care plan elements were documented during the required time frame.

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_CarePlan_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

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Long Term Services and Supports Comprehensive Care Plan and Update
Date of Submission: 11/8/2017
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 [16] differences in performance**;

OR

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

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

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

Notes

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

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly

addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions.
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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

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

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

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.17***)	**Measure Tested with Data From:**
☒ abstracted from paper record	☒ abstracted from paper record
☐ claims	☐ claims
☐ registry	☐ registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒ other: 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 <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

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

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

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

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

Characteristic	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 sociodemographic (SDS) variables 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 is on the final measure specifications. Additional detail 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:

$$\hat{K} = \frac{\rho_a - \rho_e}{1 - \rho_e}$$

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

Fifteen of the twenty-three potential data elements (15 care plan specific elements and eight additional elements related to dates, settings, and contact information) 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 comprehensive care plan completed ($\hat{\kappa}$= 0.7003), setting of care plan development, medical needs assessment documented, care plan date, emotional needs, list of all services received/expected to receive, ($\hat{\kappa}$=0.5486, 0.5049, 0.5009, 0.4899 and 0.4669,). Data elements with low reliability were first point of contact for enrollees, friend/family involvement, and assessment of social needs, ($\hat{\kappa}$= 0.1295, 0.1160, and 0.0459).

Other data elements with moderate reliability include frequency, amount and duration of services ($\hat{\kappa}$= 0.5425, 0.5362 and 0.5629) at least one enrollee goal, PCP contact information, follow up and communication schedule with care manager, key LTSS providers contact information, and an emergency backup plan ($\hat{\kappa}$= 0.5486, 0.4646, 0.4848, 0.4098 and 0.4740.). Data elements with the lowest reliability were provider name, plan for assessing progress towards goals, desired level for involvement in care planning, friend/family contact information and barriers to meeting goals, ($\hat{\kappa}$= 0.2387, 0.3597, 0.2861, 0.3859, 0.3331 and 0.1467).

Five of the seven data elements designated as “core” elements for the measure met the threshold for moderate or higher reliability ($\hat{\kappa} \geq 0.4$). Core data elements with the highest reliability were comprehensive care plan completed, at least one enrollee goal, list of all LTSS services and supports the enrollee receives, or is expected to receive, a plan for follow up and communication with the care manager, and an emergency back-up plan, ($\hat{\kappa}$= 0.7003, 0.5486, 0.4899, 0.4848 and 0.4740). Core data elements with the lowest reliability were a care plan to meet enrollee functional needs and a care plan to meet enrollee cognitive needs ($\hat{\kappa}$=0.3257, -0.0240).

A total of four of the eight data elements designated as "supplemental" elements for the measure met the threshold for moderate reliability ($\hat{\kappa} \geq 0.4$). Supplemental data elements with moderate reliability include duration of all LTSS services and supports, a care plan to meet enrollee's emotional needs, contact information for enrollee's PCP, and contact information for enrollee's key LTSS provider ($\hat{\kappa}$=0.5629, 0.4669, 0.4646, 0.4098). Supplemental data elements with the lowest reliability include documentation of plan assessing progress towards enrollee goals, documentation of barriers to meet enrollee goals, care plan to meet enrollee's social needs, and first point of contact for enrollee, ($\hat{\kappa}$=0.3597, 0.3859, 0.0459, 0.1295).

Table 3. Reliability of key data elements

Data element	Kappa statistic	Interpretation
Date of Birth	0.8426	Almost Perfect
Sex	0.8788	Almost Perfect
Place of Residence	0.4706	Moderate
Date of First Enrollment	0.7108	Substantial
Date of First Disenrollment	-0.5052	Less than Chance Agreement
Comprehensive Care Plan Completed	0.7003	Substantial
Setting of Care Plan Development (Face-to-Face, Phone, Other)	0.5486	Moderate
Care Plan Date	0.5009	Moderate
Elements Documented:		
Summary of Assessment**	NA	NA
Medical Needs*	0.5049	Moderate
Functional Needs*	0.3257	Fair
Cognitive Needs*	-0.0240	No Agreement
Emotional Needs	0.4669	Moderate
Social Needs	0.0459	Slight
List of All Services Received/Expected to Receive***	0.4899	Moderate
Frequency of Services***	0.5425	Moderate
Amount of Services***	0.5362	Moderate
Duration of Services	0.5629	Moderate
Provider Name**	0.2387	Fair
At least One Enrollee Goal*	0.5486	Moderate
Plan for Assessing Progress towards Goal(s)	0.3597	Fair
Desired Level of Involvement in Care Planning**	0.2861	Fair
Barriers to Meeting Goals	0.3859	Fair
First Point of Contact for Enrollees	0.1295	Slight

Data element	Kappa statistic	Interpretation
PCP Contact Information	0.4646	Moderate
Follow-up and Communication Schedule with Care Manager*	0.4848	Moderate
Key LTSS providers Contact Information	0.4098	Moderate
Emergency Back-up Plan*	0.4740	Moderate
Friend/Family Involved	0.1160	Slight
Friend/Family Contact Information	0.3331	Fair
Signature of Enrollee	0.6974	Substantial

* Core elements
** Removed elements
***Elements combined into one core element

Reliability of Measure Rates

Table 4 describes the ICCs for Rate 1 and Rate 2 which exceed 0.9, indicating almost perfect agreement between the samples, and showing a significant association at p<0.5 or less. 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	0.9229*	Almost Perfect
Rate 2: Core Elements + 4 or More Supplemental Documented Elements	0.9229*	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 a care plan to meet enrollee functional needs, and a care plan to meet enrollee cognitive needs. To address this limitation, we revised the measure specifications to include greater specific in the definition of these elements and reduce inter-rater variation in interpretation. For example, we revised the item on plan of care to meet cognitive needs to specifically refer to a plan of care to meet needs related to cognitive impairment or documentation of no cognitive impairment. We revised the item on plan of care to meet social needs to include community integration and gave examples of care plan elements which address social needs. We also removed some items which had low

reliability and were confusing to record reviews (i.e. desired level of involvement in care planning and list of providers).

The Interclass Correlation Coefficient for both Rates 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 care plan 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 care plan measure and four other measures being tested for the MLTSS population in this project:

- LTSS Comprehensive Assessment and Update (MLTSS-1)
- LTSS Shared Care Plan with Primary Care Practitioner (MLTSS-3)
- LTSS Re-Assessment and 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:

9. Denominator is appropriate given the intent of the measure
10. Numerator Rate 1 is appropriate given the intent of the measure
11. Numerator Rate 2 is appropriate given the intent of the measure

12. Exclusion 1 is appropriate given the intent of the measure
13. Exclusion 2 is appropriate given the intent of the measure
14. Would high performance on this measure indicate that a health plan is providing higher quality care?
15. 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:

16. Do you have any recommendations that would help strengthen the Comprehensive Care Plan 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 multistakeholder 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.

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, very few positive, significant relationships were observed. The Spearman Rank Correlation coefficient, $\rho\hat{}$, showed a significant, strong positive relationship between the two rates of this measures (Rate 1: Core Elements vs. Rate 2: Core Elements and Supplemental Elements) and the two rates in a paired measure *Comprehensive Assessment and Update* measure (Rate 1: Core Elements vs. Rate 2: Core Elements and at least 12 Supplemental elements), as shown in Table 5 (correlation of recommended measure rates). 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, Rate 1	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 the Comprehensive Care Plan and Update measure. Most TEP members 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	5	1	1	2	2	1	1
Agree	8	10	10	6	9	6	7
Disagree	0	1	0	4	1	3	3
Strongly Disagree	0	0	1	1	0	3	1
No response	0	1	1	0	1	0	1
Total % Agree	100%	85%	85%	62%	85%	54%	62%

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, the Long Term Services and Supports Comprehensive Care Plan and Update measure Rate 2, which reports the percentage of enrollees with all seven core elements and at least four supplemental elements, appears the most useful as an "aspirational" measure. Health plan performance is lower for Rate 2 relative to Rate 1 (focused on just core elements), but still yields non-zero rates for two 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. 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 Assessment of Performance Measure Score

Because all of the MLTSS measures 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 three health plans have zero rates, while for the other measure, the other two 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 in care plans - 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*)

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the*

burden of increased data collection and analysis. *Note*: ***If patient preference is an exclusion****, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.

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

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with** Click here to enter number of factors **risk factors**

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

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

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

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:

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*)

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. Note that health plan 01 demonstrated rates that differed significantly from the mean at the .05 level.

Table 7. Long Term Services and Supports Comprehensive Care Plan and Update (MLTSS-2) Performance Rates by Health Plans with Significant Difference Noted

Health Plan	Rate 1: 7 Core elements	Rate 2: 7 Core elements + 4 supplemental elements
HP 01	2.4*	2.4*
HP 02	0.7	0.7
HP 03	0.0	0.0
HP 04	0.0	0.0
HP 05	0.0	0.0
Minimum	0.0	0.0
Mean	0.6	0.6
Maximum	2.4	2.4
Standard deviation	1.1	1.1

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 poor overall performance, we do see that health plan 01 had rates that demonstrated 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.

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 care plan. 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 care plan for MLTSS enrollees. This measure assesses the percentage of Medicaid MLTSS enrollees who have a care plan developed 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 Care Plan 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. Delivering effective care coordination for complex populations, such as Medicaid MLTSS enrollees, begins with conducting and regularly updating a comprehensive assessment to identify enrollees' needs and then developing and regularly updating an individualized care plan to indicate the specific services and supports that should be provided. State Medicaid agencies have implemented numerous Medicaid MLTSS care coordination models that include care planning components. Many other programs that deliver care to individuals who are similar to (or in some cases the same as) Medicaid MLTSS enrollees require care plans and team based care, including patient-centered medical homes, Medicare managed care plans (e.g., Special Needs Plans, Financial Alignment Initiative dual eligible enrollee demonstration plans), state Medicaid home and community-based services 1915(c) waiver programs, the Program for All-Inclusive Care for the Elderly (PACE), and Medicaid Health Homes. Despite such widespread use, uniform specifications regarding the development of care plans do not exist, and performance measures used to evaluate the quality of care plans developed are not well-established. This measure would address the lack of standardization by assessing the percentage of Medicaid LTSS enrollees who have a care plan created or updated that includes specific core and supplemental elements. A standardized measure of care plan creation and update 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_CarePlan_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.

NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 3325
Measure Title: Long Term Services and Supports (LTSS) Shared Care Plan with Primary Care Practitioner
Measure Steward: Centers for Medicare & Medicaid Services, Centers for Medicaid & CHIP Services
Brief Description of Measure: This measure assesses the percentage of Medicaid Managed Long Term Services and Supports (MLTSS) Plan enrollees with a care plan for whom all or part of the care plan was transmitted to the primary care practitioner (PCP) within 30 days of the care plan's development or update.
Developer Rationale: This measure would address the lack of coordination between LTSS and medical care providers by ensuring a patient's care plan is shared with the PCP. The MLTSS enrollee population includes individuals with complex health and social support needs. Given their complex needs, they require high levels of care coordination. Effective care coordination for complex populations, such as MLTSS enrollees, includes developing and sharing individualized care plans, which are associated with numerous positive health outcomes.

Numerator Statement: Medicaid MLTSS enrollees who have a care plan (or part of a care plan) that was transmitted to their PCP within 30 days of the care plan's development or update date.
Denominator Statement: Medicaid MLTSS enrollees age 18 years and older who had a care plan developed or updated in the measurement year.
Denominator Exclusions: Exclude enrollees in the denominator who were not enrolled in an MLTSS plan for at least 30 days after a care plan's development or update date.

Exclude enrollees who refuse to have their care plan shared.

Measure Type: Process
Data Source: Management Data, Other, Paper Medical Records
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:
3319: LTSS Comprehensive Assessment and Update
3324: LTSS Comprehensive Care Plan and Update

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 Comprehensive Care Plan and Update

New Measure -- Preliminary Analysis

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

This measure would address the lack of standardization by assessing the percentage of MLTSS beneficiaries for whom all or part of the care plan was transmitted to PCPs, and the number of days between when the care plan was first developed or updated and then shared.

- The developer provides a logic model describing the steps between the process of completing a comprehensive assessment and care plan and the outcome of improvement in quality of life.
- There is no systematic review of studies of care planning in MLTSS programs. The developer conducted a targeted literature review to gather evidence in support of the measure.
    - There is evidence that PCPs who do not consistently receive communication from specialists were significantly more likely to report that their ability to provide high quality care was jeopardized (O'Malley & Reschovsky, 2011)
    - Systematic review of HBPC program evaluations found that among the nine studies that met high evidence standards, eight resulted in substantial reductions in at least one of the outcomes (emergency department visits, hospitalizations, hospital beds days of care, long-term care admissions, and long-term care bed days of care) with seven demonstrating reductions in at least two of these outcomes (Stall et al., 2014)
    - An ongoing evaluation of the Independence at Home demonstration, a HBPC program sponsored by CMS, found that during the second year of the program, all 15 participating practices improved performance on at least two of the six quality measures, and four practices met the performance thresholds for all six quality measures, including: annual documentation of patient preferences; all-cause hospital readmissions within 30 days; and avoidable hospital admissions and emergency room visits (CMS, 2017)
- Technical Expert Panel (TEP) convened in 2013 to provide insight into the priority areas for measurement and the usefulness and feasibility of the identified measures for MLTSS plans. The 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.

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?*
- *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*

Guidance from the Evidence Algorithm
Process measure not based on systematic review (Box3) -> Empirical evidence without SR or grading (Box 7) -> Empirical evidence includes all studies in the body of evidence (Box 8) -> High-moderate quality of evidence (box 9) -> 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 measure addresses the lack of coordination between LTSS and medical care providers by ensuring a patient's care plan is shared with the PCP. Performance data provided is from five MLTSS health plans representing at least two or more of the major LTSS sub-populations. The data demonstrating the proportion of beneficiaries who have a care plan shared with their PCP shows significant room for improvement:

Percent of beneficiaries with care plan shared with PCP
Mean 6.5%
Standard Deviation 10.2%
Minimum: 0%
Maximum 23.4%

Developer provided additional performance data on care plan sharing rates by enrollee type that shows approximately 30 percent of the enrollees in the measure's denominator (133 enrollees) had their care plan shared with a PCP or a key LTSS provider at least once in the measurement period (see below). Among all the 133 enrollees who had an initial care plan or care plan update shared, 63 percent were shared within 30 days, and most were shared within 10 days.

Table 1. Care plan sharing rates by enrollee type

Enrollees with	Frequency	Percentage of enrollees with a care plan eligible for measure (n =438)
Documentation of a care plan shared with an eligible provider (PCP or Key LTSS)	**133**	**30.4**
Care plan shared within 30 days from its creation	84	19.2
Care plan shared after 30 days of its creation	21	4.8
Data entry error*	28	6.4
No documentation of a care plan shared with an eligible provider	**305**	**69.6**

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

The developer cites additional literature indicating an opportunity for improvement:

- Coordination, when it occurs, is idiosyncratic and often depends on the efforts of the care coordinator to communicate with all relevant parties and to arrange for information to flow (Saucier & Burwell, 2015).
- Confusion regarding regulations protecting patient health information can often hinder necessary information exchange (McGinn-Shapiro et al., 2015)
- Technology has been shown a critical barrier to coordination between LTSS and medical care providers (NCQA, 2015)

Disparities

- The developer did not find any research on disparities in the sharing of care plans among the LTSS enrollee population, however studies have identified persistent racial and ethnic disparities regarding advanced care planning (Barwise et al., 2016; Effiong & Myrick, 2012; Garrido et al., 2014)
- 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 (CBO, 2013).
- The developer collected information about race and ethnicity during testing. However, due to the overall low rates, they do not believe additional analysis of disparities would provide meaningful information.

