Person and Family Centered Care – Post-In-Person Meeting
Additional Information Requests

Measure #/Title/Developer	Summary/Additional Information Requested	Developer Follow-Up
Committee Action: Review Developer information and public comment; discuss any outstanding issues.		
#0420 Pain Assessment and Follow-up CMS/Quality Insights Public Comment Disposition: Recommended/Exception to Evidence	Evidence: Consensus Not Reached; Passed with Exception Request: opportunity to submit additional information regarding benefits to patients of symptom assessment or if there is additional information on if this measures is contributing to quality. Either of these areas may result in a stronger rationale and evidence to support the measure so that the exception would not be needed. Public Comment: It is not obvious if any specification for what a "standard" measure of this is—e.g. is a pain scale (what is your pain on a scale from 1-10) sufficient? Also, it is interesting to think about how this gets operationalized in the context of other efforts to try to mitigate overprescribing of opioids. We agree with the need for assessment of pain and a follow-up plan where pain is present, but it is not clear what is acceptable as a follow-up plan—just a prescription and a plan to reevaluate? Referral to pain specialist, PT, etc?	The developer has updated the measure worksheet to clarify the current measure specification. See attachment submitted by the developer that summarizes the evidence and guidelines provided in support of the rationale.
Staff Notes: As indicated at the in-person meeting and in the staff preliminary analysis, this maintenance measure has had multiple changes/updates over the course of implementation. During the Public Comment period, the developers updated the measure submission/information form to ensure the most current measure description and specifications are reflected.		

Committee Action: Review the clarifying information submitted by Massachusetts General Hospital; re-vote on evaluation criteria where consensus was not reached (Reliability, Use/Usability and Overall Suitability for Endorsement).		
#2958 Informed, Patient Centered (IPC) Hip and Knee Replacement Surgery Massachusetts General Hospital Public Comment Disposition: Consensus Not Reached on Reliability, Use/Usability, and Overall Suitability for Endorsement	1. There was some confusion regarding "exclusions" – suggest reviewing measure submission, specifically the specification to ensure clarity in this area; 2. 2-year data collection timeframe – suggest reviewing/reconsider based on feedback from the Committee or providing additional data to support the timeframe 3. Establishing reliability at the practice level Public Comment: None Received on this measure specifically	See attachment submitted by the developer addressing each of the concerns raised by the committee and the updated measure submission worksheet. In summary: 1. The denominator statement and denominator exclusions were edited to clarify the target population and the exclusions due to missing responses. These changes were carried through in several places where mentioned in the worksheet/application (De3 and S7, S9, S10, S11). 2. We did not have any set time periods in the initial submission. We edited the description of sampling (S.5. and S.20) and added a clear recommendation for the timing of the survey with respect to the timing of the surgery. We also clarified the look back period for sites to collect responses. 3. We have edited the methods and results of the reliability analyses at the practice level to clarify the tests done and to include the correlation results. **Use/Usability** The measure itself is new, but it is based on a patient reported survey has been used by thousands of patients. These questions have been cognitively tested to ensure that they are consistently understood and that answers meaningfully assess patient knowledge and preferences for treatment. We have used the questions proposed in a variety of survey designs: cross-section surveys of adults 40 and older, Medicare beneficiaries known to have had procedures based on claims, and clinical settings in which patients were identified by office staff or via medical records, without any problems.
Staff Notes: In addition to the notes above, the developer provided additional reliability testing results: Specifically, to assess reliability at the practice level, they divided data within each site to samples with a minimum size of 25 and then calculated the % with IPC within each sample.		

The reliability was calculated as variability from site divided by total variability. This is a valid measure of reliability similar to the traditional method of calculation intra-class or intra-rater correlation coefficient (in this case the rater is the site). [See Fleiss 1999] • At the practice level, the total sample size is 26 (site 1 has 1 combination, site 2 has 21 combinations, site 3 has 1 combination and site 4 has 3 combinations (sample 1 vs. 2, 2 vs. 3, 1 vs. 3)) and the results of the correlation analyses were 0.805. • At the practice level, we had 14 groups (so site 1 had 2 samples, site 2 had 7, site 3 had 2 and site 4 had 3) and the reliability was 0.853

Committee Action: Review the clarifying information submitted by UDSMR; re-vote on criteria where consensus was not reached (Performance Gap, Reliability, Validity, Use/Usability, and Overall Suitability for Endorsement).		
#2776/2777/2778: Functional Change: Change in Motor Score, Self-Care, Mobility Long-Term Acute Care (LTAC) Public Comment Disposition: Consensus Not Reached on Performance Gap, Reliability, Validity, Use/Usability, and Overall Suitability for Endorsement	• Discomfort assuming use of FIM in LTAC, not a common instrument currently in this setting; • Data is limited on reliability of MEASURE across LTACs; specifically, an inter-class correlation analysis may assist in determining if the measures discriminate quality at the facility level; • Feasibility, questions were asked to better understand administration and burden in this new setting of care (for the instrument) • Low performance may be a result of small numbers Public Comment: Comments in support of measures derived from the FIM tools. Three comments received in support, all from same commenter and same comment for each of the measures (2776, 2777, 2778)	Updated the measure rationale to include the following statement: In addition, this measure also can be used to measure maintenance or decline in functional status. • #2776 Worksheet and Updated Testing • #2777 Worksheet and Updated Testing • #2778 Worksheet and Updated Testing Performance Gap – See the Testing Submission Forms – the UDS provides mean performance scores for the 39 facilities – and indicates the range exhibits opportunity for improvement. Updated Testing Submission Forms: An updated analysis was completed as requested. For this update, we have used data from a wider date range to allow for a higher number of facilities to be included in the analysis. We used data from 2002-2007, which included 39 facilities. Included in those 39 facilities were 73.039 patients who were used in the analysis. Using the updated data, the developer updated their data element level reliability statistics:

		Motor Score: the new Cronbach Alpha reliability statistic was 0.965. Self-Care: the new Cronbach Alpha reliability statistic was 0.956 Mobility: the new Cronbach Alpha reliability statistic was 0.903 In addition, UDS conducted intra-class correlation testing of the facility score: Motor Score: The ICC was 0.905, p <.001. This high ICC demonstrates that there is very high consistency for the motor measure. Self-Care: The ICC was 0.951, p <.001. This high ICC demonstrates that there is very high consistency for the self-care measure Mobility: The ICC was 0.938, p <.001. This high ICC demonstrates that there is very high consistency for the mobility measure. For each of the measures, UDS also notes: UDS also updated their validity testing with the following results: **Motor Score:** Concurrent Validity: Correlation: Our measure at both admission and discharge was highly correlated with the FIM® total, 0.876 (p < .001), and 0.905 (p <.001), respectively. Linear Regression: For all venues, when comparing our measure at admission and discharge to the respective FIM® totals, the r-square values were extremely high 0.936 (p < .001), and 0.951 (p <.001), respectively. Predictive Validity:

		Functional Gain: When comparing gain in our measure to overall FIM® gain including all items, the correlation was very high, 0.985 (p < 0.001). Discharge Disposition - Community: The logistic regression analysis shows that the gain in our measure has good predictive ability for discharge setting (community), with a C-statistic of 0.765 (95% CI .761 - .768), p < .001. **Self-Care:** Concurrent Validity: Correlation: Our measure at both admission and discharge was highly correlated with the FIM® total, 0.937 (p < .001), and 0.939 (p <.001), respectively. Linear Regression: For all venues, when comparing our measure at admission and discharge to the respective FIM® totals, the r-square values were extremely high 0.878 (p < .001), and 0.882 (p <.001), respectively. Predictive Validity: Functional Gain: When comparing gain in our measure to overall FIM® gain including all items, the correlation was moderate at 0.326 (p < 0.001). Discharge Disposition - Community: The logistic regression analysis shows that the gain in our measure has good predictive ability for discharge setting (community), with a C-statistic of 0.729 (95% CI .726 - .733), p < .001. **Mobility:**

		Concurrent Validity: Correlation: Our measure at both admission and discharge was highly correlated with the FIM® total, 0.761 ($p < .001$), and 0.847 ($p < .001$), respectively. Linear Regression: For all venues, when comparing our measure at admission and discharge to the respective FIM® totals, the r-square values were extremely high 0.936 ($p < .001$), and 0.951 ($p < .001$), respectively. Predictive Validity: Functional Gain: When comparing gain in our measure to overall FIM® gain including all items, the correlation was high, 0.867 ($p < 0.001$). Discharge Disposition - Community: The logistic regression analysis shows that the gain in our measure has good predictive ability for discharge setting (community), with a C-statistic of 0.783 (95% CI .780 - .787), $p < .001$.
Staff Notes: See updated testing results in the above right-hand column. The developer seems to have provided what the Committee requested in order to better understand between facility variation in performance scores (although it is expressed as means in the Testing worksheets and does not indicate any testing of significance), reliability and validity. The one are of concern that remains is if LTACs are using the FIM in daily practice. The developers were able to provide older data (2002 – 2007) from 39 facilities, while more current data (2010-2012) is from 6 facilities – would like to ensure that is not reflective of fewer facilities using the measures over time. This concern could impact both feasibility and usability/use. Staff will reach out to the developer to be prepared to provide information at the September 9th meeting if the Committee has additional questions.		

Re-Consideration Request

Committee Action: Consider request for reconsideration from the developer; vote on determination to re-vote on all NQF criteria; if determination made to re-vote to potentially change the recommendation for endorsement, need to determine if voting in blocks as was done at the in-person meeting (Scale, Global, etc.) or if there are any specific item-level measures that the Committee would like to remove and vote on separately.		
#2967: Home and Community Based Services Experience of Care measures Public Comment Disposition: Not Recommended	• Discussion/rationale on the length of survey, how long it takes to complete and burden on patients/caregivers; • Low survey response rates (impacts ability to discern variation and performance across programs); Understanding feasibility of getting to optimal sample size • Value of having both the global measures and the recommendation measures; • Data needed on extent of exclusions and impact on measurement • Scientific Acceptability: possibility of assessing reliability via alternate means: spearman brown prophecy formula, inter-class correlations; factor analysis • Evidence from cognitive testing – helps with validity Public Comment: 11 of the 21 comments received on the draft report were related to the HCBS EOC measures. HCBS quality measures have been identified as gaps in the current measurement spectrum by the Measures Application Dual Eligibles Measures Application Partnership (MAP) workgroup as well as an NQF Committee convened to develop a measures framework for HCBS. The comments received are in support of the PFCC Standing Committee re-considering the measure submission.	In brief, the Measure Worksheet has been enhanced with the following new information: • Since applying for NQF endorsement, the underlying survey that is the data source for the measures has been granted the CAHPS trademark and, accordingly, renamed the CAHPS® Home- and Community-Based Services Survey. Minor changes to items that were required for consistency with other CAHPS surveys are reflected in the revised worksheet and listed in the first tab of the supplemental file. • We reanalyzed the data using a larger sample that now includes proxy respondents. This improves both the performance gap and reliability results within a reasonable margin. • We took the review committee's suggestion and employed top-box scoring instead of mean scores. This approach is consistent with scoring methods for CAHPS measures and resulted in both more room for improvement on measures and improved reliability. • Statistical analysis of the correlation between global rating and recommendation items which suggests that, which related, they are measuring different constructs. We also note that the latter are typically highly valued by service providers and other stakeholders. Furthermore, the Measure Worksheet has been clarified with the following additional information: • The unit of analysis is the HCBS program and the accountable entity for Medicaid HCBS programs is the "operating" entity. The operating entity manages and oversees the quality in HCBS programs. The state Medicaid agency delegates operating authority to the operating entity, which may be another state

		agency, a non-state government entity, or a managed care organization. Clarification about the accountable entity was added to the Measures Testing form item #1.5. • The small performance gaps for the personal safety-related measures (both the scale and the individual item) reflect the fact that these are essentially "never events," which are not expected to have a substantial performance gap but are, nevertheless, critical to identify. • Similarly, the unmet need measures are expected to have low prevalence and, therefore, a small performance gap. If unmet needs are manifest, they must be taken very seriously because unmet needs in basic activities of daily living jeopardize the person's health and increase risk of institutionalization. Finally, the revised Measure Worksheet addresses the following specific requests for information by the Committee: • Rationale on the length of survey, how long it takes to complete, and burden on beneficiaries/caregivers was added to Feasibility section 3c.1; • Information on low survey response rates and understanding the feasibility of getting to optimal sample size was added to Feasibility section 3c.1; • The value of having both the global measures and the recommendation measures was added to Developer Rationale 1b.1; • Data needed on the extent of exclusions and impact on measurement was added to Denominator Exclusions section s.10; • Data from additional methods of assessing reliability such as Spearman-Brown prophecy formula, inter-class correlations, and factor analysis were added to Reliability Testing section 2a2.; and • Evidence from cognitive testing was provided in 1c.5 of Evidence, Performance Gap, Priority – Importance of Measure and Report.