Questions for the Committee:

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

**This is an important intermediate process measure. Community providers (home care, ALF, SNF, etc.) are critical care providers and linking to the PCP is essential; however, there are no existing measures to assess. The real outcome measure is that the PCP, with this knowledge, improves the patient's care - but this measure facilitates the ability to get to that phase. The literature clearly makes a case for importance and relevance and the measures from the five plans establishes that this is possible and that best practice evidence may very well exist from which to make system improvements. Suggestions: we do need clear descriptors of the setting/providers that are included AND what is included in the care plan measure. The use of technology in the collection seems to vary - concern this may alter the measure (?). Expanding beyond Medicaid may quickly become a consideration as other populations struggle with the coordination of care challenges. I do wonder with the advancement of HIE, is there a simpler way to implement this through that existing structure. And what is the burden to providers – this may be lightened via a HIE.

** "How does the evidence relate to the specific structure, process, or outcome being measured? No. Virtually all of the references provided did not reflect the specific intent of this measure and were used to extrapolate potential benefit without explicitly demonstrated linkages.

Does it apply directly or is it tangential? The evidence presented is by in large tangential at best across most of the references provided for this measure. There is an implicit assumption that improved communication is linked to tangible utilization and clinical outcomes, but that assumption is largely unproven or weak at best from the evidence provided.

How does the structure, process, or outcome relate to desired outcomes? Other than meeting administrative requirements for enrollment in government programs, no discrete health outcomes are specified in the evidence analysis provided for this measure other than hypothetical expectations.

For maintenance measures –are you aware of any new studies/information that changes the evidence base for this measure that has not been cited in the submission? N/A"

**There appears to be indirect, but not direct evidence, in support of this process measure's association with outcomes.

**Empirical data are provided and are relatively complete. Some evidence is tangential (for example, data that relates PCP - specialist communication.) Overall, this is a process measure with moderate empirical evidence relating proposed process to outcomes. I am not aware of any other data that should be included.

1b. Performance Gap

Comments

**Yes, preliminary performance data was measures with five MLTSS health plans. Receipt rate very low, supporting need to understand this measure. Full data included. One health plan demonstrated a meaningful increase in scores - worthy of further understanding. Disparity information not included and no literature to support need to do so.

**There were no specific data presented documenting the performance gap in the evidence review, just an acknowledgment that states are inconsistent and highly variable in their requirements regarding sharing of LTSS related plans of care.
Performance data provided is from five MLTSS health plans representing at least two or more of the major LTSS sub-populations. The data demonstrating the proportion of beneficiaries who have a care plan shared with their PCP shows significant room for improvement:
Percent of beneficiaries with care plan shared with PCP
Mean 6.5%
Standard Deviation 10.2%
Minimum: 0%
Maximum 23.4%
Nothing mentioned about healthcare disparities in these results.
No documentation of a care plan shared with an eligible provider* (*is this the same as a PCP?) 69.6%.
There is no specific outcome or follow on process identified or analyzed in the context of ""sharing"" the care plan with a PCP. I.e. nothing about ""what happens next"" with this process and also not specifically identified or pre-specified by the measure developer. "
**There was a gap identified in sharing care plans from pilot sites (there is no demographic data submitted re: disparities and demographics), reportedly because of low overall numbers.
**Performance gap in communication of care plans to PCP is demonstrated clearly. Impact (outcome) of potentially closing this gap is less clear.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing
2b. Validity: Testing; Exclusions; Risk-Adjustment; Meaningful Differences; Comparability Missing Data
2c. For composite measures: empirical analysis support composite approach

Reliability
2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. For maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures.
2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers. For maintenance measures – less emphasis if no new testing data provided.
Validity
2b2. Validity testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For maintenance measures – less emphasis if no new testing data provided.
2b2-2b6. Potential threats to validity should be assessed/addressed.

Complex measure evaluated by Scientific Methods Panel? ☐ **Yes** ☒ **No**
Evaluators: NQF Staff

Evaluation of Reliability and Validity (and composite construction, if applicable): Link A

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.)?*
- *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**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

2a1. Reliability- Specifications

Comments

**Yes. Very through. Outstanding questions: clear descriptors on who are included as providers; what is included in the "care plan" such as meds, risk factors, etc.; does it include contact information. Under Table 3 - do question the impact of the reliability measure with date of first disenrollment -- not clear on the process but concern this impacts the integrity of the data. Details on provider types will be warranted (only those paid by MA in the community of non-hospital facility?)
**Elements in the LTSS Shared Care Plan with Primary Care Practitioner measure were assessed too infrequently among the 144 paired assessments (<30) to allow for inter-rater reliability analysis. The data element indicating the care plan shared met the threshold for slight reliability. Disagree with preliminary rating of ""moderate""would put this as ""low"" to ""insufficient"". Noted in the Feasibility section is also this information:
• Data elements used to calculate this measure are abstracted from record by someone other than the person obtaining original information.
• This measure is primarily collected from health plan and case management records, many of which are electronic.
o Some data elements are in defined fields in electronic sources
o The data elements needed for this measure are not standardized (i.e. may be in free text rather than structured fields).
"
**limited information submitted
**Specifications seem clear, but testing samples were small.

2a2. Reliability- Testing

Comments

**see above
need a better understanding of the ""other elements"" that were assessed too infrequently to allow IRR analysis.
**Yes, believe it is low to insufficient per comments above. Disagree that it is "moderate".

**What constitutes a "care plan" might be inconsistently applied because the definition has some inherent ambiguity. Same with what constitutes an "update".
**Inadequate data for inter-rater reliability analysis as noted in the ""Scientific Acceptability"" document.
Reliability seems acceptable with regard to other elements."

2b1. Validity—Testing
2b4-7. Threats to Validity
2b4. Meaningful Differences
Comments
**Tested both performance and face validity. Panels large (clarification - that is the group included in the packet - correct?) and inclusive; however, do not see as many "providers" as I would have thought as this would be the group who would best understand validity. I did not see the TEP comments in my packet (?) so cannot fully describe fully assess response to comments. Missing data not fully described, but was discussed. Did compare with other care plan recommended measure rates which was helpful. Overall TEP recommended measure. Table 6 - row calculations would be helpful; clear that >80% agree. More discussion on harmonizing is warranted in moving forward. Patients can refuse to have their care plans shared which is confusing since this is a managed care program; curious how often this actually occurs. Any other exclusion trends?
**Do you have any concerns with the testing results?
2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data)
Yes--insufficient data to generate statistically valid inter-rater reliability.
2b4. Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality? Not demonstrated other than low initial response rates. Lack of response could be directly related to burden of meeting expectations of the measure developers.
2b5. Comparability of performance scores: If multiple sets of specifications: Do analyses indicate they produce comparable results? Not done.
2b6. Missing data/no response: Does missing data constitute a threat to the validity of this measure? Unsure.
**Transmission of a care plan does not reflect receipt of the care plan by the PCP or quality of the care plan itself (both of which are likely more important). Lack of direct evidence to link success in this process measure to improved outcomes is a barrier to validity.
**Some empirical testing was performed but was inconclusive.
Measure validity relies heavily on Technical Expert Panel for face validity assessment. Expert panel agreement regarding face validity is tempered, at best (62% and 54% agree on questions 5 and 6 in the Systematic Assessment of Face Validity). If one more expert disagreed with either question, face validity questions would be 50% or less. Is a simple majority (7 or 8 out of of 13 experts) enough to support face validity? Seems like that should be much higher.

2b2-3. Other Threats to Validity
2b2. Exclusions
2b3. Risk Adjustment
Comments
**n/a None of these threats were explicitly evaluated or mentioned in the analysis.
**Exclusions are appropriate. No risk adjustment needed.

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.

- Data elements used to calculate this measure are abstracted from record by someone other than the person obtaining original 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.

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?*

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

**A consistent approach will need to be well-defined as the data sources are originating in such a variety of setting (home care, SNFs) with small and large operations. Use of a HIE might help in a consistent mechanism to deliver to the PCP - concern if some information comes via EHR, fax, mail, email, it may make it difficult for the PCP to manage. Agree that an emeasure would be warranted for ongoing measurement

**The data is not standardized or congruent with data formatting standards making this very challenging from a measurement and evaluation standpoint. There is also no mention of the need for interoperability requirements, given that this measure evaluates a "handshake" of information and nothing more.

**Not necessarily routinely collected in a manner that makes them accessible. Not collected in a standardized way at present, but likely could be abstracted (with some effort) from the medical record kept at LTSS.

**In current environment, processes to create care plans and elements specified are not standard. Standardizing the process and elements will likely be resources intensive for health plans and potentially for providers. Unclear how much would be needed in terms of new resources to participate with and meet the measure, versus reallocation of existing resources.

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**:*

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

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

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

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

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

Improvement results

- This 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**:*

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

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

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a1. Use- Accountability and Transparency

Comments

**This measure is under initial endorsement review; no accountability program in place. This measure is useful as a springboard to the next phase of care coordination of the post-acute and community of care linking with the PCP to ensure goals of care are consistent and the team is working towards them together. Not sure the provider group has had the opportunity to provide feedback in the development; one of their core concerns will be ease of collection and ability to get to the PCP. Consideration will need to be given to the changes in plans and physician. The other consideration is that this is at 30 days - in some situations, this may be too late - the coordination may have needed to be considerably sooner. The real use is when the PCP and provider team coordinate care together; this measure in and of itself does not get us there, but it is a step. A soon thereafter measure will need to

evaluate the quality of the information being sent (is it meaningful) and did the PCP review it to adjust the care plan.
**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.
There has been no attempt to use this measure individually or in combination with other LTSS measures presented to the Committee.
The measure could be useful to improve operational efficiency in terms of assuring higher compliance/success rates from an administrative accountability standpoint.
It is very unclear from my review whether this measure can meet the standard intent of NQF, MAP and HHS to endorse ""Measures that Matter"".
**New measure, so limited info. It appears that CMS will ask states / plans to report on this measure.
**Given that this measure would likely require additional (or reallocated) HP resources to achieve success, it would be nice to see a stronger tie to outcome. In particular, this measure is inferring a patient outcome benefit related specifically to sharing HP care plan with PCP. The measure has (at best) moderate face validity by TEP, lacks direct evidence of outcome benefit, and may consume additional resources. Based on data presented, it is not clear that this measure meets goals of improving efficiency of care or reliably distinguishing higher quality or better performing by a health plan.

Criterion 5: Related and Competing Measures

Related or competing measures

Related measures include:

- 3319 LTSS Comprehensive Assessment and Update
- 3324 LTSS Comprehensive Care Plan and Update
- 3326 LTSS Re-Assessment/Care Plan Update after Inpatient Discharge

Harmonization

N/A

Committee pre-evaluation comments

Criterion 5: Related and Competing Measures

N/A

Public and member comments

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

Comments:

Morgan Buchko, Meridian Health Plan

This requires the plan to track providing an updated or new care plan to the PCP within 30 days. If we are going to be required to report on this, we will need a spec around what constitutes a significant change that requires the PCP notificatoin.

- **Zero NQF members have submitted a support/non-support choice.**

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: 3325
Measure Title: LTSS Shared Care Plan with Primary Care Practitioner

RELIABILITY

31. 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.

32. 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)

33. 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)

34. 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)

35. 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)

36. **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)

37. Was other reliability testing reported?

☐Yes (go to Question #8)

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

38. 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)

39. 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)

40. **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)

Elements in the LTSS Shared Care Plan with Primary Care Practitioner measure were assessed too infrequently among the 144 paired assessments (<30) to allow for inter-rater reliability analysis. The data element indicating the care plan shared met the threshold for slight reliability.

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

49. 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]

50. 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)

51. 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)

52. 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.

53. 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.

54. 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

55. 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.

56. 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)

57. **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	Question 5: Would high performance on this measure indicate that a health plan is providing higher quality care?	Question 6: 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	6
Disagree	2	3
Strongly Disagree	3	3
No response	0	0
Total % Agree	62%	54%

See Testing Appendix for summaries of written feedback on Systematic Assessment of Face Validity

Question 5:

- Commenters noted that the measure was specified as a process measure and does not correlate patient outcomes to a care plan.
- TEP members that supported the measure were in agreement that sharing a care plan with the PCP is a first step in improving performance.

Question 6:

- One TEP member suggested extending the 30-day transmission window to 45 days.
- One TEP member noted that some percentage of enrollees in managed LTSS plans do not have a primary care physician.
- One TEP member suggested that there was greater importance for the patient or consumer to receive a copy of the care plan than the PCP

58. 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)

59. 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)

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

60. **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

61. Was other validity testing reported?

☐Yes (go to Question #14)

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

62. Was validity testing conducted with <u>patient-level data elements</u>?

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 <u>no</u>
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.

63. 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)

64. **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).

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** TBD

Measure Title: LTSS Shared Care Plan with Primary Care Practitioner

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

Date of Submission: TBD

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: Click here to name the health outcome

☐Patient-reported outcome (PRO): Click here to name the 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*): Click here to name the intermediate outcome

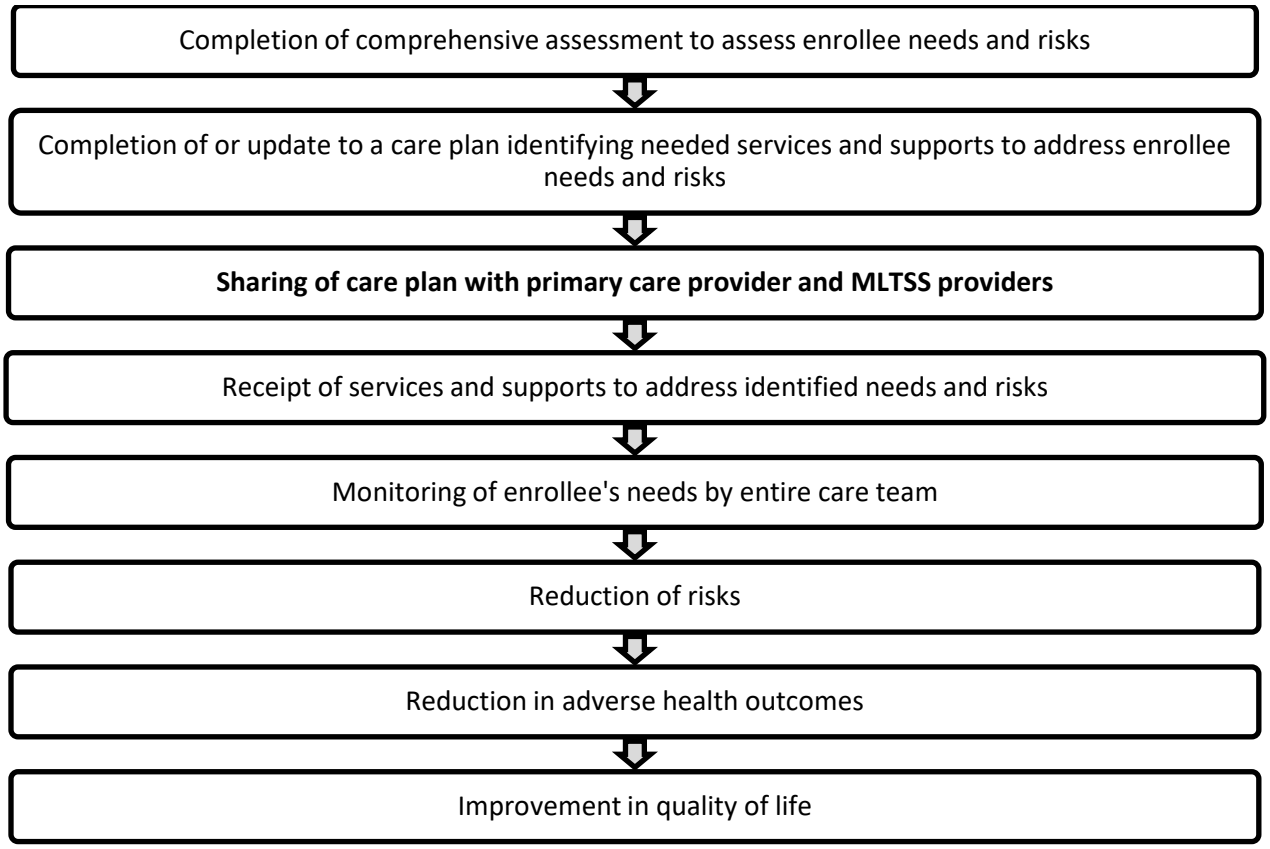
☒ Process: The percentage of Medicaid Managed Long Term Services and Supports (MLTSS) Plan enrollees with a care plan for whom all or part of the care plan was transmitted to the primary care practitioner (PCP) within 30 days of the care plan's development or update.

☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

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 shared care plans 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, Medicaid 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 Medicaid MLTSS enrollees, begins with conducting and regularly updating a comprehensive assessment to identify enrollees' needs, developing and regularly updating an individualized care plan to indicate the specific services and supports to be provided, and sharing care plans to inform care team members of services that should be coordinated (Rich et al., 2012). State Medicaid agencies have implemented numerous

Medicaid MLTSS care coordination models (Saucier & Burwell, 2015). Most models require the development of a care plan at initial enrollment and on a regular basis thereafter, as well as the use of team based care to implement the care plan. Similarly, numerous other programs that deliver care to individuals who are similar to (or in some cases the same as) Medicaid MLTSS enrollees require care plans and team based care, including patient-centered medical homes, Medicare managed care plans (e.g., Special Needs Plans, Financial Alignment Initiative dual eligible enrollee demonstration plans), state Medicaid home and community-based services 1915(c) waiver programs, the Program for All-Inclusive Care for the Elderly (PACE), and Medicaid Health Homes. In order for team-based care to be effective, providers must share the care plan and communicate changes and updates to the care plan so that all members of the care team have a complete picture of the person's needs, preferences, and services and supports provided.

Well-developed care plans are associated with numerous positive outcomes, including improving patient-provider and provider-provider communication, encouraging care team accountability, flagging potential concerns for future evaluation, and promoting individuals' and caregivers' self-management (Rich et al., 2012). Documenting goals alone has been linked to numerous positive health outcomes across different care settings, such as greater improvements in health and functioning, in a variety of MLTSS-related populations, such as those with dementia (Clare et al., 2015), coronary heart disease (Janssen et al., 2013), stroke (Warner et al., 2015), end stage renal disease (Kauric-Klein, 2012), and rehabilitation needs (Muller et al., 2011).

In all of the above studies, the care plan was shared with the clinicians providing care. The sharing of information between providers (both MLTSS and medical care providers) is a critical component of providing coordinated person-centered care and breaking down the silos that exist between medical care and MLTSS providers.

Evidence to Support Sharing of Information with Primary Care Providers (PCP)

There is no direct evidence of the impact of sharing MLTSS care plan information with the PCP on outcomes. However, there is related evidence demonstrating the importance of sharing information about other types of specialty care with PCPs.

While primary care has been demonstrated to be associated with better health outcomes and a decrease in hospital admissions and emergency department visits, lack of communication between PCPs and specialists can hinder the effectiveness of primary care (Shi, 2012; Gandhi et al., 2000; Hanlon, 2013; O'Malley & Cunningham, 2009; O'Malley & Reschovsky, 2011). For example, 28 percent of PCPs expressed dissatisfaction with the content of information they receive from specialists and 50percent of PCPs were dissatisfied with the timeliness of information they received; within two weeks of referral visits, 40 percent of PCPs received no information from specialists, and four weeks after the referral visit, 25 percent of PCPs still had no information (Gandhi et al., 2000). Another study found 81 percent of specialists said they

"always" or "most of the time" send referring PCPs notification of results and advice to patients, but only 62 percent of PCPs say they received this information (O'Malley & Reschovsky, 2011). Those PCPs who do not consistently receive communication from specialists were significantly more likely to report that their ability to provide high quality care was jeopardized.

MLTSS providers are in a unique position to provide PCPs with valuable information about an individual's risks due to their frequent presence in the patient's home. MLTSS care managers frequently conduct in-home assessments and communicate with home based care providers. They can directly observe issues such as home safety risks, potential for medication errors due to disorganized, expired or incorrect medications, food and nutrition concerns, and environmental hazards. Direct care workers, such as personal care aides, may make even more frequent home visits, sometimes daily, to provide hands-on assistance with activities of daily living such as bathing, eating and transferring, which gives them greater opportunity to observe changes in an individual's health and functional status. However, MLTSS providers may not have the authority to modify a medical care plan based on their observations. Therefore, coordination between MLTSS providers and medical care providers is critical to avoid potentially negative outcomes for individuals using MLTSS care.

Shared care planning and team-based care coordination among PCPs and home-care providers are hallmarks of home-based primary care programs (HBPC), which serve individuals with multiple chronic illnesses and functional limitations. While the frequency of communication among PCPs and other team members in the HBPC model is much greater than that in non-HBPC programs, several studies demonstrate its effectiveness in improving quality and reducing the use of intensive care. For example, a systematic review of HBPC program evaluations found that among the nine studies that met high evidence standards, eight resulted in substantial reductions in at least one of the outcomes (emergency department visits, hospitalizations, hospital beds days of care, long-term care admissions, and long-term care bed days of care) with seven demonstrating reductions in at least two of these outcomes (Stall et al., 2014). An ongoing evaluation of the Independence at Home demonstration, a HBPC program sponsored by CMS, found that during the second year of the program, all 15 participating practices improved performance on at least two of the six quality measures, and four practices met the performance thresholds for all six quality measures, including: annual documentation of patient preferences; all-cause hospital readmissions within 30 days; and avoidable hospital admissions and emergency room visits (CMS, 2017).

Variation in the Frequency and Timeliness of Sharing of Care Plans with Providers

State MLTSS program contract provisions vary in the specificity with which they require managed care plans to facilitate sharing of care plans and clinical or other information among members' providers. For example, Massachusetts' Senior Care Options (SCO) program requires that the plans "ensure linkages among the PCP, the PCT [primary care team], and any appropriate acute, long-term care, or behavioral health providers to keep all parties informed about utilization of services," and develop protocols for sharing clinical and individualized plan of care information among the enrollee's caregivers. However, most other state MLTSS contract

language is more general with regard to coordination among medical care and MLTSS providers and most are silent with respect to sharing members' care plans among providers (Rivard et al., 2013).

This measure would address the lack of standardization by assessing the percentage of MLTSS beneficiaries for whom all or part of the care plan was transmitted to PCPs, and the number of days between when the care plan was first developed or updated and then shared.

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

In the absence of a systematic review of studies of shared care plans in MMLTSS 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. We also convened a technical expert panel (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. Under the current contract in 2016, we convened a new TEP (21 members) 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 previous CMS contract (Contract No. HHSM-500-2010-00026I/HHSM-500-T0011).

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

CMS (2017). "Independence at Home Demonstration, Corrected Performance Year 2 Results." https://www.cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2017-Fact-Sheet-items/2017-01-19.html

Gandhi, T. K., Sittig, D. F., Franklin, M., Sussman, A. J., Fairchild, D. G., & Bates, D. W. (2000). Communication breakdown in the outpatient referral process. *Journal of General Internal Medicine, 15*(9), 626-631.

Hanlon, C. (2013). Measuring and Improving Care Coordination: Lessons from ABCD III. Portland, ME: The National Academy for State Health Policy.

Janssen, V., De Gucht, V., Dusseldorp, E., & Maes, S. (2013). Lifestyle modification programmes for patients with coronary heart disease: a systematic review and meta-analysis of randomized controlled trials. *European Journal of Preventive Cardiology, 20*(4), 620-640.

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/.

Kauric-Klein, Z. (2012). Improving blood pressure control in end stage renal disease through a supportive educative nursing intervention. *Nephrology Nursing Journal, 39*(3), 217-228.

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/.

Muller, M., Strobl, R., & Grill, E. (2011). Goals of patients with rehabilitation needs in acute hospitals: goal achievement is an indicator for improved functioning. *Journal of Rehabilitation Medicine, 43*(2), 145-150.

O'Malley, A. S., & Cunningham, P. J. (2009). Patient experiences with coordination of care: the benefit of continuity and primary care physician as referral source. *Journal of General Internal Medicine, 24*(2), 170-177.

O'Malley, A. S., & Reschovsky, J. D. (2011). Referral and consultation communication between primary care and specialist physicians: finding common ground. *Archives of Internal Medicine, 171*(1), 56-65.

P. Rivard, B. Jackson, J. Rachel, J. Seibert, and T. Whitworth (2013). "Environment Scan of LTSS Quality Requirements in MCO Contracts." Assistant Secretary for Planning and Evaluation (ASPE), U.S. Department of Health and Human Services. https://aspe.hhs.gov/system/files/pdf/76871/MCOcontr.pdf

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. Rockville, MD: 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.

Shi, L. (2012). The impact of primary care: a focused review. *Scientifica (Cairo), 2012*, 432892.

Stall, N., M. Nowaczynski, and S. Sinha (2014). "Systematic Review of Outcomes from Home-Based Primary Care Programs for Homebound Older Adults." *Journal of the American Geriatrics Society,* 62(12): 2243-2251.

Warner, G., Packer, T., Villeneuve, M., Audulv, A., & Versnel, J. (2015). A systematic review of the effectiveness of stroke self-management programs for improving function and participation outcomes: self-management programs for stroke survivors. *Disability and Rehabilitation*, 1-23.

NATIONAL QUALITY FORUM

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 #: 3325
Corresponding Measures:
De.2. **Measure Title:** Long Term Services and Supports (LTSS) Shared Care Plan with Primary Care Practitioner
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 Medicaid Managed Long Term Services and Supports (MLTSS) Plan enrollees with a care plan for whom all or part of the care plan was transmitted to the primary care practitioner (PCP) within 30 days of the care plan's development or update.
1b.1. **Developer Rationale:** This measure would address the lack of coordination between LTSS and medical care providers by ensuring a patient's care plan is shared with the PCP. The MLTSS enrollee population includes individuals with complex health and social support needs. Given their complex needs, they require high levels of care coordination. Effective care coordination for complex populations, such as MLTSS enrollees, includes developing and sharing individualized care plans, which are associated with numerous positive health outcomes.

S.4. **Numerator Statement:** Medicaid MLTSS enrollees who have a care plan (or part of a care plan) that was transmitted to their PCP within 30 days of the care plan's development or update date.
S.6. **Denominator Statement:** Medicaid MLTSS enrollees age 18 years and older who had a care plan developed or updated in the measurement year.
S.8. **Denominator Exclusions:** Exclude enrollees in the denominator who were not enrolled in an MLTSS plan for at least 30 days after a care plan's development or update date.

Exclude enrollees who refuse to have their care plan shared.

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 Assessment and Update
- LTSS Comprehensive Care Plan and Update

1. Evidence, Performance Gap, Priority – 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_Shared_Care_Plan_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.

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.
This measure would address the lack of coordination between LTSS and medical care providers by ensuring a patient's care plan is shared with the PCP. The MLTSS enrollee population includes individuals with complex health and social support needs. Given their complex needs, they require high levels of care coordination. Effective care coordination for complex populations, such as MLTSS enrollees, includes developing and sharing individualized care plans, which are associated with numerous positive health outcomes.

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. 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) (integrated care), Dual-Special Needs Plan (D-SNP), or Fully Integrated Dual Eligible (FIDE) SNP. 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.

The data demonstrating the proportion of beneficiaries who have a care plan shared with the PCP shows significant room for improvement (see results below).

Percent of beneficiaries with care plan shared with PCP
Mean 6.5%
Standard Deviation 10.2%
Minimum: 0%
Maximum 23.4%

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.**
There is no direct estimate for the rate of sharing care plan information between providers. However, evidence does suggest coordination and communication between LTSS providers and medical care providers is a critical gap. Among many dual eligible enrollees, which make up a large portion of LTSS enrollees, LTSS care is covered by a state Medicaid program and medical care is covered by Medicare either FFS or in a managed care arrangement. Recent research conducted in organizations providing care coordination for LTSS services found that even when financing for both Medicaid and Medicare services is integrated, care is often delivered in silos with medical and LTSS systems operating independently. One study found that establishing relationships between providers is critical for ensuring information exchange, and although technology supports such exchanges, coordinating care remains a "high touch activity." In addition, EHRs have not been widely adopted by LTSS providers, and furthermore, existing EHRs do not incorporate the type of information needed by LTSS providers. Finally, confusion regarding regulations protecting patient health information can often hinder necessary information exchange (McGinn-Shapiro et al., 2015). Coordination, when it occurs, is idiosyncratic and often depends on the efforts of the care coordinator to communicate with all relevant parties and to arrange for information to flow (Saucier & Burwell, 2015).

Technology is one critical barrier to coordination between LTSS and medical care providers. In a case study of eight organizations financially responsible for both medical and LTSS care, only one site had a fully integrated EHR system that was accessible to both medical care and LTSS care providers. Six of the sites used separate systems for care management and medical records that are not interoperable, and one site used paper records for medical and care management services and had access to the EHR at one coordinating hospital (NCQA, 2015). This barrier to coordination was echoed by stakeholders in our interviews; they stressed the importance of the care coordinator role and the need for this person to be the communication hub between all of an individual's providers.

McGinn-Shapiro, M., S. Mitchell, E. G. Walsh, M. Ignaczak, & L. Bercaw. (2015). Information exchange in integrated care models: final report. Available at https://aspe.hhs.gov/basic-report/information-exchange-integrated-care-models-final-report.

National Committee for Quality Assurance (NCQA). Policy Approaches to Advancing Person-Centered Outcome Measurement. 2015. The John A. Hartford Foundation and The SCAN Foundation. Available at https://www.ncqa.org/Portals/0/HEDISQM/Research/Policy%20Report_Final%20Report_TSF%202-1.pdf.

Saucier, P., and 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.

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 the sharing of care plans among the LTSS enrollee population. Studies have identified persistent racial and ethnic disparities regarding advanced care planning (Barwise et al., 2016; Effiong & Myrick, 2012; Garrido et al., 2014). However, most other research focuses on the identification of disparities in the need for and use of LTSS more broadly, which highlight the need for shared care plans.

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 (CBO, 2013).

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.

Barwise, A., M. Wilson, R. Kashyap, O. Gajic, & B. W. Pickering. (2016). Disparities in Advanced Care Planning in The ICU and End of Life Decision Making. Available at http://www.atsjournals.org/doi/abs/10.1164/ajrccm-conference.2016.193.1_MeetingAbstracts.A7926.

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

Effiong, A. & D. Myrick. (2012). H.R. 1589: addressing racial and ethnic disparities in advance care planning among Medicare beneficiaries. BMJ Supportive & Palliative Care, 2, 181.