Staff Notes: Please see notes above regarding Public Comment response to the "not recommended" vote for these measures; also note the identified gap for these measures as identified by the HCBS Committee (measure framework project) and the MAP Dual Eligibles Workgroup. Should also note that the Committee did pass these measures on the Importance vote, thus that does not need to be reconsidered. As noted above, the survey has now received the CAHPS trademark and the developers have converted to top-box scoring to be consistent with CAHPS reporting. The developers have provided information on the number of respondents that were excluded due to not passing the cognitive screening questions which serves as an update to threats to validity. As the developers indicate, the conversion to top-box scoring and inclusion of proxy respondents seems to have improved both the measure(s) performance and reliability estimates. The items/measures that remain of some concern regarding gap include: Personal Safety and Respect, Unmet Needs: Sufficient Staff for Toileting, and Not Hit or Hurt by Staff. As the developers indicate, while performance is high, these can all be considered "never" or "rare" events and should be considered in that light. Additional information on how the measure was cognitively tested has been provided. Updated reliability and validity analysis is provided starting on page 55 of the Measure Evaluation Worksheet.

Functional Status: Related/Competing Measures

Excerpts from the Committee Guidebook

Criterion #5: Related and Competing Measures

- NQF endorses national consensus standards—and this implies parsimony and standardization to the extent possible. Duplicative measures and/or those with similar but not identical specifications increase measurement burden and can create confusion or inaccuracy in interpreting performance results, especially if such measures produce different results for the same provider. Therefore, if a measure has met all the previous NQF evaluation criteria, the standing committee will then evaluate that measure in relation to other competing or related measures. In this evaluation, the two primary considerations will be the evidence driving the differing measure specifications the applicability of the measure (ideally, measures should include as many relevant entities as possible, based on the evidence).
- Competing measures are those measures that are intended to address the same measure focus and the same target population, while related measures are those intended to address the same measure focus or the same target population. Ideally, when evaluating competing measures, the committee will be able to identify the superior measure(s)—in which case, the committee will recommend the superior measure as suitable for endorsement but would not recommend the competing measure(s). Similarly, when evaluating related measures, the committee ideally will be able to make recommendations for harmonization (suggested alterations of related measures to make their specifications more similar).
- The dimensions of harmonization can include numerator, denominator, exclusions, calculation, and data source and collection instructions; however, the extent of harmonization depends on the relationship of the measures, the evidence for the specific measure focus, and differences in data sources. In some cases, there may be valid reasons to endorse competing measures or measures that are not harmonized to the extent possible, and measure developers have the opportunity to justify this course of action for the committee.

KEY POINTS ON RELATED AND COMPETING MEASURES

- NQF prefers endorsement of measures that assess performance for the broadest possible application (e.g., for as many possible individuals, entities, settings, and levels of analysis)
  for which the measure is appropriate, as indicated by the evidence.
- The endorsement of multiple competing measures should be by exception, with adequate justification.
- Harmonization of related measures should be done to the extent possible; differences in specifications should be justified.

Developer	UDSMR	AHCA
Measure type	Outcome	Outcome
General Description	2769 Functional Change: Change in Self-Care Score for SNF: Change in rasch derived values of self-care function from admission to discharge among adult patients treated as short term rehabilitation patients in a skilled nursing facility who were discharged alive. 2775: Functional Change: Change in Mobility Score for SNF: Change in rasch derived values of mobility function from admission to discharge among adult short term rehabilitation skilled nursing facility patients aged 18 years and older who were discharged alive	2612: The measure calculates the average change in mobility score between admission and discharge for all residents admitted to a SNF from a hospital or another post-acute care setting for therapy (i.e., PT or OT) regardless of payor status. 2613: The measure calculates the average change in self-care score between admission and discharge for all residents admitted to a SNF from a hospital or another post-acute care setting for therapy (i.e., PT or OT) regardless of payor status.
Measure level	Facility- SNF	Facility - SNF
Assessment Tool	FIM	CARE
Phase 3	2769 and 2774 identified as competing with 2612 and 2613	2612 and 2613 competing with 2769 and 2774

Assessment Items (data elements) used to calculate scores			
	The measure includes the following 12 FIM® items (Motor Skills = 12 items): <u>Self-Care (8 items):</u> Feeding, Grooming, Dressing Upper Body, Dressing Lower Body, Toileting, Bowel, Expression, Memory, <u>Mobility (4 items):</u> Transfer Bed/Chair/Wheelchair, Transfer Toilet, Locomotion Stairs.	The items included in the CARE Tool self-care subscale include: A1. Eating A3. Oral Hygiene A4. Toilet Hygiene A5. Upper Body Dressing A6. Lower Body Dressing C1. Wash Upper Body C2. Shower / Bathe C6. Putting on / taking off footwear The items included in the CARE Tool <u>Mobility subscale include:</u> B1. Lying to Sitting on Side of Bed B2. Sit to Stand B3. Chair/Bed to Chair Transfer B4. Toilet Transfer B5a & B5b. Walking or Wheelchair Mobility C3. Roll left / right	<u>The 7 self-care items (CARE) are:</u> GG 0130A. Eating GG 0130B. Oral hygiene GG 0130C. Toilet hygiene GG 0130D. Shower/bathe self GG 0130E. Upper body dressing GG 0130F. Lower body dressing GG 0130G. Putting on/taking off footwear <u>The 15 mobility items are:</u> GG 0170A. Roll left and right GG 0170B. Sit to lying GG 0170C. Lying to sitting on side of bed GG 0170D. Sit to stand GG 0170E. Chair/bed-to-chair transfer GG 0170F. Toilet transfer GG 0170G. Car transfer GG 0170I. Walk 10 feet GG 0170J. Walk 50 feet with 2 turns

		C4. Sit to Lying C5. Picking up object C7a. One Step Curb C7b. Walk 50 ft. with Two Turns C7c. Walk 12 Steps. C7d. Walk Four Steps C7e. Walking 10 ft. on Uneven surface C7f. Car Transfer	GG 0170K. Walk 150 feet GG 0170L. Walking 10 feet on uneven surfaces GG 1070M. 1 step GG 0170N. 4 steps GG 0170O. 12 steps GG 0170P. Pick up object

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 #: 0420
Measure Title: Pain Assessment and Follow-Up
Measure Steward: Centers for Medicare & Medicaid Services
Brief Description of Measure: ~~NOTE: Specification information in this section is from the 2016 Physician Quality Reporting System Manual. Testing Information is based on the specification in the 2013 (Registry Data) and specification in the 2014 (Claims Data) Physician Quality Reporting System Manual. Specifications from 2013, 2014 and 2016 are included in the attached "NQF Endorsement Measurement Submission Summary Materials"~~

Note to PFCC Standing Committee: The developer will be provided the opportunity to update their form and clarify the measure specification under consideration during this phase of work. The measure has undergone significant changes since their last endorsement review and a full history is documented. NQF staff have highlighted the sections under consideration.

Commented [DH1]: This form has been updated by QIP.

~~2014-2016 Specification Description:~~
Percentage of visits for patients aged 18 years and older with documentation of a pain assessment using a standardized tool(s) on each visit AND documentation of a follow-up plan when pain is present

~~2013 Specification Description (used in Registry Data Testing):~~
~~Percentage of visits for patients aged 18 years and older with documentation of a pain assessment through discussion with the patient including the use of a standardized tool(s) on each visit AND documentation of a follow-up plan when pain is present~~

Developer Rationale: This measure addresses a gap in care. There are disparities in care across population groups as outlined in the following statements:

The American Pain Foundation (2009) identified medically underserved populations endure a disproportionate pain burden in all health care settings.

A growing body of research reveals even more extensive gaps in pain assessment and treatment among racial and ethnic populations, with minorities receiving less care for pain than non-Hispanic whites (Green, 2003; Green, 2007; Green et al., 2011; Todd et al., 2004; Todd et al., 2007). Differences in pain care occur across all types of pain (e.g., acute, chronic, cancer-related) and medical settings (e.g., emergency departments and primary care) (Green, 2003; Green, 2007; Todd et al., 2007). Even when income, insurance status and access to health care are accounted for, minorities are still less likely than whites to receive necessary pain treatments (Green, 2003; Green, 2007; Paulson et al., 2007). Black race is associated with neighborhood socio-economic status (SES) and race plays a role in pain outcomes beyond SES (Green, 2012).

Research also shows gender differences in the experience and treatment of pain. Most chronic pain conditions are more prevalent among women; however, women's pain complaints tend to be poorly assessed and undertreated

(Green, 2003; Chronic Pain Research Alliance 2011, Weimer 2013). Although women may have higher baseline pain, differences in pain levels may not persist at one month (Peterson, 2012).

"When assessing and treating pain, practitioner sex, race, age, and duration of experience were all significantly associated with pain management decisions. These findings suggest that pain assessment and treatment decisions may be impacted by the health care providers' demographic characteristics, effects which may contribute to pain management disparities."(Bartley et al., 2015).

The aim of this quality measure is to assist eligible providers to identify patients experiencing pain and provide a follow-up plan which addresses the patients' pain in an effort to reduce or eliminate the pain. Ultimately, reducing or eliminating pain will improve a patients' quality of life, minimize the disparities that exist in the assessment and treatment of pain and reduce the cost and utilization of healthcare resources.