Garrido, M. M., S. T. Harrington, & H. G. Prigerson. (2014). End-of-life treatment preferences: a key to reducing ethnic/racial disparities in advance care planning? Cancer, 120(24), 3981-3986.http://www.ncbi.nlm.nih.gov/pubmed/25145489

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. **For maintenance of endorsement,** 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.*

IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
Medicaid MLTSS enrollees who have a care plan (or part of a care plan) that was transmitted to their PCP within 30 days of the care plan's development or update date.

S.5. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
Time Period for Data:
13 months (November 30 of the year prior to the measurement year to December 31 of the measurement year).

Numerator Details:
Medicaid MLTSS enrollees who have a care plan (or part of a care plan) that was transmitted to the PCP within 30 days of the care plan's completion or update date. Evidence of a transmitted care plan should meet the following criteria:
- Who the care plan was transmitted to.
- Date of transmittal.
- The elements of the care plan that were transmitted.

Note: If the enrollee has more than one care plan developed during the measurement year, use the most recent care plan date to assess measure.

Definitions:
Care plan: A document or electronic record which identifies enrollee needs, preferences, and risks, and contains a list of the services and supports planned to meet those needs while reducing risks.
Transmitted: Care plan may be transmitted to providers via mail, fax, secure e-mail, alert in provider portal system, mutual access to an electronic health record (EHR) with notification to the PCP, or other electronic data system.
Primary Care Practitioner (PCP): A physician, non-physician (e.g., nurse practitioner, physician assistant), or primary care practice, who offers primary care medical services. Licensed practical nurses and registered nurses are not considered PCPs.

S.6. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Medicaid MLTSS enrollees age 18 years and older who had a care plan developed or updated in the measurement year.

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 have a care plan developed or updated between November 30 of the year prior to the measurement year and December 1 of the measurement year.
- Who are 18 years and older as of the first day of the measurement year.

- Who are enrolled in an MLTSS plan for at least 30 days after the care plan's development or update date. If multiple care plan updates are documented for the year, determine continuous enrollment from the latest care plan update in the year.
- 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: The denominator for this measure may be drawn from enrollees meeting the numerator criteria for a paired measure LTSS Comprehensive Care Plan and Update.

S.8. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Exclude enrollees in the denominator who were not enrolled in an MLTSS plan for at least 30 days after a care plan's development or update date.

Exclude enrollees who refuse to have their care plan shared.

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 not enrolled in an MLTSS plan for at least 30 days after a care plan´s development or update date. These are enrollees who may have left the plan before it was shared with the PCP.

Exclude enrollees for whom there is documentation of enrollee refusal to allow care plan sharing.

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 1. Determine the eligible population of MLTSS enrollees with a care plan developed or updated in the measurement period.
Step 2. From the eligible population, draw a systematic sample.
Step 3. From the systematic sample, remove enrollees that have documentation of refusal to allow care plan sharing.

Step 4. From the systematic sample, remove enrollees who were not enrolled 30 days after the date of care plan was development or update.
Step 5. From enrollees remaining after Step 4, identify all enrollees with a care plan for whom all or part of the care plan was transmitted to the primary care practitioner within 30 days of the care plan's development or latest update date.
Step 6. Divide the number of enrollees from Step 5 by the number of enrollees remaining after Step 4 to calculate the rate.

S.15. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.
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:
- LTSS Comprehensive Assessment and Update
- LTSS Comprehensive Care Plan and Update
- LTSS Shared 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 care plan was shared during the required time frame.

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

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_Shared_Care_Plan_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

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:** TBD
Measure Title: LTSS Shared Care Plan with Primary Care Practitioner
Date of Submission: 11/8/2017
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 <u>all</u> measures, sections 1, 2a2, 2b1, 2b2, and 2b4 must be completed.**
- **For <u>outcome and resource use</u> measures**, section **2b3** also must be completed.
- If specified for **<u>multiple data sources/sets of specificaitons</u>** (e.g., claims and EHRs), section **2b5** also must be completed.
- Respond to <u>all</u> 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.

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

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

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

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

AND

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

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

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

OR

- rationale/data support no risk adjustment/ stratification.

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

OR

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

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

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

Notes

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

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality

indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions.
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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

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

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

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

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 <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

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

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

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

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	

Characteristics	Percentage of enrollees in the testing sample (n=715)
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

Characteristic	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

Characteristic	Percentage of enrollees in the testing sample (n=715)
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 sociodemographic (SDS) variables 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 is on the final measure specifications. Additional detail on the recipients and timing of the shared care plan 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:

$$\hat{\kappa} = \frac{\rho a - \rho e}{1 - \rho}$$

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*)
The single data element in the measure met the threshold for slight reliability at ($\hat{\kappa}$=.1667)

Table 3. Reliability of key data elements

Measure (elements)	Data element	Kappa statistic	Interpretation
General	Date of Birth	0.8426	Almost Perfect
(4)	Sex	0.8788	Almost Perfect
	Place of Residence	0.4706	Moderate
	Date of First Enrollment	0.7108	Substantial
	Date of First Disenrollment	-0.5052	Less than Chance Agreement
LTSS Shared Care Plan with Primary Care Practitioner	Care Plan Shared?	0.1667	Slight

Source: Mathematica analysis of paired data from 144 MLTSS enrollees, representing five health plans.

Notes: Interpretation of Kappa statistic 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)

*The remainder of LTSS Shared Care Plan with Primary Care Practitioner elements were assessed too infrequently among the 144 paired assessments (<30) to allow for IRR analysis.

Reliability of Measure Rates

ICCs for the measure rate 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 Measure Rates

Measure	ICC statistic	Interpretation
LTSS Shared Care Plan with Primary Care Practitioner		
Rate: Shared with PCP without Consideration for Share Method	0.9668**	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<0.01 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?*)
The Interclass Correlation Coefficient for the measure rate exceeds 0.9, indicating almost perfect agreement between the samples for the single data element indicating that the care plan was shared. However, the other elements in the LTSS Shared Care Plan with Primary Care Practitioner measure were assessed too infrequently among the 144 paired assessments (<30) to allow for inter-rater reliability analysis.

2b1. VALIDITY TESTING

2b1.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Performance measure score**

☒ **Empirical validity testing**

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

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

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 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 care plan sharing measure with four other measures that were tested for the MLTSS population in this project:

- LTSS Comprehensive Assessment and Update
- LTSS Comprehensive Care Plan and Update Measure
- LTSS Re-Assessment and Care Plan Update after Inpatient Discharge
- Falls: Screening, Risk-Assessment and Plan of Care to Prevent Future Falls

We examined the correlation of different versions of the five MLTSS measures with each other using the Spearman Rank Correlations. As with this assessment measure, we tested several variations of each of the other MLTSS measures. Although we analyzed validity among all versions of the five measures for ease of review we present results for only the most promising versions of the measures in Table 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 is appropriate given the intent of the measure
3. Exclusion is appropriate given the intent of the measure
4. Would high performance on this measure indicate that a health plan is providing higher quality care?
5. 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:

6. Do you have any recommendations that would help strengthen the Shared Care Plan with Primary Care Practitioner 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 multistakeholder 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.

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, very few positive, significant relationships were observed. The Spearman Rank Correlation coefficient, ρˆ, showed no significant relationship between the LTSS Shared Care Plan with Primary Care Practitioner measure and any other MLTSS measure as shown in Table 5 (correlation of recommended measure rates).

Table 5. Correlation of recommended measure rates

Measures	MLTSS-1, Rate 3	MLTSS-1, Rate 4	MLTSS-2, Rate 3	MLTSS-2, Rate 4	MLTSS-3	MLTSS-4, Rate 1	MLTSS-4, Rate 2	MLTSS-5 Screening	MLTSS-5 Assess	MLTSS-5 Care Plan
MLTSS-3, Rate: Shared with PCP, w/o consideration for Share Method	-0.0574	-0.0574	0.5000	0.5000	--	-0.4472	-0.4472	-0.8944*	-0.6250	NA

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

*Significant association, at $p < 0.05$

MLTSS-1 = LTSS Comprehensive Assessment and Update
Rate 3 = Assessment with 9 core elements documented
Rate 4 = Assessment with at least 12 supplemental elements documented
MLTSS-2 = LTSS Comprehensive Care Plan and Update
Rate 3 = Care plan with 7 core elements documented
Rate 4 = 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 and 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 the voting results from the survey. Overall, the majority of TEP members supported the denominator, numerators, and exclusions for the Long Term Services and Supports (LTSS) Shared Care Plan with Primary Care Practitioner measure. The majority of TEP members also agreed that the measure is reflective of quality and had the potential to distinguish performance.

Table 6

Response	Denominator is appropriate given the intent of the measure	Numerator Rate is appropriate given the intent of the measure	Exclusion 1 is appropriate given the intent of the measure	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	2	1	1
Agree	11	10	10	7	6
Disagree	0	0	1	2	3
Strongly Disagree	0	2	0	3	3
No response	1	0	0	0	0
Total % Agree	92%	85%	92%	62%	54%

Additional Face Validity Feedback

The TEP noted that his 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 in care plans - including non-medical needs. The measurement team agrees that sharing a care plan with a PCP is an important initial activity to determine that organizations are assessing the needs of beneficiaries, and as documentation of the core elements improves among plans over time, performance is likely to become more meaningful and useful.

Additionally, feedback from the public comment period was generally supportive of the measure. 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

As documentation of the core elements improves among plans over time, performance is likely to become more meaningful and useful and the internal validity of the measures should also improve accordingly.

This measure assesses the percentage of enrollees age 18 and older whose care plan was shared with their PCP within 30 days of development or update. This measure does not require documentation of the care plan's transmission method. Testing results indicate marked variation in care plan sharing practices. The refined measure specification that focuses on sharing of information with the PCP should represent a reasonable benchmark that plans can work to meet over time. As performance improves, it may be useful to add back the requirement for documentation of the method for sharing the care plan; however, we do not recommend doing so at this time.

Testing results primarily highlighted the overall sub-optimal performance even for the simplified measure (all but two plans had rates of zero). However, testing results suggested the existence of meaningful variation (one plan with meaningful and significantly higher performance). Participants in the TEP, the workgroup and public commenters all supported moving forward with this measure.

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 measures, and an outcome measure would better serve this population. The measurement team agrees that outcome measures are a long-term goal, but before useful data on outcomes can be collected, organizations must assess and document 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 in care plans - including non-medical needs.

Additional Face Validity Feedback
Stakeholder input suggests 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** Click here to enter number of factors **risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

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

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

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

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

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

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 on the measure is presented in Table 6. Health plan 01 demonstrated rates that differed significantly from the mean at the .05 level.

Table 6. Performance rates by health plans with significant difference noted

Health Plan	Care Plan Shared with PCP
HP 01	**23.4***
HP 02	**0.0**
HP 03	**9.2**
HP 04	**0.0**
HP 05	**0.0**
Minimum	**0.0**
Mean	**6.5**
Maximum	**23.4**
Standard deviation	**10.2**

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.

*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 poor overall performance, health plan 01 had performance rate that demonstrated 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.

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

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

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

Not applicable. There is only one set of specifications for this measure.

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 sharing of a care plan with the primary care provider. 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. The extent of missing data for key data 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 regarding practices for sharing care plans with primary care providers. This measure assesses the percentage of Medicaid MLTSS enrollees who have care plan shared with a primary care provider within 30 days of development or update, and in doing so, should help address this lack of standardization.

Appendix: Additional Testing Data

Methods:

Prior to testing reliability and validity, we assessed the individual elements included in the measure and tested alternative versions of calculating the measure.

1) We analyzed how often new or updated care plans were shared with key LTSS providers and PCPs, and the number of days between when the care plan was first developed or updated and then shared.
2) We analyzed which types of providers were more likely to be sent care plans. The measure as specified prior to beta testing required documentation of sharing the care plan with both the PCP and at least one key LTSS provider.
3) Finally we analyzed who the care plan with was shared by exploring three different possible rates.
    - Original Rate: Percentage of enrollees with a care plan transmitted to the primary care practitioner (PCP) and other key LTSS providers
    - Version 1 - PCP Only: Percentage of enrollees with a care plan transmitted to the primary care practitioner (PCP) with documentation of the method of the care plan's transmission (e.g., fax, email, EHR)
    - Version 2 - PCP only - no requirement of transmission method: Percentage of enrollees with a care plan transmitted to the primary care practitioner (PCP) without any documentation of method of transmission requirement

Results:

Frequency of Shared Care Plan

Approximately 30 percent of the enrollees in the measure's denominator (133 enrollees) had their care plan shared with a PCP or a key LTSS provider at least once in the measurement period (Table 1). Among all the 133 enrollees who had an initial care plan or care plan update shared, 63 percent were shared within 30 days, and most were shared within 10 days.

Table 1. Care plan sharing rates by enrollee type

Enrollees with	Frequency	Percentage of enrollees with a care plan eligible for MLTSS-3 (n =438)
Documentation of a care plan shared with an eligible provider (PCP or Key LTSS)	**133**	**30.4**
Care plan shared within 30 days from its creation	84	19.2
Care plan shared after 30 days of its creation	21	4.8
Data entry error*	28	6.4
No documentation of a care plan shared with an eligible provider	**305**	**69.6**

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

Types of Providers Care Plan was Shared With

We also analyzed which types of providers were more likely to be sent care plans. The measure as specified prior to beta testing required documentation of sharing the care plan with both the PCP and at least one key LTSS provider.

Of the 133 enrollees that had an initial care plan or care plan update shared, 58 percent of the enrollees had the care plan shared with the enrollee's PCP (Table 2), while only 9 percent of these enrollees had the care plan shared with both the PCP *and* one or more LTSS providers. This represents 18 and 3 percent, respectively, of all 438 enrollees in the measure's denominator. Among enrollees with share care plans or care plan updates, "other provider" was the most frequently selected choice (35 percent) for key LTSS providers receiving the care plan, followed by providers of personal care in the home (14 percent).

Regarding the care plan's transmission method, among enrollees with an initial care plan or care plan update shared, approximately 50 percent had the method of sharing (fax, email, EHR) documented in the enrollee's record. This represents only 15 percent of all enrollees in the measure's denominator. The most common method for transmitting a care plan or care plan update was through fax, while the least common was a notification through the electronic health record system.

Table 2. Documented type of provider with whom the care plan or care plan update was shared

Type of provider receiving the shared care plan	Frequency	Percentage of enrollees with a shared care plan (n =133)
Care plan or care plan update shared with Primary Care Practitioner (PCP)	**77**	**57.9**
Care plan or care plan update shared with LTSS Provider		
Other Provider*	47	35.3
Provider of Personal Care in the Home	19	14.2
Provider of Residential and Habilitation Center	3	2.3
Physical or Occupational Therapist	0	0.0
Skilled Nurse	0	0.0
Care plan or care plan update shared with PCP and one LTSS Provider	**12**	**9.0**

*Primarily LTSS provider agencies and adult day programs

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

Health Plan Performance

While our results indicate that there is documentation of components for the majority of care plans that are shared (85 percent), the percentage of care plans actually shared with the PCP was quite low, with only one health plan showing a non-zero measure rate (Health Plan 3) at about 3 percent for the LTSS Shared Care Plan with Primary Care Practitioner Original Rate (Table 2) during beta testing.

Table 2. Performance rates by health plan

Health Plan	Original Rate: Shared with PCP and key LTSS Providers	Version 1: Shared with PCP only, required documentation of method of transmission	Version 2: PCP only - no requirement of transmission method
HP 01	0.0	0.0	23.4
HP 02	0.0	0.0	0.0
HP 03	3.1	7.7	9.2
HP 04	0.0	0.0	0.0

HP 05	0.0	0.0	0.0
Minimum	0.0	0.0	0.0
Mean	0.6	1.5	6.5
Maximum	3.1	7.7	23.4
Standard deviation	1.4	3.4	10.2

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.