Numerator Statement: ~~2013 Specification Numerator Statement (used in Registry Data Testing):~~
~~Percentage of visits for patients aged 18 years and older with documentation of a pain assessment through discussion with the patient including the use of a standardized tool(s) on each visit AND documentation of a follow-up plan when pain is present (Testing completed on Registry Data)~~

~~2014 and 2016 Numerator Statement (used in Claims Data Testing):~~
Percentage of visits for patients aged 18 years and older with documentation of a pain assessment using a standardized tool(s) on each visit AND documentation of a follow-up plan when pain is present.
Denominator Statement: All visits for patients aged 18 years and older
Denominator Exclusions: : Not Eligible – A patient is not eligible if one or more of the following reason(s) is documented:

Severe mental and/or physical incapacity where the person is unable to express himself/herself in a manner understood by others. For example, cases where pain cannot be accurately assessed through use of nationally recognized standardized pain assessment tools

Patient is in an urgent or emergent situation where time is of the essence and to delay treatment would jeopardize the patient's health status

Measure Type: Process
Data Source: Administrative claims, Paper Medical Records
Level of Analysis: Clinician : Group/Practice, Clinician : Individual

IF Endorsement Maintenance – Original Endorsement Date: Jul 31, 2008 **Most Recent Endorsement Date:** Jul 31, 2008

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 *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.

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 indicated they have updated the evidence since the last endorsement review and stated the following rationale supporting the measure: *Utilization of validated pain assessment tools facilitates the monitoring of the patient's health status and the differentiation of treatment approaches in order to improve the patient's pain level.*
- Three clinical practice guidelines were provided to support the measures: Assessment and Management of Chronic Pain (2013), Adult Acute and Subacute Low Back Pain (2012) and Clinical Practice Guidelines Linked to the International Classification of Functioning, Disability, and Health from the Orthopaedic Section of the American Physical Therapy Association (2012).
- This measure is a process measure that has a more global target population of all adults, two out of the three guidelines cited focus on low back pain. One of the low back pain guidelines is more specific to imaging used in diagnostics versus the pain assessment and follow-up plans

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: See above, the developer submitted a new evidence form. This measure was originally recommended for time-limited endorsement in 2008. The steering Committee that reviewed the measure recommended the changes the developer has made since that time (clarity on standardized assessment, documentation of follow-up plan).

Exception to evidence

Based on the information provided, is there rationale to support this measure with an exception to evidence? As a process measure, the evidence requirement is a systematic assessment and grading of the quantity, quality and consistency of the body of evidence that measured process leads to a desired health outcome. The developers provide clinical guideline recommendations for adult pain and low back pain, and specifically on the importance of assessment. We are specifically looking for evidence that the assessment and documentation of a treatment plan for pain leads to improved outcomes. The lack of systematic assessment of evidence may be an oversight versus the lack of evidence.

Guidance from the Evidence Algorithm

For a process measure, is it based on systematic review and grading of the BODY of empirical evidence where specific focus of the evidence matches what is being measured (box 2): No → is empirical evidence submitted but without systematic review and grading of the evidence (box 7): No → Are there, or could there be , performance measures of a related health outcome or evidence-based intermediate clinical outcome or process (box 10): No → is there evidence of systematic assessment of expert opinion that the benefits outweigh potential harms (box 11): Yes → Does the SC agree that it is okay to hold the providers accountable for performance in the absence of empirical evidence?: if yes – Rate as insufficient evidence with exception; if No – rate as insufficient.

Questions for the Committee:

If the developer provided updated evidence for this measure:

- *Questions specific to the measure information provided on evidence*
  - *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?*

- *For possible exception to the evidence criterion:*
  - *Are there, or could there be, performance measures of a related health outcome, OR evidence-based intermediate clinical outcomes, intervention/treatment?*
  - *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*
  - *Does the SC agree that it is acceptable (or beneficial) to hold providers accountable without empirical evidence?*

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

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

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

The developer provides the following summary of performance data from PQRS:
A. Quality Indicator Performance 1/1/2014 through 12/31/2014
1. Total Claims Submitted- 10,555,143
2. Valid Denominator Criteria - 9,515,468/ 90.2% of total
3. Performance Exclusion – 341,159/ 3.5% of valid
4. Measure Performance Rate- 7,627,424 / 9,174,309 83.1%

B. Performance Variation by Eligible Professional 1/1/2014 through 12/31/2014: Describes the variation of measure scores by discrete National Provider Identification (NPI).
• N (# of NPIs) – 59,722
• Mean Measure Score – 81.9%
• Standard Deviation - .35
• Min/Max – 0/100%
• 1st percentile – 0.0%
• 5th percentile – 0.0%
• 10th percentile – 0.0%
• 25th percentile – 90.6%
• 50th percentile – 100.0%

The developer also notes: Reporting for the measure is voluntary and providers who report may not be representative of all eligible professionals. In 2014 only 10.7% of eligible providers reported this measure. Reported performance rates from this group cannot be generalized to the total eligible population

Disparities
Disparities in performance based on race/ethnicity, urban/rural status, gender and age were identified. Analysis of claims from 1/1/2014 through 12/31/2014 reveal statistically significant differences in measure performance between genders and age groups with larger differences observed between urban/rural providers and patient race/ethnic group.

Performance rates by categories:
Rural 87.3%, Urban 81.8% (X2 = 34753.95, N = 9,159,741 p < .0001)
Female 83.7%, Male 82.2 % (X2 = 3424.87, N = 9,174,309 p < .0001)
White 84.2%, Non-white 70.6% (X2 = 85850.38, N = 9,002,090 p < .0001)
Asian 76.2%, Black 68.2%, Hispanic 79.1%, Native 73.6%, White 84.2%, Other 79.6%, Unknown 86.1%(X2 = 95281.16, N = 9,174,309 p < .0001))
Age Under 50 years 80.0%, 50-64 years 80.9%, 65-69 years 85.4%, 70-74 years 84.6%, >=75 81.7% (X2 = 23394.64, N = 9,174,309, p < .0001)

Questions for the Committee:

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

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

Committee pre-evaluation comments

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

1.a. Evidence to Support Measure Focus

Comments:

**Providing clinical guidelines only supports the premise that assessment and a plan of treatment is important, in that it is necessary but not sufficient to improve pain. The developers did not provide evidence that assessing pain and documenting a plan resumed in improved pain scores, or improved quality of life or function. There is no way of knowing if the plan documented is evidence based or effective. The guidelines sipped are tangentially related, rather than directly related. I am not aware of any studies that either support or refute that better assessment results in improved health outcomes.

**The measure developer sites guidelines that recommend screening for pain and there was a comment as to whether the screening and development of a plan improved patients' outcomes for pain management.
An article published in 2007 questioned the Accuracy of the Pain Numeric Rating Scale as a Screening Test in Primary Care:
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2305860/

**The evidence supports the assessment of pain in adults with low back pain. The evidence does suggest that treatment and improvement of pain is a goal worthy of investigation. However, one source notes, "1.Increase the identification of patients who are in the early stages of a serious illness who would benefit from palliative care.
2.Improve the effectiveness and comfort level of primary care clinicians in communicating the necessity and benefits of palliative care with those patients with a serious illness.
3.Improve the assessment of the identified patient's palliative care needs, utilizing the domains of palliative care. 4.Increase the percentage of patients in the early stages of a serious illness who have a care plan identified and/or documented.
5.Improve the ongoing reassessment and adjustment of the patient's plan of care as the condition warrants, utilizing the domains of palliative care.
6.Increase the completion, documentation and ongoing utilization of advance directives for patients with a serious illness."
https://www.icsi.org/guidelines__more/catalog_guidelines_and_more/catalog_guidelines/catalog_palliative_care_guidelines/palliative_care/

A second source concurs that assessment and planning should identify the type and source of chronic pain and the plan should match the finding based on the assessment. They also note the aims as follows:
Aims
1. Improve the function of patients age 18 years and older with chronic pain. (Annotations #2, 14)
2. Improve the assessment and reassessment of patients age 18 years and older with chronic pain diagnosis utilizing the biopsychosocial model. (Annotations #2, 3, 12)
3. Improve the appropriate use of Level I and Level II treatment approaches for patients age 18 years and older with chronic pain. (Annotations #14, 19, 25)
4. Improve the effective use of non-opioid medications in the treatment of patients age 18 years and older with chronic pain. (Annotations #15, 19)
5. Improve the effective use of opioid medications in the treatment of patients age 18 years and older with chronic pain. (Annotations #15, 19)
https://www.icsi.org/_asset/bw798b/ChronicPain.pdf

However, the Faces Pain Scale (FPS) was designed for use in children and does not include instructions on assessing intensity, quality of pain, etc. http://www.iasp-pain.org/Education/Content.aspx?ItemNumber=1519
In addition, the rationale specifically states that the goal is assessment of all types of chronic pain, yet the evidence several discussions limited to the assessment and treatment of chronic low back pain.

Given that the acceptable measures include the faces scale, which is a 1-10 pain scale

1b. Performance Gap

Comments:
**There does appear to be an ongoing performance gap between urban and rural providers, and patient ethnic group. Black patients remain under assessed and treated compared to white patients, with other non-white patients displaying smaller gaps compared to whites.
**The developer supplied data showing variation in results although overall good performance. Since it is a voluntary measure it is possible higher performing groups chose to submit. Only about 10% of eligible providers submitted.
**While the resources, do, support the use of a performance measure related to chronic pain, the measure, as it is proposed, does not assess the outcome of the treatment. From a patient and family centric view of this measure, pain assessment and planning has little value without producing some benefit.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications
Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Administrative Claims data

Specifications:

- Satisfactory reporting criteria are met by valid submission of one of six G codes on claims that meet denominator criteria
- The measure is reported via G-codes (numerator and exclusions) and CPT codes (denominator)
- The numerator reporting options are performance met, pain assessment not documented patient not eligible, and pain assessment not documented reason not given (all reported via G-codes)
- This is a process measure and is not risk adjusted

Questions for the Committee :

- *Specific questions on the specifications, codes, definitions, etc.*
- *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?*

2a2. Reliability Testing Attachment
Maintenance measures – less emphasis if no new testing data provided

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.

Describe any updates to testing

Because of the updates to the specifications over time, and the ability to gather data through PQRS, the developer updated testing to reflect the current measure specifications (use of G-Codes)

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☒ **Data element** ☒ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing

Critical Data Element Testing: Quality Insights of Pennsylvania (Quality Insights) oversees the abstraction of 405 randomly generated Medicare Part B claims records for all 74 unique NPIs/eligible professionals who reported one of the G-codes for the measure during the 1/1/2014 – 12/31/2014 time period. Quality Insights requests the medical record documentation from the NPI/eligible professional for the randomly selected encounter date. The documentation is abstracted and a G-code is assigned by two registered nurse (RN) abstractors, one from Quality Insights and one from an independent reviewer contracted with Quality Insights, according to the measure specifications.

Agreement rates between independent reviewers were calculated (inter-rater reliability) as well as the rate of agreement between the numerator code submitted with the claim and an independent reviewer (critical data element validity. See 2b2. Validity testing). Crude agreement, prevalence adjusted kappa (PAK), Cohen´s kappa values and corresponding confidence intervals were calculated.

Performance Score: reliability is estimated with a beta-binomial model. The beta-binomial model is appropriate for measuring the reliability of pass/fail measures such as those proposed.

Results of reliability testing

Inter-Rater Reliability:
Numerator crude agreement 95.0%
Prevalence adjusted kappa .90 (CI .86 – .94)
Kappa .87 (CI -.81 – .93)

Performance measure score (1/1/2013 – 12/31/2013):

Data source	N	Between-provider variance	Reliability mean	Reliability median	Reliability Std dev	Reliability min/max
Claims	29,398	.105	.994	1.0	.020	.457 - 1.0
Registry	5,639	.214	.996	1.0	.012	.817 – 1.0

Guidance from the Reliability Algorithm

Are specifications precise and complete (box 1): Yes →Was empirical reliability testing conducted (box 2): Yes → Was reliability testing conducted with computed performance measure scores for measured entity (box 4): Yes → Was method described appropriate (box 5): Yes → Based on reliability statistics and scope – what is level of certainty or confidence that the performance measure scores are reliable (box 6): High

Questions for the Committee:

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

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

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☐ **Yes** ☒ **Somewhat** ☐ **No**

Specification not completely consistent with evidence

- We would like the committee to discuss; while evidence form was submitted and contained clinical recommendations, there may be additional evidence to support this measure that was not submitted. Based on what was on the evidence

form, staff would rate this as "somewhat" met; however, it seems appropriate that a pain assessment would be conducted and follow-up plan documented and this was the recommendation of past committees.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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, summarize the validity testing from the prior review:
Note: the prior measure testing forms were not found thus information is updated in this form.

Describe any updates to validity testing
The developer indicated on their measure checklist that they did not update validity testing, but noted in their testing form that patient level data elements were assessed. This is described below.

SUMMARY OF TESTING
Validity testing level ☐ **Measure score** ☒ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☐ **Face validity only**
- ☒ **Empirical validity testing of the measure score**

Validity testing method:

Quality Insights of Pennsylvania (Quality Insights) oversees the abstraction of 405 randomly generated Medicare Part B claims records for all 74 unique NPIs/eligible professionals who reported one of the G-codes for the measure during the 1/1/2014 – 12/31/2014 time period. Quality Insights requests the medical record documentation from the NPI/eligible professional for the randomly selected encounter date. The documentation is abstracted and a G-code is assigned by two registered nurse (RN) abstractors, one from Quality Insights and one from an independent reviewer contracted with Quality Insights, according to the measure specifications.

Agreement rates between independent reviewers were calculated (inter-rater reliability) as well as the rate of agreement between the numerator code submitted with the claim and an independent reviewer (critical data element validity). Crude agreement, prevalence adjusted kappa (PAK), Cohen´s kappa values and corresponding confidence intervals were calculated.

Validity testing results:
Critical data element testing:
Overall Reliability of Claims vs. Independent Review:
Numerator crude agreement 85.9%
Prevalence adjusted kappa .72 (.66 - .79)
Kappa .55 (86% CI .45 - .65)

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*
- *Other specific question of the validity testing?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- A patient is not eligible if one or more of the following reason(s) is documented:

- Severe mental and/or physical incapacity where the person is unable to express himself/herself in a manner understood by others. For example, cases where pain cannot be accurately assessed through use of nationally recognized standardized pain assessment tools
- Patient is in an urgent or emergent situation where time is of the essence and to delay treatment would jeopardize the patient's health status

QIP analyzed 10,555,143 claims submitted for this measure. Of those 9,515,468 (90.2%) met the denominator criteria for patient age and relevant CPT codes as defined in the measure specifications. It was from that pool the sample for reliability testing was drawn. Two independent clinical reviewers abstracted 405 cases from 74 providers to assess validity of exclusion criteria in claims reporting for encounters from 1/1/2014 to 12/31/2014.

3.6 % of the total number of valid claims were reported as exclusions.

Testing of exclusion criteria agreement demonstrated high reliability in measure reporting. Reliability between two independent clinical reviewers was almost perfect with a PAK = .98, (95% CI=.96 - 1.0) and crude agreement= 99.0%; similarly the "gold standard" clinical reviewer vs. claims agreement was almost perfect with a PAK = .98 (99% CI .97 - 1.00), crude agreement=99.2%.

Questions for the Committee:

- *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)?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:
Reported provider performance variation (2014):
N – 59,722
Mean – 81.9%
Min – 0.0%, Max – 100.0% Std Deviation .35
50th percentile – 100.0%
25th percentile – 90.6%
10th percentile – 0.0%
1st percentile – 0.0%

- *The overall performance rate reported via claims for the period 1/1/2014 to 12/31/2014 was 83.1%. The average provider performance rate was 81.9%.*
- *Average reported performance rates are above 80% however the need for improvement can be seen for the lowest 10% reporting (10th percentile 0.0%). It should also be noted that the measure is reported voluntarily and those eligible professionals who chose to report may not be representative of the total population of eligible providers.*

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

N/A

2b7. Missing Data

The number of eligible providers reporting the measure is about 10.7% (3.6% in 2010, 4.5% in 2011, 1.8% in 2012, and 7.4% in 2013).

Because reporting is voluntary the reporting population cannot be said to be representative of the total eligible population. Generalizations to the overall eligible population should not be made.

Guidance from the Reliability Algorithm

Measure specifications consistent with evidence (Box 1): Yes →All relevant potential threats to validity assessed (Box 2): Yes → empirical validity testing using measure as specified (Box 3): Yes → Validity tested at computed performance measure score (Box 6): Yes → Method described appropriate (Box 7): Yes → Based on results and scope of testing and analysis of potential threats, level of certainty/confidence that measure scores are a valid indicator of quality (Box 8): Moderate (some questions about direct evidence support for measure as specified; face validity information not particularly useful, yet exclusion testing and overall validity of measure seemed sound)

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

Committee pre-evaluation comments

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

2a.1 & 2b.1 Specifications

Comments:

**All specifications clear. No risk adjustment since this is a process not an outcome measure. The specifications are consistent with the evidence, in that the guidelines recommend validated tools for assessment. Documentation of plan is not as well defined in the guidelines, with the exception of whether imaging is indicated in radicular pain.

**Specifications are clear and the reliability and validity were assessed.

**Numerous measures are offered as meeting the requirements for a valid and reliable measure. I am somewhat concerned that no evidence is offered about the appropriateness of the measures related to various diagnoses. Many are specific to low back, yet the measure under review does not limit its usability to that population. This raises validity concerns.

2a.2 Reliability Testing

Comments:

**Measure score and data level testing were both performed. Reliability was tested between independent reviewers, and between reviewer and submitter. A sufficient n of encounters where included, from all unique participating providers. Appropriate testing methods were used for a pass/fail measure. Sufficient reliability testing demonstrated.

**Chart reviews were done on a sample of the results submitted and were found to be consistent.

**Unclear given the disparity noted in 2b.1.

2b.2 Validity Testing

Comments:

**Again, assessment and plan for treatment of pain are necessary but not sufficient to improve patient's lives. Agree with past committees that its reasonable to perform these first steps, without which, quality care cannot be provided. Data elements tested against the gold standard only. Adequate scope and entities included for reliability testing, with correct method used.

**The results support whether an assessment and plan were done which are consistent with accepted guidelines. What is less clear is whether this results in a patient centered outcome of either less pain or increased function.

**If I understand correctly, this measure does not evaluate the quality and appropriateness of the tool or the plan. It only assesses whether or not a tool, of any variety, was used, and a plan, also of any variety was created.

2b.3.-2b7. Testing (Related to Potential Threats to Validity)

Comments:

**Exclusion seem reasonable and are sufficiently rare (about 3%.)Exclusions were also reliably identified. Exclusion groups narrow, meaning the vast majority of patients would be included. No patient groups unfairly excluded. The analysis supports that the bottom 10% percentile have lots of room for improvement.

**Exclusions noted along with frequency. Stratification was done to show gaps related to a number of factors.

**In addition, the high floor value and low ceiling suggest that process measure will not be useful in identifying the disparities in chronic pain care it seeks to tease out.

B. Performance Variation by Eligible Professional 1/1/2014 through 12/31/2014: Describes the variation of measure scores by discrete National Provider Identification (NPI).
- N (# of NPIs) – 59,722
- Mean Measure Score – 81.9%
- Standard Deviation - .35
- Min/Max – 0/100%
- 1st percentile – 0.0%
- 5th percentile – 0.0%
- 10th percentile – 0.0%
- 25th percentile – 90.6%
- 50th percentile – 100.0%

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

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

The measure is collected primarily via administrative data (claims), but has an option for medical record abstraction.

Questions for the Committee:
- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*

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

Committee pre-evaluation comments
Criteria 3: Feasibility

3. Feasibility

Comments:

**Pain scales of 1-10 almost always generated and captured, but more meaningful scales (VAS, Wong Baker) remain clinically underutilized. A plan of care is rarely documented in any systematic way, and usually involves concerted effort to develop a template that "forces" documentation of a clinically meaningful plan. However, once that process is established, the required data elements can be easily documented in an EHR, and extracted from there with only moderate burden. The upfront investment can be burdensome in other words.
**The groups (10% of eligible) that submitted showed the measure to be feasible. There are concerns that overall feasibility across more clinicians, especially in primary care, could be challenging given the number of things primary care is already expected to do in a given visit. This barrier could be one of the reasons more groups didn't submit this through PQRS.
**No issues noted.

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

4. Usability and 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.

Current uses of the measure
PQRS

Publicly reported? ☒ **Yes** ☐ **No**

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

Accountability program details
The measure is currently in use in the PQRS program; in 2014, there were 573,233 (10.7%) Eligible Professionals who could report NQF# 0420. In 2013, NQF #0420 was the 6th most reported measure within PQRS with 664,929 (7.4%) eligible professionals participating in reporting this measure.

Improvement results
Provider and Patients Statistics for program year 2014 (from "2014 Physician Quality Reporting System Program Monitoring and Evaluation Report"):
Average Performance Rates by Year (PQRS – all reporting methods):
2009 – 97.4%
2010 – 97.3%
2011 – 94.8%
2012 – 86.9%
2013 – 85.7%
2014 - 88.5%

Unexpected findings (positive or negative) during implementation
The developer indicated no unexpected findings

Potential harms
For the overall measure, none noted. For low back pain, it was noted that a standardized, back-specific pain assessment could potentially prevent unwarranted imaging studies.

Feedback :
None; measure was not on the most recent MUC list (2015-6 MAP proceedings)

***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**

Committee pre-evaluation comments
Criteria 4: Usability and Use

4. Usability and Use
Comments:
**Voluntary reporting through PQRS. Continued emphasis on at least assessing and planning to treat and follow progress is undeniably useful, but hope that the process measure eventually becomes an outcome measure, where provider must ensure the plan is evidence based and demonstrating improvement in wellbeing and function.

Documenting these elements might result in more efficient care, if the plan is adequate and results in less resource use, or less cost to the system, with better outcomes.

Criterion 5: Related and Competing Measures

Related or competing measures

0383 : Oncology: Plan of Care for Pain – Medical Oncology and Radiation Oncology (paired with 0384)
0676 : Percent of Residents Who Self-Report Moderate to Severe Pain (Short-Stay)
0677 : Percent of Residents Who Self-Report Moderate to Severe Pain (Long-Stay)
1628 : Patients with Advanced Cancer Screened for Pain at Outpatient Visits
1634 : Hospice and Palliative Care -- Pain Screening
1637 : Hospice and Palliative Care -- Pain Assessment

Harmonization

The developer reports that all measures listed above (and a similar list of measures related, but not endorsed) have not been harmonized, and provided rationale and analysis of differences in measures. Staff review indicates relation to list of measures, and agrees that not competing, mainly due to variations in target population and numerator requirements.