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

After discussing these results with advisory panel stakeholders, we determined that plan abstractors found the term "key LTSS provider" subjective and confusing. We then surveyed workgroup members to determine which types of providers should always be sent care plans for MLTSS enrollees. Among the six workgroup members who completed the survey, PCPs ranked highest as the provider type that should always receive the care plan (67 percent), followed by personal care assistants (50 percent), and skilled nurses (33 percent). We also noted, verification of which key LTSS providers enrollees were currently receiving care (and therefore should have their care plan shared with) was not possible in the current measure (i.e., there was no way to identify if the enrollee had a skilled nurse who should be receiving the care plan).

Interpretation:

These results, in combination with challenges associated with defining "key LTSS providers," led to a recommendation to limit the measure to focus only on sharing the care plan with the PCP, without any requirement as to which components or elements must be shared (Version 2 above).

Systematic Assessment of Face Validity: Summaries of Written Feedback

1. Denominator is appropriate given the intent of the measure

The majority of TEP members (12 out of 13, or 92%) agreed that the denominator is appropriate for the measure. One TEP member asked for clarification as to how MLTSS plan enrollees were defined, as LTSS state plan benefits can differ from HCBS waiver services.

2. Numerator Rate is appropriate given the intent of the measure

The majority of TEP members (11 out of 13, or 85%) agreed that the numerator is appropriate for the measure. Comments from the TEP included the suggestion that a summary of the care plan would be more efficient for the PCP, as care plans can be lengthy. While one TEP member noted that for LTSS measures the PCP may not be involved in the execution of the care plan, the measurement team agrees that the PCP should be informed regarding the care plan.

3. Exclusion 1 is appropriate given the intent of the measure

Most TEP members (12 out of 13 or 92%) thought that the preferences of enrollees regarding the sharing of a care plan with the PCP was an appropriate exclusion for this measure. One TEP member noted that this exclusion was consistent with recognizing the rights of enrollees to make decisions regarding their own healthcare.

4. Would high performance on this measure indicate that a health plan is providing higher quality care?

While most of the TEP (8 out of 13, or 62%) responded that good performance on this measure is indicative of high quality care, several TEP members had comments regarding the disposition of care plans once transmitted to the provider. Members noted that sharing a care plan with a PCP was not an indicator that the PCP had reviewed the care plan, or the plan was sent to the appropriate provider, or that members were receiving needed services specified in the plan. While we agree with the members that this measure does not indicate whether the PCP acted on the care plan, sending the care plan is first step of coordination between LTSS and medical care providers and is critical to ensuring PCPs are aware of the care being provided by the LTSS providers. Unfortunately, it is not possible to determine with current documentation whether the care plan was reviewed by the PCP.

Another TEP member questioned if all LTSS services outlined in a care plan are healthcare related and appropriate to share with the provider. To address this concern the measure was specified to allow only a portion of the care plan to be transmitted to the PCP, which will allow for the MLTSS plan to determine which components of the care plan are most relevant for the PCP.

5. In the future, do you think that performance scores on this measure will distinguish between good and poor performance?

The TEP was mostly (7 out of 13, or 54%) supportive of the measure's ability to distinguish between good and poor performance. Commenters noted that the measure was specified as a process measure and does not correlate patient outcomes to a care plan. While the measurement team agrees with the value of outcome measures for this population, MLTSS is an evolving area of measurement and process measures can drive the standardization of data needed to develop outcome measures. TEP members that supported the measure were in agreement that sharing a care plan with the PCP is a first step in improving performance.

6. Do you have any recommendations that would help strengthen the Long Term Services and Supports (LTSS) Shared Care Plan with Primary Care Practitioner measure?

TEP members that were supportive of the measure noted that this was a good first step to ensure that plans were reaching out to the PCP. One TEP member suggested extending the 30-day transmission window to 45 days. One TEP member noted that some percentage of enrollees in managed LTSS plans do not have a primary care physician. One TEP member suggested that there was greater importance for the patient or consumer to receive a copy of the care plan than the PCP.

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 Long Term Services and Supports (LTSS) Shared Care Plan with Primary Care Practitioner 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.

Delivering effective care coordination for complex populations, such as Medicaid MLTSS enrollees, begins with conducting and regularly updating a comprehensive assessment to identify enrollees' needs, developing and regularly updating an individualized care plan to indicate the specific services and supports to be provided, and sharing care plans to inform care team members of services that should be coordinated. State Medicaid agencies have implemented numerous Medicaid MLTSS care coordination models. Most models require the development of a care plan at initial enrollment and on a regular basis thereafter, as well as the use of team based care to implement the care plan. Similarly, numerous other programs that deliver care to individuals who are similar to (or in some cases the same as) Medicaid MLTSS enrollees require care plans and team based care, including patient-centered medical homes, Medicare managed care plans (e.g., Special Needs Plans, Financial Alignment Initiative dual eligible enrollee demonstration plans), state Medicaid home and community-based services 1915(c) waiver programs, the Program for All-Inclusive Care for the Elderly (PACE), and Medicaid Health Homes. In order for team-based care to be effective, providers must share the care plan and communicate changes and updates to the care plan so that all members of the care team have a complete picture of the person's needs, preferences, and services and supports provided.

This measure would address the existing gap in information sharing by assessing the percentage of Medicaid LTSS enrollees who have a care plan developed and shared with their PCP. A standardized measure of care plan sharing 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_Shared_Care_Plan_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.

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 3326
Measure Title: Long Term Services and Supports (LTSS) Re-Assessment/Care Plan Update after Inpatient Discharge
Measure Steward: Centers for Medicare & Medicaid Services, Centers for Medicaid & CHIP Services
Brief Description of Measure: The measure has two rates:

Rate 1 (LTSS Re-Assessment after Inpatient Discharge Rate): The percentage of discharges from inpatient facilities in the measurement year for Medicaid Managed Long Term Services and Supports (MLTSS) Plan enrollees resulting in a LTSS re-assessment within 30 days of discharge.

Rate 2 (LTSS Re-Assessment and Care Plan Update after Inpatient Discharge Rate): The percentage of discharges from inpatient facilities in the measurement year for MLTSS plan enrollees resulting in a LTSS re-assessment and care plan update within 30 days of discharge.

Developer Rationale: This measure addresses continuity of care following a discharge from an acute (i.e. hospital) or non-acute (i.e. nursing facility) inpatient setting for Medicaid MLTSS enrollees. The MLTSS enrollee population includes individuals with complex health and social support needs. Given their complex needs, they require high levels of care coordination. Re-assessment and the updating of a care plan following discharge is a

critical step in ensuring enrollees return to the community with the needed services and supports that address their goals, preferences, and needs, which may have changed following an inpatient admission.

The Technical Expert Panel (TEP) convened in 2013 to advise on these measures originally suggested a measure assessing whether the member care plan is updated after a major change in health status or living situation. Since "major change in health status or living situation" could not be feasibly identified from claims or enrollment records, the measure development team and TEP decided to use a care transition from an acute or post-acute facility as a proxy for major change in health status or living situation. The TEP advised that although the desired result is to ensure that the care plan is revised to reflect changes in the individual's situation, goals and preferences and assessment is necessary to determine such changes. Therefore, the measure that was tested looked for both assessment and care plan update conducted face-to-face after discharge for an unplanned hospitalization or from a post-acute care facility.

Through our field test, we determined that requiring both a re-assessment and care plan update within 30 days would not result in meaningful performance rates (i.e., three out of five plans had a 0% performance rate). Follow-up analysis identified that often re-assessment is done without a care plan update (see Appendix: Additional Testing Results).

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 truly comprehensive care plan updates, we decided to break the measure into two rates:

1) LTSS Re-Assessment after Inpatient Discharge Rate, and
2) LTSS Re-Assessment and Care Plan Update after Inpatient Discharge Rate

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 updated care plans after transitions in care. 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 (LTSS Re-Assessment after Inpatient Discharge Rate): Discharges from inpatient facilities in the measurement year for MLTSS plan enrollees resulting in a LTSS re-assessment within 30 days of discharge.

Rate 2 (LTSS Re-Assessment and Care Plan Update after Inpatient Discharge Rate): Discharges from inpatient facilities in the measurement year for MLTSS plan enrollees resulting in a LTSS re-assessment and care plan update within 30 days of discharge.

Denominator Statement: Acute and non-acute inpatient facility discharges for Medicaid MLTSS enrollees age 18 years and older. The denominator is based on discharges, not enrollees. Enrollees may appear more than once in a sample.

Denominator Exclusions: For Rate 2, enrollees who refuse care planning are excluded.

For both rates:
- Pregnancy-related or other perinatal hospital discharges are excluded.
- Enrollees who refuse re-assessment are excluded.
- Exclude planned hospital admissions from the measure denominator. A hospital stay is considered planned if it meets any of the following criteria:
  - Hospital stays with a principal diagnosis of pregnancy or condition originating in the perinatal period are
  - A principal diagnosis of maintenance chemotherapy (Chemotherapy Value Set).

- A principal diagnosis of rehabilitation (Rehabilitation Value Set).
- An organ transplant (Kidney Transplant Value Set, Bone Marrow Transplant Value Set, Organ Transplant Other Than Kidney Value Set).
- A potentially planned procedure (Potentially Planned Procedures Value Set) without a principal acute diagnosis (Acute Condition Value Set).

Measure Type: Process
Data Source: Claims, Management Data, Other, Paper Medical Records
Level of Analysis: Health Plan

New Measure -- Preliminary Analysis

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 care plan and the outcome of improvement in quality of life.
- There is no systematic review assess assessing the impact of re-assessment and care plan update after a discharge from an inpatient (or other type of facility) on outcomes. Developer conducted a targeted literature review to gather evidence in support of this measure:
  - When individuals with multiple chronic conditions or disabilities do not receive transitional care support, they are more likely to receive duplicative medical services, experience medication errors, and have avoidable re-hospitalizations (Coleman & Berenson, 2004; Arbaje et al., 2014)
  - Poor communication between inpatient and outpatient clinicians, medication changes (both intentional and unintentional), discharge with incomplete diagnostic work-ups and inadequate enrollee understanding of diagnoses, medication, and follow up needs contribute to ineffective care transitions (Rennke et al., 2013)
  - A randomized controlled trial among 750 community-dwelling older adults found that individuals receiving care coordination encouraging "continuity across settings and guidance from a transition coach" experienced a reduction in re-hospitalization at 30 days (8.3 percent versus 11.9 percent, p=0.048) and 90 days (16.7 percent versus 22.5 percent, p=0.04) and lower mean hospital costs ($2058 versus $2546) than controls (Coleman et al., 2006).

- Additional randomized trials found that "nurse-led transition care programs" can reduce preventable readmission rates by up to 56 percent (Parry et al., 2003; Parry et al., 2008; Naylor et al., 2003; Naylor et al., 2004; Naylor, 2003).

- Technical Expert Panel (TEP) convened in 2013 to provide insight into the priority areas for measurement and the usefulness and feasibility of the identified measures for MLTSS plans. The 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.

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?*
- *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*

Guidance from the Evidence Algorithm

Process measure not based on systematic review (Box3) -> Empirical evidence without SR or grading (Box 7) -> Empirical evidence includes all studies in the body of evidence (Box 8) -> High-moderate quality of evidence (box 9) -> Moderate

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

RATIONALE:

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.

This measure addresses continuity of care following a discharge from an acute (i.e. hospital) or non-acute (i.e. nursing facility) inpatient setting for Medicaid MLTSS enrollees. The MLTSS enrollee population includes individuals with complex health and social support needs. Re-assessment and the updating of a care plan following discharge is a critical step in ensuring enrollees return to the community with the needed services and supports that address their goals, preferences, and needs, which may have changed following an inpatient admission.

The Technical Expert Panel (TEP) convened in 2013 to advise on these measures originally suggested a measure assessing whether the member care plan is updated after a major change in health status or living situation. The developer states that "major change in health status or living situation" could not be feasibly identified from claims or enrollment records, therefore care

transition from an acute or post-acute facility was used as a proxy for major change in health status or living situation.

Performance data indicates low rates of re-assessment and care plan update post discharge.

Rate	Rate 1 - Re-assessment only, no face-to-face requirement	Rate 2 - Re-assessment and care plan update, no face-to-face requirement
Mean	22.4%	5.2%
Standard Deviation	12.5%	6.0%
Minimum	7.4%	0%
Maximum	40.0%	14.3%

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 found that older black and Hispanic individuals have higher rates of functional impairment than whites (CBO, 2013).

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 demonstrates a need for improved communication between the developers of the Care Plan and PCP but due to the lack of standardization and variation in the tools and the outcome goals that the plans use, it is difficult to assess how submitting an updated care plan will result in an improved outcomes.

**It simply does not seem that empirical data exists to support the logic model provided by the developer that links a comprehensive care plan to improved outcomes and quality of life. The literature findings support the idea of proper care coordination, which I think is agreeable as it relates to a higher quality and cheaper healthcare system. But does the presence of a care plan actually lead to better care coordination? The connection from a care plan to better care coordination is where the lack of empirical data and the literature review both fall short. I personally agree that a co-designed care plan that is revisited post-discharge would improve care coordination and transitions, but there is a lack of empirical evidence to support that fact.

And does revisiting the care plan after an inpatient discharge actually improve outcomes? Some outcomes might include better adherence to the care plan, or improved communication with primary care providers. If the care plan were to be updated a majority of the time after discharge - it be evidence to the fact that post-discharge is an effective point in the care continuum to re-visit a care plan.
But does continuous re-evaluation of a care plan after an inpatient discharge actually improve some of these outcomes? The empirical evidence is limited in supporting this idea.
**According to the review there was evidence to support the measure although there were several related references.

1b. Performance Gap

Comments

**Disparities in Care were not reported. However, the review demonstrated performance gap in care and opportunity for improvement.
**The performance data provided by the developer does show a gap in the regular re-evaluation of and care plan updates post-discharge. There is no disparities data provided, but the performance gap is large overall, so any subsets of the population would be assumed to be low also.
But again, the question is: if a health plan is scoring 100% on re-assessment after discharge does that have a direct impact on adherence to the care plan or physician communication? Or are there other indicators out there for major changes in health status or living situation?
**I question if this measure is a reflection of MLTSS quality or the outcomes of an acute or sub-acute stay. The information provided indicates poor performance with completing care plans post discharge.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing
2b. Validity: Testing; Exclusions; Risk-Adjustment; Meaningful Differences; Comparability Missing Data
2c. For composite measures: empirical analysis support composite approach

Reliability

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

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

Validity

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

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

Complex measure evaluated by Scientific Methods Panel? ☐ **Yes** ☒ **No**
Evaluators: NQF Staff

Evaluation of Reliability and Validity (and composite construction, if applicable):
Link A

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.)?*
- *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**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

2a1. Reliability- Specifications
Comments
**Likelihood of measure being consistently implemented is low. Without standardized data or tools there is no way to evaluate the impact on the desired outcomes.
**It seems that all data elements are clearly defined here. I am a bit confused as to why the discharge date is only slightly reliable when considering the key data elements. Why would this slump in reliability exist? Seems like a pretty important element to have reliable.

2a2. Reliability- Testing
Comments
**meaningful differences about quality not accessed.
**It seems that reliability drops when tested at the measure rates.

2b1. Validity—Testing
2b4-7. Threats to Validity
2b4. Meaningful Differences
Comments
**Missing data and variation in data sources constitute a moderate threat.
**The TEP that was utilized for validity testing even expressed a decrease in agreement when asked about "high performance on this measure indicating a health plan is providing higher quality care."