Pre-meeting public and member comments

- We support the pain assessment measure but it is not obvious if any specification for what a "standard" measure of this is—e.g. is a pain scale (what is your pain on a scale from 1-10) sufficient? Also, it is interesting to think about how this gets operationalized in the context of other efforts to try to mitigate overprescribing of opioids. We agree with the need for assessment of pain and a follow-up plan where pain is present, but it is not clear what is acceptable as a follow-up plan—just a prescription and a plan to reevaluate? Referral to pain specialist, PT, etc.?

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0420

Measure Title: Pain Assessment and Follow-Up

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: 3/30/2016

All the information in this form is updated from last endorsement of NQF 0420 in September 2011. This NQF evidence form was not in existence in 2010/2011. Evidence continues to support measure focus.

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

☐ Health 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

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☒ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

N/A

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

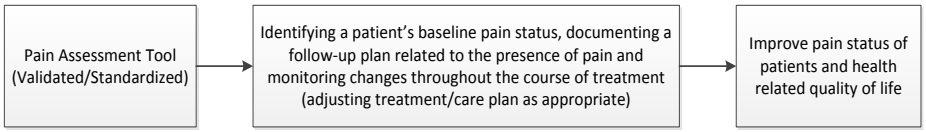
N/A

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

Utilization of validated pain assessment tools facilitates the monitoring of the patient's health status and the differentiation of treatment approaches in order to improve the patient's pain level.



1. Assess for the presence of pain using a standardized tool in all patients aged 18 years and older
2. Identification of pain (positive screen) results in the documentation of a follow-up plan related to the presence of pain and the management of it to reduce pain intensity.
3. Follow-up recommendation and intervention strategies for treating pain can lead to decreased level of pain, thus improving the health and well-being of the patient and can help to reduce the use of healthcare resources and/or lost productivity.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1. Hooten, W.M., Timming, R., Belgrade, M., Gaul, .J, Goertz, M., Haake, B., … Walker, N.(2013). Assessment and management of chronic pain. *Institute for Clinical Systems Improvement* (6th ed.). Retrieved from https://www.icsi.org/_asset/bw798b/ChronicPain.pdf

2. Goertz, M., Thorson, D., Bonsell, J., Bonte, B., Campbell,.R, Haake B., …, Timming, R. (2012). Adult Acute and Subacute Low Back Pain. *Institute for Clinical Systems Improvement* (15th ed). Retrieved from https://www.icsi.org/_asset/bjvqrj/LBP.pdf

3. Delitto, A., George, S.Z., Van Dillen, L.R., Whitman, J.M., Sowa, G., Shekelle, P., & Denninger, T.R. (2012). Low back pain. Clinical Practice Guidelines Linked to the International Classification of Functioning, Disability, and Health from the Orthopaedic Section of the American Physical Therapy Association. *Journal of Orthopedic Sports Physical Therapy*, 42(4), A1-A57.

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1. ICSI Guideline: Assessment and Management of Chronic Pain (Hooten et al. (2013)) – The assessment and management algorithms are found on pages 1 and 2 of guideline.

A. Assessment Algorithm Annotations (p.12)

Critical First Step: Assessment

Recommendations:

- A clinician should complete an adequate pain assessment on all patients that includes documentation of pain location, intensity, quality, onset/duration/variations/rhythms, manner of expressing pain, pain relief, exacerbation triggers, effects of pain and response to previous treatments.
    - Musculoskeletal assessment - Rasmussen, 2004 [*Low Quality Evidence*]
    - Multidimensional assessment tools – Cleeland, 1994 *[Low Quality Evidence*], Smith, 1997 [Low Quality Evidence], Galer, 1997 *[Low Quality Evidence],* Savedra, 1989 *[Low Quality Evidence*], VanCleve, 1993 [*Low Quality Evidence),* Penny, 1999 *[Low Quality Evidence]*

      General approach to use of pain assessment tools in chronic pain:
        - On initial visit, use a multidimensional tool such as the Brief Pain Inventory to obtain a comprehensive picture of the pain experience. The patient should complete this assessment tool before the physician visit.
        - With follow-up visits, continue to use a multidimensional pain assessment tool filled out by the patient before seeing the physician.
        - Use specific tools such as the Neuropathic Pain Scale (NPS) when appropriate.
        - Avoid the use of single-dimensional pain assessment tools in chronic pain except to rate the intensity of specific pain episodes.

      (American Pain Society, 2005 [*Low Quality Evidence*]; Herr, 2004 [*Guideline*]; Kaiser Permanente Medical Care Program, 2004 [*Guideline*]; McCaffery, 1999 [Guideline]; Daut, 1983 *[Low Quality Evidence])*

2. ICSI Guideline: Adult Acute and Subacute Low Back Pain (Goertz et al., 2012) **–** Algorithms for Core Treatment of Non-specific Low Back Pain, Red Flags and Radicular Pain are located on pages 1-3 of guideline

A. Recommendations Table for the assessment and treatment of acute and subacute low back pain (p.7)

Topic	Quality of Evidence	Recommendation	Strength of Recommendation	Annotation Number	Relevant References
Education	Moderate	Clinicians should educate patients as an adjunct to other treatment. No standardized form of education is suggested.	Strong	11, 16, 17, 18	*Engers, 2008; Heymans, 2004*

B. Core Treatment of Non-specific Low Back Pain Algorithm Annotations: B. Initial Evaluation and Data Set: Recommendation (p.12)

- Clinicians should not recommend imaging (including computed tomography [CT], magnetic resonance imaging [MRI] and X-ray) for patients with non-specific low back pain *(Strong Recommendation, Moderate Quality Evidence) (Chou 2011; French 2010; Chou 2009b).*

Note: The supportive documentation for this recommendation advises the use of pain assessment tools instead of imaging to influence medical decision-making in the first 6 weeks of onset of non-specific low back pain (p.12).

C. Reevaluation (p. 16)
   - Reevaluation of low back pain should include the following:
     - Pain reassessed with a repeat Visual Analog Scale and Oswestry Disability Questionnaire

3. Low Back Pain: Clinical Practice Guidelines (Delitto et al. (2012))

A. CLINICAL COURSE (p. A2): The clinical course of low back pain can be described as acute, subacute, recurrent, or chronic. Given the high prevalence of recurrent and chronic low back pain and the associated costs, clinicians should place high priority on interventions that prevent (1) recurrences and (2) the transition to chronic low back pain. (Recommendation based on theoretical/foundational evidence.)
B. EXAMINATION – OUTCOME MEASURES (p. A2): Clinicians should use validated self-report questionnaires, such as the Oswestry Disability Index and the Roland-Morris Disability Questionnaire. These tools are useful for identifying a patient's baseline status relative to pain, function, and disability and for monitoring a change in a patient's status throughout the course of treatment. (Recommendation based on strong evidence.)
C. EXAMINATION – ACTIVITY LIMITATION AND PARTICIPATION RESTRICTION MEASURES (p. A2): Clinicians should routinely assess activity limitation and participation restriction through validated performance-based measures. Changes in the patient's level of activity limitation and participation restriction should be monitored with these same measures over the course of treatment. (Recommendation based on expert opinion.)

1a.4.3. Grade assigned to the quoted recommendation <u>with definition</u> of the grade:

1. ICSI Guideline: Assessment and Management of Chronic Pain (Hooten et al., 2013). See section 1a.4.2 for grade and 1a.4.4 for definition.
2. ICSI Guideline: Adult Acute and Subacute Low Back Pain (Goertz et al., 2012). Strong Recommendation; Moderate Quality Evidence. Definition: see section 1a.4.4
3. Low Back Pain: Clinical Practice Guidelines (Delitto et al. (2012))
   A. CLINICAL COURSE: Recommendation E - (Theoretical/foundational evidence): A preponderance of evidence from animal or cadaver studies, from conceptual models/principles, or from basic science/ bench research supports this conclusion
   B. EXAMINATION – OUTCOME MEASURES**:** Recommendation A - (Strong evidence): A preponderance of level I and/or level II studies support the recommendation
   C. EXAMINATION – ACTIVITY LIMITATION AND PARTICIPATION RESTRICTION MEASURES: Recommendation F - (Expert opinion): Best practice based on the clinical experience of the guideline development team

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1. & 2. ICSI Guidelines Assessment and Management of Chronic Pain Adult Acute and Subacute Low Back Pain use the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system as a method of assessing the quality of evidence and writing recommendations**.** See below for definitions

Grading of Recommendations Assessment, Development and Evaluation (GRADE)

Category	Quality Definitions	Strong Recommendation	Weak Recommendation
High Quality Evidence	Further research is very unlikely to change the work group's confidence in the estimate of effect.	The work group is confident that the desirable effects of adhering to this recommendation outweigh the undesirable effects. This is a strong recommendation for or against. This applies to most patients.	The work group recognizes that the evidence, though of high quality, shows a balance between estimates of harms and benefits. The best action will depend on local circumstances, patient values or preferences.
Moderate Quality Evidence	Further research is likely to have an important impact on the work group's confidence in the estimate of effect and may change the estimate.	The work group is confident that the benefits outweigh the risks, but recognizes that the evidence has limitations. Further evidence may impact this recommendation. This is a recommendation that likely applies to most patients.	The work group recognizes that there is a balance between harms and benefit, based on moderate quality evidence, or that there is uncertainty about the estimates of the harms and benefits of the proposed intervention that may be affected by new evidence. Alternative approaches will likely be better for some patients under some circumstances.
Low Quality Evidence	Further research is very likely to have an important impact on the work group's confidence in the estimate of effect and is likely to change. The estimate or any estimate of effect is very uncertain.	The work group feels that the evidence consistently indicates the benefit of this action outweighs the harms. This recommendation might change when higher quality evidence becomes available.	The work group recognizes that there is significant uncertainty about the best estimates of benefits and harms.

3. Low Back Pain: Clinical Practice Guidelines (Delitto et al. (2012)) - uses criteria described by the Centre for Evidence-Based Medicine, Oxford for grading the recommendations. See below for definitions.

Oxford Centre for Evidence-Based Medicine

Recommendation Grades

Recommendation A. (Strong evidence): A preponderance of level I and/or level II studies support the recommendation. This must include at least one level I study

Recommendation B. (Moderate evidence): A single high-quality randomized controlled trial or a preponderance of level II studies support the recommendation

Recommendation C. (Weak evidence): A single level II study or a preponderance of level III and IV studies, including statements of consensus by content experts, support the recommendation

Recommendation D. (Conflicting evidence): Higher-quality studies conducted on this topic disagree with respect to their conclusions. The recommendation is based on these conflicting studies

Recommendation E. (Theoretical/foundational evidence): A preponderance of evidence from animal or cadaver studies, from conceptual models/principles, or from basic science/ bench research supports this conclusion

Recommendation F. (Expert Opinion): Best practice based on the clinical experience of the guideline development team

Levels of Evidence:

I. Evidence obtained from high-quality diagnostic studies, prospective studies, or randomized controlled trials
II. Evidence obtained from lesser-quality diagnostic studies, prospective studies, or randomized controlled trials (eg, weaker diagnostic criteria and reference standards, improper randomization, no blinding, <80% follow-up)
III. Case-controlled studies or retrospective studies
IV. Case series
V. Expert Opinion

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

For "Low Back Pain: Clinical Practice Guidelines" (Delitto et al. (2012)) - uses criteria described by the Centre for Evidence-Based Medicine, Oxford for grading the recommendations:

Low Back Pain: Clinical Practice Guidelines Centre for Evidence-Based Medicine, Oxford, United Kingdom URL: (http://www.cebm.net/index.aspx?o=1025)

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☒ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

N/A

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

N/A

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

N/A

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

N/A

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

N/A

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

N/A

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1. **ICSI Guideline: Assessment and Management of Chronic Pain** (Hooten et al., 2013) – **"**This guideline discusses the assessment and management of chronic pain. It is intended for primary care clinicians to help with diagnosis and management of primarily four types of biological markers for pain: neuropathic, muscle, inflammatory and mechanical/compressive. Although opioid use is discussed in this guideline, it is not a comprehensive discussion of the usage of opioids in chronic pain."