2b2-3. Other Threats to Validity
2b2. Exclusions
2b3. Risk Adjustment

Comments
**N/A
**Considering the intent of the measure, the exclusions seem appropriate and there is no need for risk adjustment since this is a process measure. Agree that the denominator should be based off discharges and not enrollees.
**This measure is not risk adjusted.

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: Non standardized data elements, many of which are in free text fields.

Committee pre-evaluation comments

Criteria 3: Feasibility

3. Feasibility
Comments
**Low feasibility in the absence of standardized tools
**Due to the unstandardized fields in the electronic health plan and case management records, it may be difficult to extrapolate what is needed. But I agree with the developer that health plans and case management records should start to have structured fields for care plans in the future. I stand at a low-moderate primary rating for feasibility.
**There are questions about the methodology which is used across MLTSS including reporting and data collection methodologies.

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
N/A

Additional Feedback:
N/A

***Questions for the Committee**:*
- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*

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

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
N/A

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

Potential harms
N/A

Additional Feedback
None

***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
**Not currently publicly reported.
**There measure is not currently in use. The biggest unintended benefit in my mind would be another taxation on the health system's bandwidth if it is not providing any benefit (e.g. if care plans are not actually being updated regularly at this point in the care continuum).
I see no plan for implementation beyond the feasibility rational provided.
**There are plans to report this publicly but it is unclear the relevance to the public from my view.

Criterion 5: Related and Competing Measures

Related or competing measures

Related measures include:

- 3319 LTSS Comprehensive Assessment and Update
- 3324 LTSS Comprehensive Care Plan and Update
- 3325 LTSS Shared Care Plan with Primary Care Practitioner

Harmonization

N/A

Committee pre-evaluation comments

Criterion 5: Related and Competing Measures

N/A

Public and member comments

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

Comments:

Morgan Buchko, Meridian Health Plan

From the LTSS reassessment care elements, it seems like this means an LTSS reassessment is performing a new CA. If it is a new CA/HRA entirely, that would be a large lift to complete a new one after every discharge, even considering the exlcusions. There are two rates for this measure: LTSS reassessment after discharge and LTSS reassessment and care plan update after discharge. We are seeking clarification on when a member would fall only into the first rate. If we are completing a new assessment with them, we would update the care plan. The second rate requires the plan of care to have 7 core elements which would be a manual investigation to ensure they are completed in the POC for us or new logic built. Additionally, with the lack of EDT feeds directly from facilities, we anticipate that would be a barrier to completign the 30 day timeframe

- **Zero NQF members have submitted a support/non-support choice.**

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: 3226
Measure Title: Long Term Services and Supports (LTSS) Re-Assessment/Care Plan Update after Inpatient Discharge

RELIABILITY

41. 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.

42. 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)

43. 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)

44. 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)

45. 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 using Inter-class correlation coefficient (ICC) was used to determine agreement of reliability of measure performance rates.

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

46. **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)

Results of reliability of both rates (re-assessment and assessment and care plan) after inpatient discharge showed only fair agreement. Developer suggests result may be

caused by low numbers in discharges that were followed by a re-assessment or care plan update.

47. Was other reliability testing reported?
    ☒Yes (go to Question #8)
    ☐No (rate Question #11: OVERALL RELIABILITY as LOW and proceed to the VALIDITY SECTION)

48. 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)

49. 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 to evaluate inter-rater reliability of data elements.
    ☐No (if no, please explain below and rate Question #10 as INSUFFICIENT)

50. **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)
    3 of the 4 data elements had substantial to moderate reliability, however discharge date was only slight.

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

65. 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]

66. 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)

67. 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)

68. 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.

69. 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)

70. 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

71. Was <u>empirical</u> 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)

72. Was <u>face validity</u> 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)

73. **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	0	0
Agree	8	8
Disagree	4	4
Strongly Disagree	1	1
No Response	0	0
Total Percent Agree	62%	62%

74. 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)

75. 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)

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

76. **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

77. Was other validity testing reported?
☐Yes (go to Question #14)
☐No (please explain below and rate Question #17: OVERALL VALIDITY as LOW)

78. 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)

79. 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)

80. **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).

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

TIPS: Consider the following: Do the component measures fit the quality construct? Are the objectives of parsimony and simplicity achieved while supporting the quality construct?

☐High

☐Moderate

☐Low (please explain below)

☐Insufficient (please explain below)

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** TBD

Measure Title: LTSS Re-Assessment/Care Plan Update After Inpatient Discharge

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

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: Click here to name the health outcome

☐Patient-reported outcome (PRO): Click here to name the 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*): Click here to name the intermediate outcome

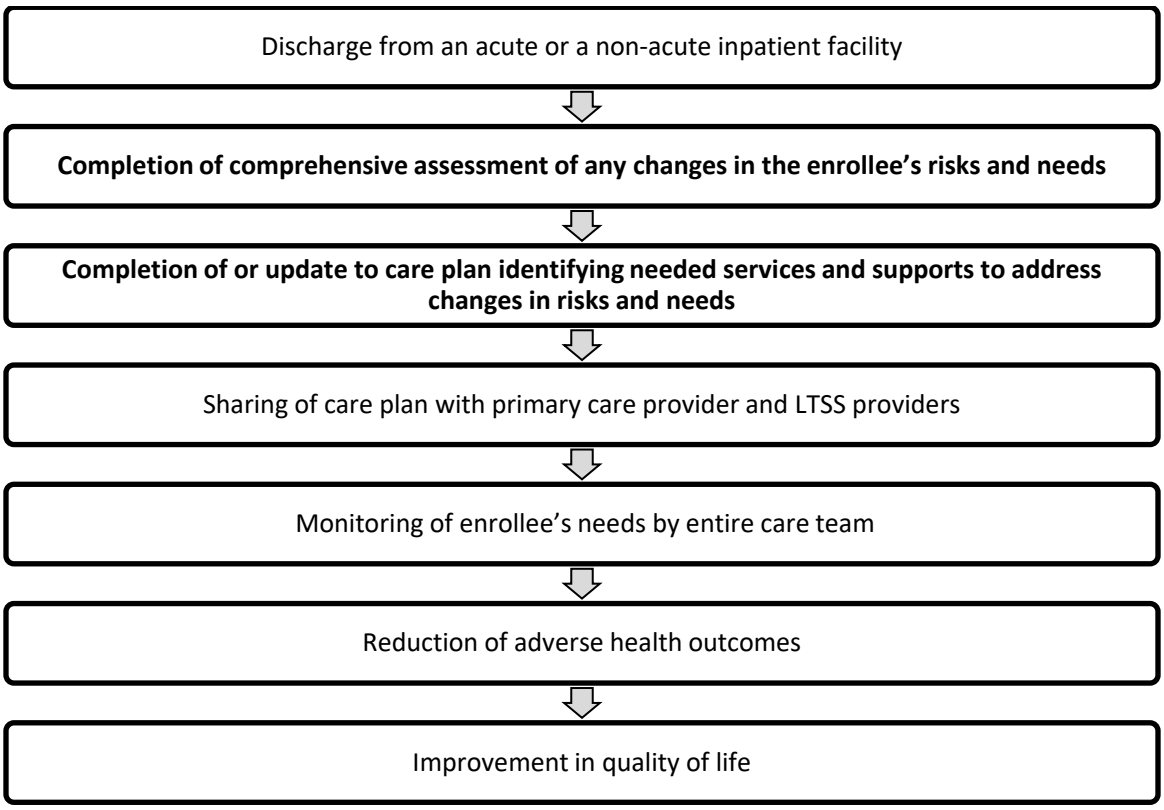
☒ Process: The rate of discharges from acute and non-acute inpatient facilities for MLTSS enrollees age 18 and older that were followed by a reassessment within 30 days.

☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

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, 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, they often receive care from multiple providers and settings (Ujvari et al., 2015). MLTSS enrollees are also more likely to be Medicare-Medicaid enrollees, whose benefits are not aligned (MACPAC, 2014; Saucier & Burwell, 2015). As a result, they often experience highly fragmented care and are at risk for numerous adverse health care utilization patterns and outcomes, including hospitalizations and readmissions (Ujvari et al., 2015; Toles et al., 2012; Naylor et al., 2009; Saucier & Burwell, 2015; Freedman & Spillman, 2014; Allen et al., 2014; Komisar et al., 2005; Sands et al., 2006; Gaugler et al., 2007).

To adequately meet their needs, MLTSS enrollees require high levels of care coordination (Saucier & Burwell, 2015). Effective care coordination for complex populations, such as MLTSS enrollees, begins with conducting and regularly updating comprehensive assessments to identify enrollees' needs, developing and regularly updating care plans to indicate the services and supports to be provided, and sharing care plans to inform care team members about the individual's needs, preferences, care goals, and services to be coordinated (Rich et al., 2012). CMS and other care coordination experts agree that service decisions for MLTSS enrollees should be based on current assessments and fully developed care plans, particularly during care transitions (Ujvari et al., 2015; CMS, 2013).

Evidence to Support Care Transition Interventions from Hospital to Home

We were unable to find a systematic review assessing the impact of re-assessment and care plan update after a discharge from an inpatient or other type of facility on outcomes. However, there is extensive evidence to support the importance to care quality of interventions that assess or re-assess patient care needs and development of a new or updated care plan following a transition of care from hospital to home.

When individuals with multiple chronic conditions or disabilities do not receive transitional care support, they are more likely to receive duplicative medical services, experience medication errors, and have avoidable re-hospitalizations (Coleman & Berenson, 2004; Arbaje et al., 2014). Discharge from an inpatient facility can be followed by multiple care setting transitions in a short period of time, for example to a rehabilitation facility, a nursing facility, and potentially a hospital readmission. Each transition risks a disruption in the enrollee's care. Poor communication between inpatient and outpatient clinicians, medication changes (both intentional and unintentional), discharge with incomplete diagnostic work-ups and inadequate enrollee understanding of diagnoses, medication, and follow up needs contribute to ineffective care transitions (Rennke et al., 2013).

Transitions of care interventions such as risk assessment, transition plans, timely follow-up, and self-management support have been shown in numerous studies to reduce hospital readmissions and lower overall healthcare costs (Coleman et al., 2006). One meta-analysis including 18 studies among patients with congestive heart failure demonstrated that comprehensive discharge planning and post-discharge support reduced readmission rates by 25 percent (Epstein, 2009; Phillips et al., 2004). A randomized controlled trial among 750 community-dwelling older adults found that individuals receiving care coordination encouraging "continuity across settings and guidance from a transition coach" experienced a reduction in re-hospitalization at 30 days (8.3 percent versus 11.9 percent, p=0.048) and 90 days (16.7 percent versus 22.5 percent, p=0.04) and lower mean hospital costs ($2058 versus $2546) than controls (Coleman et al., 2006). Additional randomized trials found that "nurse-led transition care programs" can reduce preventable

readmission rates by up to 56 percent (Parry et al., 2003; Parry et al., 2008; Naylor et al., 2003; Naylor et al., 2004; Naylor, 2003).

A number of care transition models have been developed and implemented in the past decade, such as the Transitional Care Model (Naylor et al., 2003), Care Transitions Program (Coleman et al., 2006), Project RED (Berkowitz et al., 2013), and Project BOOST (Hansen et al., 2013), in an effort to avoid or reduce adverse outcomes. Research is ongoing to identify the exact components of these transitional care models that best improve outcomes for at-risk populations (PCORI, 2015).

For MLTSS enrollees, transitions are a particularly vulnerable time due to the level of care they may require in the home following a discharge, such as personal care assistance, home modifications, durable medical equipment, home health services, meal and transportation assistance, and overall coordination of care across providers (Alliance for Home Health Quality and Innovation, 2014). In order to avoid or reduce the risk of readmission to an acute facility, or to a nursing home or other institution, and to ensure continuity of care, it is critical that LTSS providers: 1) know a enrollee is being discharged, 2) proactively assess or reassess any changes in the enrollee's physical, mental, and social health needs, and 3) develop or update a care plan that documents changes in the enrollee goals, preferences, needs, and the services that will be provided to address those needs. This measure will address these critical steps in care coordination for MLTSS enrollees.

Evidence to Support Care Transition Interventions from Non-Acute Settings to Home

While transitions from hospital to home are the focus of many studies and interventions, a large proportion of LTSS enrollees are discharged into post-acute care settings. In 2013, among Medicare enrollees, 20 percent of discharges were to skilled nursing facilities, 4 percent were to inpatient rehabilitation facilities, and 1 percent was to long term care hospitals. This suggests that among older adults and people with disabilities, one-in-four are not discharged directly from the hospital to home but receive care in another acute or non-acute care facility (MACPAC, 2015). The rate of post-acute care use is likely to be higher among Medicaid MLTSS enrollees, including those who are dually eligible, because LTSS enrollees often have more functional limitations than the overall population of Medicare beneficiaries. For example, in 2012, 31 percent of dual eligible enrollees had 3 or more limitations in activities of daily living, three times the rate (10 percent) of non-dual Medicare beneficiaries (MedPAC and MACPAC, 2017).

Transitions from nursing facilities to home can be equally risky for MLTSS enrollees. Many of the same potential risks of hospital to home transitions apply to nursing facility to home transitions, such as poor communication, incomplete transfer of information, inadequate education of patients and their caregivers, limited access to essential services, and the absence of a single point of contact (Naylor & Keating, 2008). Unsuccessful transitions from a nursing facility to home increase the risk of a hospital admission, or re-admission to a nursing facility. A

study in New Jersey of 1,354 long-term nursing home residents who were transitioned to the community found that the highest predictors of nursing home readmission were being male, single, dissatisfied with one's living situation, and falling within eight to 10 weeks after discharge (Howell et al., 2007). The study authors concluded that transition care managers should work one-on-one with nursing facility residents to understand their unique needs and situations and identify where particular services, such as falls risk prevention programs, are necessary.

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. Under the current contract in 2016, we convened a new TEP (21 members) 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 previous CMS contract (Contract No. HHSM-500-2010-00026I/HHSM-500-T0011).

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

Allen, S. M., E. R. Piette, & V. Mor. (2014). The Adverse Consequences of Unmet Need Among Older Persons Living in the Community: Dual-Eligible Versus Medicare-Only Enrollees. *The Journals of Gerontology: Psychological Sciences*, 69(1), S51-S58.

Alliance for Home Health Quality and Innovation. (2014). Improving Care Transitions Between Hospital and Home Health: A Home Health Model of Care Transitions. Available at http://ahhqi.org/images/uploads/AHHQI_Care_Transitions_Tools_Kit_r011314.pdf.

Arbaje, A.I., Kansagara, D.L., Salanitro, A.H., Englander, H.L., Kripalani, S., Jencks, S.F., Lindquist, L.A. (2014). Regardless of Age: Incorporating Principles from Geriatric Medicine to Improve Care Transitions for patients with Complex Needs. *Journal of General Internal Medicine,* 29(6), 932-939.

Berkowitz, R. E., Fang, Z., Helfand, B. K., Jones, R. N., Schreiber, R., & Paasche-Orlow, M. K. (2013). Project ReEngineered Discharge (RED) lowers hospital readmissions of patients discharged from a skilled nursing facility. *Journal of the American Medical Directors Association*, *14*(10), 736-740.

Centers for Medicare & Medicaid Services (CMS). (2013). Guidance to States using 1115 Demonstrations or 1915(b) Waivers for Managed Long Term Services and Supports Programs. Available at https://www.medicaid.gov/medicaid-chip-program-information/by-topics/delivery-systems/downloads/1115-and-1915b-mltss-guidance.pdf.