2. **ICSI Guideline: Adult Acute and Subacute Low Back Pain** (Goertz et al., 2012) – "Adult patients age 18 and over in primary care who have symptoms of low back pain or radiculopathy. The focus is on the acute (pain for up to 7 weeks) and subacute (pain for between 7 and 12 weeks) phases of low back pain. It includes the ongoing management, including indications for spine specialist referral within the first 12 weeks of onset."

3. **Low Back Pain: Clinical Practice Guidelines** (Delitto et al., 2012) "The purpose of these low back pain clinical practice guidelines, in particular, is to describe the peer-reviewed literature and make recommendations related to (1) treatment matched to low back pain subgroup responder categories, (2) treatments that have evidence to prevent recurrence of low back pain, and (3) treatments that have evidence to influence the progression from acute to chronic low back pain and disability."

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1. **ICSI Guideline: Assessment and Management of Chronic Pain** (Hooten et al., 2013). In the guideline, individual study evidence quality was also graded. These evidence grades, when present, are identified in section 1a.4.2

2. **ICSI Guideline: Adult Acute and Subacute Low Back Pain** (Goertz et al., 2012). In the guideline, individual study evidence quality was also graded. These evidence grades, when present, are identified in section 1a.4.2
   Definitions of GRADE: Same as above

3. **Low Back Pain: Clinical Practice Guidelines** (Delitto et al., 2012). Guideline uses criteria by the Centre for Evidence-Based Medicine, Oxford for grading the recommendations. In the guideline, individual study evidence quality was also graded. These evidence grades are identified in section 1a.4.2. Refer to section 1a.4.3 for definitions.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

Refer to section 1a.4.2 and 1a.4.3 for grades and definitions.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range:

1. ICSI Guideline: Assessment and Management of Chronic Pain (Hooten et al., 2013) - August 2011-August 2013
2. ICSI Guideline: Adult Acute and Subacute Low Back Pain (Goertz et al., 2012) - May 2011- June 2012.
3. Low Back Pain: Clinical Practice Guidelines (Delitto et al., 2012) - 1966-2010

Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

This information is not provided within the ICSI guideline: Assessment and Management of Chronic Pain, ICSI guideline: Adult Acute and Subacute Low Back Pain or in Low Back Pain: Clinical Practice Guidelines.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

1. This information is not provided within the ICSI Guideline: Assessment and Management of Chronic Pain. The literature search was divided into two stages to identify systematic reviews and randomized controlled trials, meta-analysis and other literature.
2. ICSI Guideline: Adult Acute and Subacute Low Back Pain: The literature search was limited to systematic reviews, meta-analysis and randomized control trials. No further information is provided in guideline.
3. Low Back Pain: Clinical Practice Guidelines: The strength of the body of evidence varies from theoretical/foundational evidence to expert opinion to strong evidence. Definitions for the level of evidence include the following:
   I. Evidence obtained from high-quality diagnostic studies, prospective studies, or randomized controlled trials
   II. Evidence obtained from lesser-quality diagnostic studies, prospective studies, or randomized controlled trials (eg, weaker diagnostic criteria and reference standards, improper randomization, no blinding, <80% follow-up)
   III. Case-controlled studies or retrospective studies
   IV. Case series
   V. Expert Opinion

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1. ICSI Guideline: Assessment and Management of Chronic Pain (Hooten et al. (2013)- Not addressed
2. ICSI Guideline: Adult Acute and Subacute Low Back Pain (Goertz) - Not addressed
3. Low Back Pain: Clinical Practice Guidelines (Delitto et al. (2012)) – Not addressed

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

1. ICSI Guideline: Assessment and Management of Chronic Pain (Hooten et al.,2013)- No harms reported
2. ICSI Guideline: Adult Acute and Subacute Low Back Pain - (Goertz et al., (2012)
   - Harm:
     - No Imaging First Six Weeks with Radicular Pain; Use Core Treatment Plan - Recommendation: Clinicians should not recommend imaging (including CT, MRI or X-ray) for patients in the first six weeks of radicular pain *[Strong Recommendation, Moderate Quality Evidence]*.
       - "Most patients with radiculopathy supported by exam findings consistent with history will recover within several weeks of onset. The majority of disc herniations regress or reabsorb by eight weeks from onset. In the absence of red flags or progressive neurologic deficit there is no evidence that the delaying surgery worsens outcomes. The use of the core treatment plan is recommended. Refer to Annotation #11, Core Treatment Plan. With this in mind, in the face of radiculopathy there is no benefit and there is possible harm in obtaining an MRI prior to six weeks. The exception to this is a progressing neurologic deficit or persistent disabling pain. If the patient has demonstrable leg weakness that is disabling or is worsening, further evaluation with imaging and consultation with a spine specialist would also be indicated" (p.29)
3. Low Back Pain: Clinical Practice Guidelines (Delitto et al., 2012) – No harms reported

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

N/A

1a.8 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.

N/A

1a.8.1 What process was used to identify the evidence?

N/A

1a.8.2. Provide the citation and summary for each piece of evidence.

N/A

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 subcriterion 1b).

Brief Measure Information

NQF #: 0420
De.2. **Measure Title:** Pain Assessment and Follow-Up
Co.1.1. **Measure Steward:** Centers for Medicare & Medicaid Services
De.3. **Brief Description of Measure:** ~~NOTE: Specification information in this section is from the 2016 Physician Quality Reporting System Manual. Testing Information is based on the specification in the 2013 (Registry Data) and specification in the 2014 (Claims Data) Physician Quality Reporting System Manual. Specifications from 2013, 2014 and 2016 are included in the attached "NQF Endorsement Measurement Submission Summary Materials"~~

~~2014-2016 Specification Description:~~
Percentage of visits for patients aged 18 years and older with documentation of a pain assessment using a standardized tool(s) on each visit AND documentation of a follow-up plan when pain is present

~~2013 Specification Description (used in Registry Data Testing):~~
~~Percentage of visits for patients aged 18 years and older with documentation of a pain assessment through discussion with the patient including the use of a standardized tool(s) on each visit AND documentation of a follow-up plan when pain is present~~
1b.1. **Developer Rationale:** This measure addresses a gap in care. There are disparities in care across population groups as outlined in the following statements:

The American Pain Foundation (2009) identified medically underserved populations endure a disproportionate pain burden in all health care settings.

A growing body of research reveals even more extensive gaps in pain assessment and treatment among racial and ethnic populations, with minorities receiving less care for pain than non-Hispanic whites (Green, 2003; Green, 2007; Green et al., 2011; Todd et al., 2004; Todd et al., 2007). Differences in pain care occur across all types of pain (e.g., acute, chronic, cancer-related) and medical settings (e.g., emergency departments and primary care) (Green, 2003; Green, 2007; Todd et al., 2007). Even when income, insurance status and access to health care are accounted for, minorities are still less likely than whites to receive necessary pain treatments (Green, 2003; Green, 2007; Paulson et al., 2007). Black race is associated with neighborhood socio-economic status (SES) and race plays a role in pain outcomes beyond SES (Green, 2012).

Research also shows gender differences in the experience and treatment of pain. Most chronic pain conditions are more prevalent among women; however, women's pain complaints tend to be poorly assessed and undertreated (Green, 2003; Chronic Pain Research Alliance 2011, Weimer 2013). Although women may have higher baseline pain, differences in pain levels may not persist at one month (Peterson, 2012).

"When assessing and treating pain, practitioner sex, race, age, and duration of experience were all significantly associated with pain management decisions. These findings suggest that pain assessment and treatment decisions may be impacted by the health care providers' demographic characteristics, effects which may contribute to pain management disparities."(Bartley et al., 2015).

The aim of this quality measure is to assist eligible providers to identify patients experiencing pain and provide a follow-up plan which addresses the patients' pain in an effort to reduce or eliminate the pain. Ultimately, reducing or eliminating pain will improve a patients' quality of life, minimize the disparities that exist in the assessment and treatment of pain and reduce the cost and utilization of healthcare resources.

S.4. **Numerator Statement:** ~~2013 Specification Numerator Statement (used in Registry Data Testing):~~

~~Percentage of visits for patients aged 18 years and older with documentation of a pain assessment through discussion with the patient including the use of a standardized tool(s) on each visit AND documentation of a follow-up plan when pain is present (Testing completed on Registry Data)~~

~~2014 and 2016 Numerator Statement (used in Claims Data Testing):~~
Percentage of visits for patients aged 18 years and older with documentation of a pain assessment using a standardized tool(s) on each visit AND documentation of a follow-up plan when pain is present.
S.7. **Denominator Statement:** All visits for patients aged 18 years and older
S.10. **Denominator Exclusions:** Not Eligible – A patient is not eligible if one or more of the following reason(s) is documented:

Severe mental and/or physical incapacity where the person is unable to express himself/herself in a manner understood by others. For example, cases where pain cannot be accurately assessed through use of nationally recognized standardized pain assessment tools

Patient is in an urgent or emergent situation where time is of the essence and to delay treatment would jeopardize the patient's health status

De.1. **Measure Type:** Process
S.23. **Data Source:** Administrative claims, Paper Medical Records
S.26. **Level of Analysis:** Clinician : Group/Practice, Clinician : Individual

IF Endorsement Maintenance – Original Endorsement Date: Jul 31, 2008 **Most Recent Endorsement Date:** Jul 31, 2008

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?** n/a

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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
NQF_0420_MeasSubm_Evidence_033016.docx

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., the benefits or improvements in quality envisioned by use of this measure)*
This measure addresses a gap in care. There are disparities in care across population groups as outlined in the following statements:

The American Pain Foundation (2009) identified medically underserved populations endure a disproportionate pain burden in all health care settings.

A growing body of research reveals even more extensive gaps in pain assessment and treatment among racial and ethnic populations, with minorities receiving less care for pain than non-Hispanic whites (Green, 2003; Green, 2007; Green et al., 2011; Todd et al., 2004; Todd et al., 2007). Differences in pain care occur across all types of pain (e.g., acute, chronic, cancer-related) and medical settings (e.g., emergency departments and primary care) (Green, 2003; Green, 2007; Todd et al., 2007). Even when income, insurance status and access to health care are accounted for, minorities are still less likely than whites to receive

necessary pain treatments (Green, 2003; Green, 2007; Paulson et al., 2007). Black race is associated with neighborhood socio-economic status (SES) and race plays a role in pain outcomes beyond SES (Green, 2012).