CMS. 2015. Medicaid Managed Care Enrollment and Program Characteristics, 2014. Mathematica Policy Research, prepared for CMS. https://www.medicaid.gov/medicaid-chip-program-information/by-topics/data-and-systems/medicaid-managed-care/downloads/2014-medicaid-managed-care-enrollment-report.pdf.

Coleman, E.A., Berenson, R.A. (2004). Lost in Transition: Challenges and Opportunities for Improving the Quality of Transitional Care. *Annals of Internal Medicine,* 141(7), 533-536.
Coleman, E.A., Parry, C., Chalmers, S., et al. (2006). The Care Transitions Intervention: Results of A Randomized Controlled Trial. *Arch Intern Med*.166(17):1822-8.
Epstein, A.M. (2009). Revisiting Admissions – Changing the Incentives for Shared Accountability. *New England Journal of Medicine*. 360(14)1457-59.
Freedman, V. & B. C. Spillman. (2014). Disability and Care Needs Among Older Americans. *The Milbank Quarterly*, 92(3), 509-541.
Gaugler, J. E., S. Duval, K. A. Anderson, & R. L. Kane. (2007). Predicting Nursing Home Admission in the U.S.: A Meta-Analysis." *BMC Geriatrics*, 7(1), 1.
Hansen, L. O., Greenwald, J. L., Budnitz, T., Howell, E., Halasyamani, L., Maynard, G., ... & Williams, M. V. (2013). Project BOOST: effectiveness of a multihospital effort to reduce rehospitalization. *Journal of Hospital Medicine*, *8*(8), 421-427.
Howell, S., et al. (2007). Determinants of remaining in the community after discharge: Results from New Jersey's nursing home transition program. *The Gerontologist*, 47(4), 535-547.
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/.
Komisar, H. L, J. Feder, & J. D. Kasper. (2005). "Unmet Long-Term Care Needs: An Analysis of Medicare-Medicaid Dual Eligibles." *Inquiry*, (42)2, 171-182.
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/.
MACPAC. (2014). Report to the Congress on Medicaid and CHIP. Chapter 2. Medicaid's Role in Providing Assistance with Long-Term Services and Supports.
MACPAC. (2015). Chapter 7. Medicare's post-acute care: Trends and ways to rationalize payments. Available at http://www.medpac.gov/documents/reports/chapter-7-medicare's-post-acute-care-trends-and-ways-to-rationalize-payments-(march-2015-report).pdf?sfvrsn=0.
MedPAC and MACPAC. (2017). Data Book: Beneficiaries Dually Eligible for Medicare and Medicaid. Available at https://www.macpac.gov/wp-content/uploads/2017/01/Jan17_MedPAC_MACPAC_DualsDataBook.pdf.
Naylor, M. D., E. T. Kurtzman, & M. V. Pauly. (2009). Transitions of Elders Between Long-Term Care and Hospitals." *Policy, Politics, and Nursing Practice*, 10(3), 187-194.
Naylor, M., & S. A. Keating. (2008). Transitional care: moving patients from one care setting to another. *The American Journal of Nursing*, 108(9 Suppl), 58.
Naylor, M.D. (2003). Transitional Care of Older Adults. *Annual Review of Nursing Research*. 20:127-47.
Naylor, M.D., Brooten, D.A., Campbell, R., et al. (2003). Comprehensive Discharge Planning and Home Follow-Up of Hospitalized Elders. *Journal of the American Medical Association.* 281:613-20.
Naylor, M.D., Brooten, D.A., Campbell, R.L., Maislin, G., McCauley, K.M., Schwartz, J.S. (2004). Transitional Care of Older Adults Hospitalized with Heart Failure: A Randomized, Controlled Trial. *Journal of the American Geriatrics Society*. 52:675-84.
Parry, C., Coleman, E.A., Smith, J.D., Frank, J., Kramer, A.M. (2003). The Care Transitions Intervention: A Patient-Centered Approach to Ensuring Effective Transfers Between Sites of Geriatric Care. *Home Health Care Services Quarterly*. 22(3):1-17.
Parry, C., Mahoney, E., Chalmers, S.A., Coleman, E.A. (2008) Assessing the Quality of Transitional Care: Further Applications of the Care Transitions Measure. *Medical Care,* 46(3), 317-322.
Patient-Centered Outcomes Research Institute (PCORI). (2015). Project ACHIEVE (Achieving Patient-Centered Care and Optimized Health in Care Transitions by Evaluating the Value of Evidence.

Available at http://www.pcori.org/research-results/2014/project-achieve-achieving-patient-centered-care-and-optimized-health-care.
Phillips, C.O., Wright, S.M., Kern, D.E., Singa, R.M., Shepperd, S., Rubin, H.R. (2004). Comprehensive Discharge Planning with Post Discharge Support for Older Patients with Congestive Heart Failure: A Meta-Analysis. *Journal of the American Medical Association*. 291:1358-67.
Rennke, S., Nguyen, O.K., Shoeb, M.H., Magan, Y., Wachter, R.M., Ranji, S.R. (2013). Hospital-Initiated Transitional Care Interventions as a Patient Safety Strategy: A Systematic Review. *Annals of Internal Medicine,* 158(5, Part 2), 433-440.
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. Rockville, MD: 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.
Sands, L. P., Y. Wang, G. P. McCabe, K. Jennings, C. Eng, & K. E. Covinsky. (2006). Rates of Acute Care Admissions for Frail Older People Living with Met Versus Unmet Activity of Daily Living Needs. *Journal of the American Geriatrics Society*, 53(2), 339-344.
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.
Toles, M. P., Abbott, K. M., Hirschman, K. B., & Naylor, M. D. (2012). Transitions in Care among Older Adults Receiving Long Term Services and Supports. *Journal of Gerontological Nursing*, *38*(11), 40–47. http://doi.org/10.3928/00989134-20121003-04
Ujvari, K., W. Fox-Grage, & L Hendrickson. (2015). Effective transitions between settings. Washington, DC: AARP Public Policy Institute. Available at http://longtermscorecard.org/~/media/Microsite/Files/2015/AARP987_EffectiveCareTransitions_June2015.pdf.

NATIONAL QUALITY FORUM

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 #: 3326
Corresponding Measures:
De.2. **Measure Title:** Long Term Services and Supports (LTSS) Re-Assessment/Care Plan Update after Inpatient Discharge
Co.1.1. **Measure Steward:** Centers for Medicare & Medicaid Services, Centers for Medicaid & CHIP Services
De.3. **Brief Description of Measure:** The measure has two rates:

Rate 1 (LTSS Re-Assessment after Inpatient Discharge Rate): The percentage of discharges from inpatient facilities in the measurement year for Medicaid Managed Long Term Services and Supports (MLTSS) Plan enrollees resulting in a LTSS re-assessment within 30 days of discharge.

Rate 2 (LTSS Re-Assessment and Care Plan Update after Inpatient Discharge Rate): The percentage of discharges from inpatient facilities in the measurement year for MLTSS plan enrollees resulting in a LTSS re-assessment and care plan update within 30 days of discharge.
1b.1. **Developer Rationale:** This measure addresses continuity of care following a discharge from an acute (i.e. hospital) or non-acute (i.e. nursing facility) inpatient setting for Medicaid MLTSS enrollees. The MLTSS enrollee population includes individuals with complex health and social support needs. Given their complex needs, they require high levels of care coordination. Re-assessment and the updating of a care plan following discharge is a critical step in ensuring enrollees return to the community with the needed services and supports that address their goals, preferences, and needs, which may have changed following an inpatient admission.

The Technical Expert Panel (TEP) convened in 2013 to advise on these measures originally suggested a measure assessing whether the member care plan is updated after a major change in health status or living situation. Since "major change in health status or living situation" could not be feasibly identified from claims or enrollment records, the measure development team and TEP decided to use a care transition from an acute or post-acute facility as a proxy for major change in health status or living situation. The TEP advised that although the desired result is to ensure that the care plan is revised to reflect changes in the individual's situation, goals and preferences and assessment is necessary to determine such changes. Therefore, the measure that was tested looked for both assessment and care plan update conducted face-to-face after discharge for an unplanned hospitalization or from a post-acute care facility.

Through our field test, we determined that requiring both a re-assessment and care plan update within 30 days would not result in meaningful performance rates (i.e., three out of five plans had a 0% performance rate). Follow-up analysis identified that often re-assessment is done without a care plan update (see Appendix: Additional Testing Results).

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 truly comprehensive care plan updates, we decided to break the measure into two rates:

1) LTSS Re-Assessment after Inpatient Discharge Rate, and
2) LTSS Re-Assessment and Care Plan Update after Inpatient Discharge Rate

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 updated care plans after transitions in care. 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 (LTSS Re-Assessment after Inpatient Discharge Rate): Discharges from inpatient facilities in the measurement year for MLTSS plan enrollees resulting in a LTSS re-assessment within 30 days of discharge.

Rate 2 (LTSS Re-Assessment and Care Plan Update after Inpatient Discharge Rate): Discharges from inpatient facilities in the measurement year for MLTSS plan enrollees resulting in a LTSS re-assessment and care plan update within 30 days of discharge.

S.6. **Denominator Statement:** Acute and non-acute inpatient facility discharges for Medicaid MLTSS enrollees age 18 years and older. The denominator is based on discharges, not enrollees. Enrollees may appear more than once in a sample.

S.8. **Denominator Exclusions:** For Rate 2, enrollees who refuse care planning are excluded.

For both rates:
- Pregnancy-related or other perinatal hospital discharges are excluded.
- Enrollees who refuse re-assessment are excluded.
- Exclude planned hospital admissions from the measure denominator. A hospital stay is considered planned if it meets any of the following criteria:
  - Hospital stays with a principal diagnosis of pregnancy or condition originating in the perinatal period are
  - A principal diagnosis of maintenance chemotherapy (Chemotherapy Value Set).
  - A principal diagnosis of rehabilitation (Rehabilitation Value Set).
  - An organ transplant (Kidney Transplant Value Set, Bone Marrow Transplant Value Set, Organ Transplant Other Than Kidney Value Set).
  - A potentially planned procedure (Potentially Planned Procedures Value Set) without a principal acute diagnosis (Acute Condition Value Set).

De.1. **Measure Type:** Process
S.17. **Data Source:** Claims, 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:

De.4. **IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** This measure is not currently paired or grouped.

1. Evidence, Performance Gap, Priority – 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_ReAssess_CarePlan_Update_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.

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.
This measure addresses continuity of care following a discharge from an acute (i.e. hospital) or non-acute (i.e. nursing facility) inpatient setting for Medicaid MLTSS enrollees. The MLTSS enrollee population includes individuals with complex health and social support needs. Given their complex needs, they require high levels of care coordination. Re-assessment and the updating of a care plan following discharge is a critical step in ensuring enrollees return to the community with the needed services and supports that address their goals, preferences, and needs, which may have changed following an inpatient admission.

The Technical Expert Panel (TEP) convened in 2013 to advise on these measures originally suggested a measure assessing whether the member care plan is updated after a major change in health status or living situation. Since "major change in health status or living situation" could not be feasibly identified from claims or enrollment records, the measure development team and TEP decided to use a care transition from an acute or post-acute facility as a proxy for major change in health status or living situation. The TEP advised that although the desired result is to ensure that the care plan is revised to reflect changes in the individual's situation, goals and preferences and assessment is necessary to determine such changes. Therefore, the measure that was tested looked for both assessment and care plan update conducted face-to-face after discharge for an unplanned hospitalization or from a post-acute care facility.

Through our field test, we determined that requiring both a re-assessment and care plan update within 30 days would not result in meaningful performance rates (i.e., three out of five plans had a 0% performance rate). Follow-up analysis identified that often re-assessment is done without a care plan update (see Appendix: Additional Testing Results).

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 truly comprehensive care plan updates, we decided to break the measure into two rates:

1) LTSS Re-Assessment after Inpatient Discharge Rate, and
2) LTSS Re-Assessment and Care Plan Update after Inpatient Discharge Rate

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 updated care plans after transitions in care. 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 elderly, 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) (integrated care), Dual-Special Needs Plan (D-SNP), or Fully Integrated Dual Eligible (FIDE) SNP. 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.

The data shown below demonstrating the proportion of beneficiaries with re-assessment only after discharge and the proportion beneficiaries who receive an assessment and care plan update after discharge, aligns with feedback from alpha testing, which indicates that care coordinators only conduct a care plan update if the nature of the discharge warrants it.

Rate 1. Re-assessment only, no face-to-face requirement
Mean: 22.4 %
Standard Deviation: 12.5%
Minimum: 7.4%
Maximum: 40.0%

Rate 2. Re-assessment and care plan update, no face-to-face requirement
Mean: 5.2%
Standard Deviation: 6.0%
Minimum: 0%
Maximum: 14.3%

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**
Not applicable.

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 and the development and sharing of care plans among the MLTSS enrollee population post hospitalization. However, most other research focuses on the identification of disparities in the need for and use of LTSS more broadly, which highlight MLTSS enrollees' vulnerabilities during care transitions.

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 (CBO, 2013).

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

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)*
Attachment **Attachment:** LTSS_ReAssess_CarePlan_Exclusions_Value_Set.xlsx

S.2c. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.
No, this is not an instrument-based measure **Attachment:**

S.2d. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.
Not an instrument-based measure

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

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

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The measure has two rates. The numerators for the two rates are as follows:

Rate 1 (LTSS Re-Assessment after Inpatient Discharge Rate): Discharges from inpatient facilities in the measurement year for MLTSS plan enrollees resulting in a LTSS re-assessment within 30 days of discharge.

Rate 2 (LTSS Re-Assessment and Care Plan Update after Inpatient Discharge Rate): Discharges from inpatient facilities in the measurement year for MLTSS plan enrollees resulting in a LTSS re-assessment and care plan update within 30 days of discharge.

S.5. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The numerator details for the two rates are as follows.

Rate 1 (LTSS Re-Assessment after Inpatient Discharge Rate): Discharges from inpatient facilities in the measurement year for MLTSS plan enrollees resulting in a LTSS re-assessment within 30 days of discharge.

LTSS Re-Assessment: The LTSS re-assessment must include a discussion with the enrollee using a structured or semi-structured tool that assesses the status of, and any changes to, the enrollee's health status and needs. The LTSS re-assessment must document current enrollee status on nine (9) core elements. The date of the LTSS re-assessment must be documented.

LTSS Re-Assessment 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)(e.g., concentration, memory, problem solving abilities)
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)mood, affect, anxiety)
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 (e.g., home, nursing facility, assisted living).
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 (LTSS Re-Assessment and Care Plan Update after Inpatient Discharge Rate) Discharges from inpatient facilities in the measurement year for MLTSS plan enrollees resulting in a LTSS re-assessment (described above) AND care plan update within 30 days of discharge (described below).

LTSS Care Plan Update: The LTSS care plan is a document or electronic record which identifies enrollee needs, preferences, and risks, and contains a list of the services and supports planned to meet those needs while reducing risks. There must be documentation that the care plan update after discharge was completed with input from the enrollee during a discussion between the individual responsible for creating the care plan (care manager) and enrollee. The LTSS re-assessment and care plan update after discharge may be done during the same encounter or during different encounters. A care plan update may be called a service plan update in certain Medicaid MLTSS plans. Per its definition, the LTSS care plan update must include:
- Documentation on whether family or friend caregiver(s) were involved in the development of the care plan, and the contact information for said caregiver(s). If there is no friend or family caregiver involved in care-planning, the lack of informal caregiver availability must be documented to meet this element.
- Documentation of enrollee (or power of attorney) agreement to comprehensive care plan, or appeal of care plan. Documentation of agreement includes: verbal agreement from the enrollee, or power of attorney (POA), received by phone or in person OR written agreement from the enrollee, or POA, received by mail (e.g., a signature). Documentation that a care plan was discussed or reviewed is not sufficient to meet this measure. The documentation must indicate that the enrollee (or POA) agreed to the care plan or the care plan is being appealed.

The care plan update after discharge must include documentation of seven (7) core elements:

LTSS Care Plan Update Core Elements
1. Care planned to meet enrollee medical needs. Documentation must include either plan for addressing need or documentation of no need.

2. Care planned to meet enrollee functional needs. Documentation must include either plan for addressing need or documentation of no need.
3. Care planned to meet enrollee needs due to cognitive impairment or documentation of no cognitive impairment. Example of care to meet cognitive impairment needs includes support for behavioral difficulties, caregiver support or education to address cognitive impairment, or support for keeping individual cognitive engaged in activities. Documentation must include either plan for addressing need related to cognition (or cognitive impairment/dementia) or documentation of no need.
4. List of all LTSS services and supports the enrollee receives, or is expected to receive in the next month, in the home (paid or unpaid) or in other settings (e.g., adult day health center, nursing facility), including amount (e.g., hours, days) and frequency (e.g., every day, once a week). Documentation of no LTSS services is sufficient to meet the numerator criteria.
5. At least one enrollee (and family as appropriate) individualized goal (medical or non-medical goals).
6. A plan for follow-up and communication with the care manager (i.e., documentation of follow-up and communication schedule with care manager)
7. Plan for ensuring enrollee needs are met if an emergency occurs (e.g., if a personal care assistant or home health aide is unable to get to home, natural disaster). Must include at a minimum the name of an individual at the MLTSS plan or contracted provider to contact in case of an emergency.

S.6. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Acute and non-acute inpatient facility discharges for Medicaid MLTSS enrollees age 18 years and older. The denominator is based on discharges, not enrollees. Enrollees may appear more than once in a sample.

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 all qualifying discharges from acute and non-acute inpatient facilities (e.g., hospitals, skilled nursing facilities, inpatient rehabilitation, custodial nursing facilities, inpatient psychiatric care facilities) between January 1 and December 1 of the measurement year for enrollees who meet the following criteria:
- 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 30 days after the qualifying discharge.
- Who have either of the following benefits: 1) long-term services and supports: home and community based or 2) long-term services and supports: institutional care AND 3) medical care: inpatient care.

The time frame for the denominator allows for 30 days to conduct the LTSS re-assessment and care plan update in the measurement year. The denominator for this measure is based on discharges, not enrollees. If enrollees have more than one discharge, include all discharges in the measurement year.

The MLTSS plan may use its own method to identify discharges from acute or non-Acute inpatient facilities.

S.8. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
For Rate 2, enrollees who refuse care planning are excluded.

For both rates:
- Pregnancy-related or other perinatal hospital discharges are excluded.
- Enrollees who refuse re-assessment are excluded.
- Exclude planned hospital admissions from the measure denominator. A hospital stay is considered planned if it meets any of the following criteria:

- Hospital stays with a principal diagnosis of pregnancy or condition originating in the perinatal period are
- A principal diagnosis of maintenance chemotherapy (Chemotherapy Value Set).
- A principal diagnosis of rehabilitation (Rehabilitation Value Set).
- An organ transplant (Kidney Transplant Value Set, Bone Marrow Transplant Value Set, Organ Transplant Other Than Kidney Value Set).
- A potentially planned procedure (Potentially Planned Procedures Value Set) without a principal acute diagnosis (Acute Condition Value Set).

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.)*
For Rate 2, enrollees who refuse care planning are excluded.

For both rates:
- Pregnancy-related or other perinatal hospital discharges are excluded.
- Enrollees who refuse re-assessment are excluded.
- Exclude planned hospital admissions from the measure denominator. A hospital stay is considered planned if it meets any of the following criteria:
- Hospital stays with a principal diagnosis of pregnancy or condition originating in the perinatal period are
- A principal diagnosis of maintenance chemotherapy (Chemotherapy Value Set).
- A principal diagnosis of rehabilitation (Rehabilitation Value Set).
- An organ transplant (Kidney Transplant Value Set, Bone Marrow Transplant Value Set, Organ Transplant Other Than Kidney Value Set).
- A potentially planned procedure (Potentially Planned Procedures Value Set) without a principal acute diagnosis (Acute Condition Value Set).

See attachment: LTSS ReAssess CarePlan Exclusions Value Sets.

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 of qualifying discharges.

Step 1b. From the eligible population, draw a systematic sample of discharges that occur between January 1 and December 1 of the measurement year.
Step 1c. Exclude discharges for planned admissions and pregnancy-related or other perinatal hospital stays. Exclude discharges where the enrollee refused LTSS re-assessment.

Rate 1 (LTSS Re-Assessment after Inpatient Discharge Rate):
Step 1d. From the remaining qualifying discharges, identify through documentation in the medical or care management record if the enrollee had a LTSS re-assessment within 30 days of the qualifying discharge date.
Step 1e. Divide the number of discharges in Step 1d by the number of discharges in Step 1c to calculate the rate.

Rate 2 (LTSS Re-Assessment and Care Plan Update after Inpatient Discharge Rate):
Step 2a. From discharges identified in Step 1d, exclude discharges where the enrollee refused care planning.
Step 2b. Identify through documentation in the medical or care management record if the enrollee had a LTSS re-assessment and care plan update within 30 days of discharge.
Step 2c. Divide the number of discharges in Step 2b by the remaining number of discharges in Step 2a to calculate the rate.

S.15. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.
MLTSS plans should identify a systematic sample of 411 discharges for enrollees who meet the eligible population criteria.

S.16. **Survey/Patient-reported data** *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*
Specify calculation of response rates to be reported with performance measure results.
Not applicable.

S.17. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.18.
Claims, 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 re-assessment or care plan update were completed within the required time frame.

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_ReAssess_CarePlan_Update_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

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:** TBD
Measure Title: LTSS Re-Assessment/Care Plan Update After Inpatient Discharge
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** [16] **differences in performance**;

OR

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

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

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

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-

item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

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

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions.
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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

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

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

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

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. Forty-six enrollees were removed from the sample for inconsistent information, and the total sample included in the testing and analysis contained 715 enrollees. The total number of inpatient admissions of enrollees was 354. Of the

354 inpatient admissions and discharges, 35 discharges were excluded from the sample as they were for planned admissions. The remaining sample of 319 discharges were for unplanned or unknown reasons. 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

Characteristics	Percentage of enrollees in the testing sample (n=715)
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

Characteristic	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	

Characteristic	Percentage of enrollees in the testing sample (n=715)
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. The total number of inpatient admissions of enrollees age 18 or older for unplanned or unknown reasons included in the sample was 319.

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 was 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 is on the final measure specifications. Additional detail on assessment domains and data elements documents 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:

$$\widehat{\mathrm{K}} = \frac{\rho_a - \rho_e}{1 - \rho_e}$$

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

Seven of the nine potential data elements (four measure specific elements and five additional elements related to dates, settings, and contact information) 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 type of facility that the patient was discharged from, type of stay and if an assessment took place, ($\hat{\kappa}$=0.6364, 0.6258 and 0.4561). The data element with the lowest reliability was discharge date at $\hat{\kappa}$=0.240.

Table 3. Reliability of key data elements

Measure (elements)	Data element	Kappa statistic	Interpretation
General	Date of Birth	0.8426	Almost Perfect
(5)	Sex	0.8788	Almost Perfect
	Place of Residence	0.4706	Moderate
	Date of First Enrollment	0.7108	Substantial
	Date of First Disenrollment	-0.5052	Less than Chance Agreement
LTSS Re-Assessment and Care Plan Update After Inpatient Discharge*	Discharged from acute/non-acute inpatient facility	0.6364	Substantial
(4)			
	Discharge date	0.0240	Slight
	Type of stay	0.6258	Substantial
	Assessment occurred after discharge	0.4561	Moderate

Source: Mathematica analysis of paired data from 144 MLTSS enrollees, representing five health plans.

Notes: Interpretation of Kappa statistic 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)

*The remainder of the LTSS Re-Assessment and Care Plan Update After Inpatient Discharge measure's elements (i.e., update to care plan) were assessed too infrequently among the 144 paired assessments (<30) to allow for IRR analysis.

Reliability of Measure Rates

ICCs for Rate 1 and Rate 2 are within the range for fair agreement (0.4 to 0.59).

Table 4. Reliability of recommended measure rates

Measure	ICC statistic	Interpretation
LTSS Re-Assessment and Care Plan Update After Inpatient Discharge		

Measure	ICC statistic	Interpretation
Rate 1: Assessment Updated after Discharge, without Consideration for Location	0.3831	Fair
Rate 2: : Assessment and Care Plan Updated after Discharge, without Consideration for Location	0.2326	Fair

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)

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 the data elements had high reliability, discharge date had low reliability.

The results for reliability of the LTSS Re-Assessment/Care Plan Update After Inpatient Discharge measure rate showed only fair agreement. This result may be caused by the relatively low numbers of discharges that were followed by a re-assessment or care plan update.

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 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 measure and the four measures being tested for the MLTSS population in this project:

- LTSS Comprehensive Assessment and Update
- LTSS Comprehensive Care Plan and Update Measure

- LTSS Shared Care Plan with Primary Care Practitioner
- Falls: Screening, Risk-Assessment and Plan of Care to Prevent Future Falls

We examined the correlation of different versions of the five MLTSS measures with each other using the Spearman Rank Correlations. As with this assessment measure, we tested several variations of each of the other MLTSS measures. Although we analyzed validity among all versions of the five measures for ease of review we present results for only the most promising versions of the measures in Table 3 (reliability of key data elements).

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 ReAssessment/Care Plan Update after Inpatient Discharge 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 multistakeholder 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.

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, very few positive, significant relationships were observed. The Spearman Rank Correlation coefficient, $\hat{\rho}$, showed a significant, strong positive relationship between the two rates in *LTSS Re-Assessment and Care Plan Update After Inpatient Discharge* and the three rates in *Falls: Screening, Risk-Assessment and Plan of Care to Prevent Future Falls*.

The remaining relationships were negative, which is likely due to the high prevalence of zero rates across the measures.

Table 5. Correlation of recommended measure rates

Measures	MLTSS-1, Rate 1	MLTSS-1, Rate 2	MLTSS-2, Rate 1	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-4, Rate 1: Assessment Updated after Discharge, w/o consideration for Location	-0.2368	-0.2368	-0.2868	-0.2868	-0.4472	--	0.4104	0.3000	0.4588	NA
MLTSS-4, Rate 2: Assessment & Care Plan Updated after Discharge, w/o consideration for Location	-0.0513	-0.0513	-0.2236	-0.2236	-0.4472	0.4104	--	0.9747**	0.8944*	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 and 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 the voting results from the survey. Overall, the majority of TEP members supported the denominator, numerators, and exclusions for the Long Term Services and Supports (LTSS) Re-Assessment/Care Plan Update after Inpatient Discharge measure. The majority of TEP members also agreed that the measure is reflective of quality and had the potential to distinguish performance.

Table 6

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	Exclusion 2 is appropriate given the intent of the measure	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	1	1	3	0	0
Agree	12	8	9	10	7	8	8
Disagree	0	4	3	2	0	4	4
Strongly Disagree	0	0	0	0	0	1	1
No Response	0	0	0	0	3	0	0
Total Percent Agree	100%	69%	78%	85%	78%	62%	62%

Additional Face Validity Feedback

Feedback from the public comment period was generally supportive of the measure. 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

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, LTSS Re-Assessment and Care Plan Update After Inpatient Discharge Rate 2, which reports the percentage of enrollees with re-assessment and care plan update, with no face-to-face requirement, appears the most useful as a second "aspirational" measure. Health plan performance is lower for Rate 2 relative to Rate 1 (focused on just assessment), 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.

Because all of the MLTSS measures 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, this measure has a strong positive relationship with the Falls: Screening, Risk-Assessment and Plan of Care to Prevent Future Falls measure, but a significant, strong negative relationship

with the *LTSS Comprehensive Assessment and Update, LTSS Comprehensive Care Plan and Update, and LTSS Shared Care Plan with Primary Care Practitioner* measures, reflecting the fact that for one measure three health plans have zero rates, while for the other measure, the other two measures have zero rates. As reporting improves the internal validity of the measures should also improve accordingly.

Systematic Assessment of Face Validity
While the voting results suggest that this is a valid measure, TEP members noted that the presence of an updated care plan without available network providers to provide services may still result in re-admission and assessment is not the only driver of improvement. The measurement team agrees that future measure goals should include indicators that high quality care was provided, but establishing standard practices for assessments and care plans must take place first. 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 in care plans - including non-medical needs.

Additional Face Validity Feedback
Stakeholder input suggests 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*

Thirty-five admissions were planned, and excluded from the sample of 354 admissions.

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*)

N/A

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

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

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

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

2b3.3a. Describe the conceptual/clinical <u>and</u> statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?
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 <u>or</u> 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 <u>2b3.9</u>

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

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

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

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 on the two measure rates are presented in Table 6. Health plans 04 and 05 demonstrated rates that differed significantly from the mean at the .05 level.

Table 6. Performance rates by health plan with significant differences noted

Health Plan	Rate 1: Re-assessment only, no face-to-face requirement	Rate 2: Re-assessment and care plan update, no face-to-face requirement
HP 01	20.5	0.0
HP 02	15.4	7.7
HP 03	7.4	0.0
HP 04	40.0*	3.9
HP 05	28.6	14.3*
Minimum	7.4	0.0
Mean	22.4	5.2
Maximum	40.0	14.3

Health Plan	Rate 1: Re-assessment only, no face-to-face requirement	Rate 2: Re-assessment and care plan update, no face-to-face requirement
Standard deviation	**12.5**	**6.0**

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?*)
Between the two measure rates, we can observe a range in performance across the five health plans. Additionally, health plans 04 and 05 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*)

Not applicable. There is only one set of specifications for this measure.

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 re-assessment and care plan updates after inpatient discharge. 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. The extent of missing data for key data 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 regarding practices for conducting re-assessments and updating care plans after inpatient discharges among MLTSS enrollees. This measure assesses the percentage of discharges from inpatient facilities for Medicaid MLTSS enrollees that result in a re-assessment or care plan update within 30 days, 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)*.
The LTSS ReAssess Care Plan Exclusions Value Sets can be used to identify discharges for planned acute care hospital admissions, and includes CPT codes, which are proprietary. CMS has an existing license agreement with the AMA, covering the use of these codes, as follows:

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 Re-Assessment and Care Plan Update After Inpatient Discharge 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. For MLTSS enrollees, transitions are a particularly vulnerable time due to the level of care they may require in the home following a discharge, such as personal care assistance, home modifications, durable medical equipment, home health services, meal and transportation assistance, and overall coordination of care across providers. In order to avoid or reduce the risk of readmission to an acute facility, or to a nursing home or other institution, and to ensure continuity of care, it is critical that LTSS providers: 1) know a enrollee is being discharged, 2) proactively assess or reassess any changes in the enrollee's physical, mental, and social health needs, and 3) develop or update a care plan that documents changes in the enrollee goals, preferences, needs, and the services that will be provided to address those needs. This measure will address these critical steps in care coordination for MLTSS enrollees.

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_ReAssess_CarePlan_Update_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.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.