Research also shows gender differences in the experience and treatment of pain. Most chronic pain conditions are more prevalent among women; however, women's pain complaints tend to be poorly assessed and undertreated (Green, 2003; Chronic Pain Research Alliance 2011, Weimer 2013). Although women may have higher baseline pain, differences in pain levels may not persist at one month (Peterson, 2012).

"When assessing and treating pain, practitioner sex, race, age, and duration of experience were all significantly associated with pain management decisions. These findings suggest that pain assessment and treatment decisions may be impacted by the health care providers' demographic characteristics, effects which may contribute to pain management disparities."(Bartley et al., 2015).

The aim of this quality measure is to assist eligible providers to identify patients experiencing pain and provide a follow-up plan which addresses the patients' pain in an effort to reduce or eliminate the pain. Ultimately, reducing or eliminating pain will improve a patients' quality of life, minimize the disparities that exist in the assessment and treatment of pain and reduce the cost and utilization of healthcare resources.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
This PQRS measure is designed to encourage and improve the documentation and reporting of standardized pain assessments. It is scored as a simple count of valid submissions on payment claims in the time frame where Part B Medicare claims were available for analysis.

The measure is constructed so that a performance score can be easily derived by dividing the number of claims with codes indicating that the recommended processes were followed (or that the patient was ineligible) by the total number of numerator G codes submitted.

2014 Performance Scores: Claims data consists of all Medicare Part B claims submitted from 1/1/2014 to 12/31/2014 with one of the numerator G codes for this measure. The numerator G code submissions are voluntary and providers who report may not be representative of all eligible professionals. Performance rates cannot be generalized to the population.

A. Quality Indicator Performance 1/1/2014 through 12/31/2014
1. Total Claims Submitted- 10,555,143
2. Valid Denominator Criteria - 9,515,468/ 90.2% of total
3. Performance Exclusion – 341,159/ 3.5% of valid
4. Measure Performance Rate- 7,627,424 / 9,174,309 83.1%

B. Performance Variation by Eligible Professional 1/1/2014 through 12/31/2014: Describes the variation of measure scores by discrete National Provider Identification (NPI).
- N (# of NPIs) – 59,722
- Mean Measure Score – 81.9%
- Standard Deviation - .35
- Min/Max – 0/100%
- 1st percentile – 0.0%
- 5th percentile – 0.0%
- 10th percentile – 0.0%
- 25th percentile – 90.6%
- 50th percentile – 100.0%

Performance scores for the majority of reporting providers skew high (90.6% at the 25th percentile) but drop off sharply for the below the 25th percentile (0% at the 10th percentile). As the eligible provider pool has expanded average performance rates decreased (97.4% in 2009, 88.5% in 2014).

Reporting for the measure is voluntary and providers who report may not be representative of all eligible professionals. In 2014 only 10.7% of eligible providers reported this measure. Reported performance rates from this group cannot be generalized to the total eligible population

Provider and Patients Statistics for program year 2014 (from "2014 Physician Quality Reporting System Program Monitoring and Evaluation Report"):
Average Performance Rates by Year (PQRS – all reporting methods):
2009 – 97.4%
2010 – 97.3%
2011 – 94.8%
2012 – 86.9%
2013 – 85.7%
2014 - 88.5%

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.**
n/a

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 endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Disparities in performance based on race/ethnicity, urban/rural status, gender and age were identified. Analysis of claims from 1/1/2014 through 12/31/2014 reveal statistically significant differences in measure performance between genders and age groups with larger differences observed between urban/rural providers and patient race/ethnic group.

Performance rates by categories:
Rural 87.3%, Urban 81.8% (X2 = 34753.95, N = 9,159,741 p < .0001)
Female 83.7%, Male 82.2 % (X2 = 3424.87, N = 9,174,309 p < .0001)
White 84.2%, Non-white 70.6% (X2 = 85850.38, N = 9,002,090 p < .0001)
Asian 76.2%, Black 68.2%, Hispanic 79.1%, Native 73.6%, White 84.2%, Other 79.6%, Unknown 86.1%(X2 = 95281.16, N = 9,174,309 p < .0001))
Age Under 50 years 80.0%, 50-64 years 80.9%, 65-69 years 85.4%, 70-74 years 84.6%, >=75 81.7% (X2 = 23394.64, N = 9,174,309, p < .0001)

Refer to section IV. Analysis of Claims Data in attached "NQF Endorsement Measurement Submission Summary Materials" document

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
n/a

1c. High Priority (previously referred to as High Impact)
The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, High resource use, Patient/societal consequences of poor quality, Severity of illness
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

The American Pain Foundation (2009) identified pertinent facts related to the impact of pain as follows:

• Uncontrolled pain is a leading cause of disability and diminishes quality of life for patients, survivors, and their loved ones. It interferes with all aspects of daily activity, including sleep, work, social and sexual relations.

• Under-treated pain drives up costs – estimated at $100 billion annually in healthcare expenses, lost income, and lost productivity– extending length of hospital stays, as well as increasing emergency room trips and unplanned clinic visits.

• Medically underserved populations endure a disproportionate pain burden in all health care settings

• Disparities exist among racial and ethnic minorities in pain perception, assessment, and treatment for all types of pain, whether chronic or acute.

The Institute Of Medicine's (IOM) Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education and Research (2011) report suggests that chronic pain rates will continue to increase as a result of:

• More Americans will experience a disease in which chronic pain is associated (diabetes, cardiovascular disease, etc.)

• Increase in obesity which is associated with chronic conditions that have painful symptoms

• Progress in lifesaving techniques for catastrophic injuries for people who would have previously died leads to a group of young people at risk for lifelong chronic pain

• Surgical patients are at risk for acute and chronic pain

• The public has a better understanding of chronic pain syndromes and new treatments and therefore may seek help when they may not have sought help in the past

Gaskin and Richard (2012) studied the economic costs of pain in the United States estimates and reported the national cost of pain ranged from $560 to $635 billion, exceeding the annual costs of heart disease, cancer and diabetes. This study also reported chronic pain affects approximately 100 million adults in the USA. Chronic pain impacts the working lives of those affected as well as Independent Activities of Daily Living (IADLs), sleep and the family as noted by Prefontaine and Rochette (2013). Low back pain and neck pain are two of the diseases with the largest number of years lived with a disability (YLDs) in 2010 as reported by The State of US Health, 1990-2010, Burden of Diseases, Injuries, and Risk Factors (Murray et al., 2013). Inflation adjusted ($2010) biennial expenditures on ambulatory services for chronic back pain increased by 129% from $15.6 billion in 2000-2001 to $35.7 billion in 2006-2007 (Smith, 2013). It is clear the enormous pain-related costs, in both dollars and quality of life, represent a great challenge and an opportunity in terms of improving the quality and cost-effectiveness of care.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

American Pain Foundation (2009). Pain resource guide: Getting the help you need. Retrieved from http://www.peacehealthlabs.org/GeneralPurposeDocuments/Pain%20Resource%20Guide.pdf

Gaskin, D. and Richard, P. (2012). The Economic Costs of Pain in the United States. The Journal of Pain, 13(8), 715-724.

Institute of Medicine (2011). A blueprint for transforming prevention, care, education, and research. Relieving pain in america (269-276). Washington, DC: The National Academies Press. Retrieved from: http://www.nap.edu/download.php?record_id=13172#

Murray, C.J., Abraham, J., Ali, M.K., Alvarado, M., Atkinson, C., Baddour, L.M.…Lopez, A.D. (2013). The State of US Health, 1990-2010, Burden of Diseases, Injuries, and Risk Factors. JAMA; 310(6), 591-608. doi:10.1001/jama.2013.13805

Prefontaine, K. & Rochette, A. (2013). A literature review on chronic pain: the daily overcoming of a complex problem. British Journal of Occupational Therapy, 76(6), 280-286. DOI: 10.4276/030802213X13706169932905

Smith, M., Davis, M.A., Stano, M., &, Whedon, J. M. (2013). Aging baby boomers and the rising cost of chronic back pain: secular trend analysis of longitudinal medical expenditures panel survey data for years 2000 to 2007. Journal of Manipulative and Physiological Therapeutics, 36(1), 1-9.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

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 subcriteria 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):*
Musculoskeletal, Musculoskeletal : Low Back Pain, Prevention : Screening

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Functional Status, Health and Functional Status, Health and Functional Status : Functional Status, Prevention, Prevention : Screening

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
https://www.cms.gov/apps/ama/license.asp?file=/PQRS/Downloads/2016_PQRS_IndMeasuresSpecs_ClaimsRegistry_010716.zip
https://www.cms.gov/apps/ama/license.asp?file=/PQRS/Downloads/2016_PQRS_IndivMeasures_SingleSource_12182015.xlsx

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:** Data_Dictionary_033016.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
2011 Measure Specification: The Instructions were updated to reflect the term "eligible professional" in place of 'non-MD/DO clinicians'; the numerator statement now includes the word "therapy" as stated in the title, for consistency; definition of "Standardized Tool" was updated to include examples of standardized tools; updated description of G-Codes by substituting the word "therapy" for the word 'treatment."

2012 Measure Specification: the title was updated from "Pain Assessment Prior to Initiation of Patient Therapy and Follow-up" to "Pain Assessment and Follow-Up" to avoid confusion regarding the term "prior to the initiation of therapy;" minor language changes to the Description, Numerator and Instructions; added definition of Pain Assessment; updated Definition of Not Eligible, Standardized Tool, Follow-Up Plan and Not Eligible; deleted definition of Qualifying Visit; added Wellness codes G0402, G0438 and G0439 HCPCS code G0101 (cervical or vaginal cancer screening; pelvic and clinical breast examination) and 'office or outpatient visit for the evaluation of new or established patient' codes to Denominator Coding to allow a broader base of providers to report; deleted Denominator CPT Code 99211, this is a five minute office or outpatient visit; replaced Numerator Option codes G-Code G8440, G8508, and G8441 with G8730, G8731, and G8732 which contained more specific descriptions of the quality action performed.

2013 Measure Specification: Minor language changes to Description, added clarifying language to the Instructions linking the follow-up plan to the presence of pain; minor language change to Denominator Statement; added denominator codes for treatment of speech, language, voice, communication, and/or auditory processing disorder, treatment of swallowing dysfunction and/or oral function for feeding and a code for development of cognitive skills to improve attention, memory, and problem solving to allow eligible provider reporting; added quality action numerator code G8939 - Pain assessment documented, follow-up plan not documented, patient not eligible/appropriate for improved reporting; updated psychiatric diagnostic evaluation codes; minor language changes to Numerator Definitions including the removal of "patient refuses to participate" and 'diagnosis/condition/illness is not situationally related to pain" from the definition of Not Eligible; G-code description language

added for ease of reporting and minor language changes to the G-code definitions which do not change the intent of the quality action code.

2014 Measure Specification: Updated description by removing the phrase 'through discussion with the patient'; provided additional example of a follow-up in the Instructions; added ophthalmological, physical therapy, occupational therapy, dental and neuropsychological testing CPT codes to the denominator coding to broaden eligible provider reporting; added Numerator Note to assist providers with the documentation of the use of a standardized pain assessment tool and included an exception to this documentation; updated Numerator Definitions of Pain Assessment and Follow-Up; all G-code definitions updated by providing more detail.

2015 Measure Specification: Addition of health and behavior assessment denominator CPT code, 96151.

2016 Measure Specification: Updated National Quality Strategy Domain to "Communication and Care Coordination".

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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
~~2013 Specification Numerator Statement (used in Registry Data Testing):~~
~~Percentage of visits for patients aged 18 years and older with documentation of a pain assessment through discussion with the patient including the use of a standardized tool(s) on each visit AND documentation of a follow-up plan when pain is present (Testing completed on Registry Data)~~

~~2014 and 2016 Numerator Statement (used in Claims Data Testing):~~
Percentage of visits for patients aged 18 years and older with documentation of a pain assessment using a standardized tool(s) on each visit AND documentation of a follow-up plan when pain is present.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
This measure is to be reported for each visit occurring during the reporting period for patients seen during the reporting period. The reporting period is 12 months from January 1st to December 31st.

S.6. **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, 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.
~~2016 Numerator Details (Note: 2013 and 2014 Numerator Details are similar with minor language edits):~~
Definitions:
Pain Assessment – Documentation of a clinical assessment for the presence or absence of pain using a standardized tool is required. A multi-dimensional clinical assessment of pain using a standardized tool may include characteristics of pain; such as: location, intensity, description, and onset/duration.

Standardized Tool – An assessment tool that has been appropriately normed and validated for the population in which it is used. Examples of tools for pain assessment, include, but are not limited to: Brief Pain Inventory (BPI), Faces Pain Scale (FPS), McGill Pain Questionnaire (MPQ), Multidimensional Pain Inventory (MPI), Neuropathic Pain Scale (NPS), Numeric Rating Scale (NRS), Oswestry Disability Index (ODI), Roland Morris Disability Questionnaire (RMDQ), Verbal Descriptor Scale (VDS), Verbal Numeric Rating Scale (VNRS) and Visual Analog Scale (VAS).

Follow-Up Plan – A documented outline of care for a positive pain assessment is required. This must include a planned follow-up appointment or a referral, a notification to other care providers as applicable OR indicate the initial treatment plan is still in effect. These plans may include pharmacologic and/or educational interventions.

Not Eligible – A patient is not eligible if one or more of the following reason(s) is documented:
• Severe mental and/or physical incapacity where the person is unable to express himself/herself in a manner understood by others. For example, cases where pain cannot be accurately assessed through use of nationally recognized standardized pain assessment tools

• Patient is in an urgent or emergent situation where time is of the essence and to delay treatment would jeopardize the patient's health status

NUMERATOR NOTE: The standardized tool used to assess the patient's pain must be documented in the medical record (exception: A provider may use a fraction such as 5/10 for Numeric Rating Scale without documenting this actual tool name when assessing pain for intensity).

G-codes are defined as Quality Data Codes (QDCs), which are subset of HCPCs II codes. QDCs are non-billable codes that providers will use to delineate their clinical quality actions, which are submitted with Medicare Part B Claims. There are 6 G-code options for this measure.

Numerator Quality-Data Coding Options for Reporting Satisfactorily:
Pain Assessment Documented as Positive AND Follow-Up Plan Documented
(One quality-data code [G8730 or G8731] is required on the claim form to submit this numerator option)
Performance Met: G8730: Pain assessment documented as positive using a standardized tool AND a follow-up plan is documented
OR
Pain Assessment Documented as Negative, No Follow-Up Plan Required
Performance Met: G8731: Pain assessment using a standardized tool is documented as negative, no follow-up plan required

OR

Pain Assessment not Documented Patient not Eligible
(One quality-data code [G8442 or G8939] is required on the claim form to submit this numerator option)
Other Performance Exclusion: G8442: Pain assessment NOT documented as being performed, documentation the patient is not eligible for a pain assessment using a standardized tool
OR
Pain Assessment Documented as Positive, Follow-Up Plan not Documented, Patient not Eligible
Other Performance Exclusion: G8939: Pain assessment documented as positive, follow-up plan not documented, documentation the patient is not eligible

OR

Pain Assessment not Documented, Reason not Given
(One quality-data code [G8732 or G8509] is required on the claim form to submit this numerator option)
Performance Not Met: G8732: No documentation of pain assessment, reason not given
OR
Pain Assessment Documented as Positive, Follow-Up Plan not Documented, Reason not Given
Performance Not Met: G8509: Pain assessment documented as positive using a standardized tool, follow-up plan not documented, reason not.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
All visits for patients aged 18 years and older

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
~~Lists of individual codes with descriptors for the 2013, 2014, and 2016 measure specifications are provided in an Excel file at S.2b~~

~~2013 Specification (used in Registry Data Testing):~~
~~Denominator Criteria (Eligible Cases): Patient encounter during the reporting period (CPT or HCPCS): 90791, 90792, 92507, 92508, 92526, 96116, 96150, 97001, 97003, 97532, 98940, 98941, 98942, 99201, 99202, 99203, 99204, 99205, 99212, 99213, 99214, 99215, G0101, G0402, G0438, G0439~~

~~2014 Specification (used in Claims Data Testing):~~
~~Denominator Criteria (Eligible Cases): Patient encounter during the reporting period (CPT or HCPCS): 90791, 90792, 92002, 92004, 92012, 92014, 92507, 92508, 92526, 96116, 96118, 96150, 97001, 97002, 97003, 97004, 97532, 98940, 98941, 98942, 99201, 99202, 99203, 99204, 99205, 99212, 99213, 99214, 99215, D7140, D7210, G0101, G0402, G0438, G0439~~
~~(Denominator codes for ophthalmological, physical therapy, occupational therapy, dental and neuropsychological testing were added: CPT codes 92002, 92004, 92012, 92014, D7140, D7210, 97002, 97004 and 96118)~~

~~2016 Specification:~~
Denominator Criteria (Eligible Cases): Patient encounter during the reporting period (CPT or HCPCS): 90791, 90792, 92002, 92004, 92012, 92014, 92507, 92508, 92526, 96116, 96118, 96150, 96151, 97001, 97002, 97003, 97004, 97532, 98940, 98941, 98942, 99201, 99202, 99203, 99204, 99205, 99212, 99213, 99214, 99215, D7140, D7210, G0101, G0402, G0438, G0439

Lists of individual codes with descriptors for the measure specifications are provided in an Excel file at S.2b

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Not Eligible – A patient is not eligible if one or more of the following reason(s) is documented:

Severe mental and/or physical incapacity where the person is unable to express himself/herself in a manner understood by others. For example, cases where pain cannot be accurately assessed through use of nationally recognized standardized pain assessment tools

Patient is in an urgent or emergent situation where time is of the essence and to delay treatment would jeopardize the patient's health status

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
Pain Assessment not Documented Patient not Eligible
(One quality-data code [G8442 or G8939] is required on the claim form to submit this numerator option)
Other Performance Exclusion: G8442: Pain assessment NOT documented as being performed, documentation the patient is not eligible for a pain assessment using a standardized tool
OR
Pain Assessment Documented as Positive, Follow-Up Plan not Documented, Patient not Eligible
Other Performance Exclusion: G8939: Pain assessment documented as positive, follow-up plan not documented, documentation the patient is not eligible

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
All eligible patients are subject to the same numerator criteria

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
n/a

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
n/a

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
Satisfactory reporting criteria are met by valid submission of one of six G codes on claims that meet denominator criteria.
A rate of quality performance is calculated by dividing the number of records with G codes indicating that the quality actions were performed or that the patient was not eligible by total number of valid G code submissions.

THIS SECTION PROVIDES DEFINITIONS & FORMULAS FOR THE NUMERATOR (A), TOTAL DENOMINATOR POPULATION (TDP), DENOMINATOR EXCLUSIONS (B) CALCUATION & PERFORMANCE DENOMINATOR (PD) CALCULATION.

NUMERATOR (A): HCPCS Clinical Quality Codes G8730, G8731

TOTAL DENOMINATOR POPULATION (TDP): Patient aged 18 years and older on the date of the encounter of the 12-month reporting period, with denominator defined encounter codes & Medicare Part B Claims reported HCPCS Clinical Quality Codes G8730, G8731, G8442, G8939, G8732, G8509

DENONINATOR EXCLUSION (B): HCPCS Clinical Quality Code G8442, G8939

DENOMINATOR EXCLUSION CALCULATION: Denominator Exclusion (B): # of patients with valid exclusions # G8442+G8939 / # TDP

PERFORMANCE DENOMINATOR CALCULATION: Performance Denominator (B): Patients meeting criteria for performance denominator calculation # A / (# TDP - # B)

(Refer to section V. Measure Logic Flow Diagram for Performance Rate Calculation in attached "NQF Endorsement Measurement Submission Summary Materials" Document)

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
n/a

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
n/a

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
n/a

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.

Administrative claims, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
The data source is the patient medical record. Medicare Part B claims data and registry data is provided for test purposes.

S.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Group/Practice, Clinician : Individual

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic, Ambulatory Care : Outpatient Rehabilitation, Behavioral Health/Psychiatric : Outpatient
If other:

S.28. **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.)*
n/a

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
NQF_0420_Testing_Attachment_033016.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 0420
Measure Title: Pain Assessment and Follow-Up
Date of Submission: 3/30/2016
All the information in this form is updated from last endorsement of NQF 0420 in September 2011. This NQF testing form was not in existence in 2010/2011. Testing continues to support measure specification.

Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☒ Process
☐ 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, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** 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 (2b2-2b6) 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 20 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 sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering 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 **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. 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 **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [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]

2b4. 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 sociodemographic 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.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), 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.

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.23***)	**Measure Tested with Data From:**
☒ abstracted from paper record	☒ abstracted from paper record
☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

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*).

2014 Part B Medicare claims data for HCPCS codes G8730, G8731, G8442, G8939, G8732, G8509.

2013 PQRS Administrative Data for claims and registry

1.3. What are the dates of the data used in testing? Registry/Claims: 1/1/2013 – 12/31/2013, Claims: 1/1/2014 – 12/31/2014

Part B Medicare claims data for encounters from 1/1/2014 to 12/31/2014 were analyzed for performance gaps and variation.

Performance data aggregated at the provider level from PQRS Administrative Data for claims and registry for encounters from 1/1/2013 to 12/31/2013 were analyzed for signal to noise reliability.

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:	**Measure Tested at Level of:**

(*must be consistent with levels entered in item S.26*)	
☒ 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*)

Data element reliability/validity sample (1/1/2014 – 12/31/2014):
A total of 59,722 unique NPIs reported the measure on 10,555,143 claims.

NPIs that had fewer than ten claims were removed from the dataset. A simple random sample of 160 NPIs was drawn from 46,001 remaining NPIs in the claims database. The records were then stratified by the business location address listed in the NPI registry so that the maximum number of records from each business location was limited to 10 records. This limitation was set so that the providers would not see this task as too burdensome and would be more likely to send in their records. The resulting sample was comprised of 761 claims.

Providers were mailed a letter requesting that they provide the documentation to support the assignment of the numerator code that they had submitted on the claim.

Documentation for 405 claims from 74 providers was received and reviewed.
Records Requested/Returned/Reviewed 761/416/405
Providers Requested/Returned/Reviewed 160/75/74
Provider response rate 46.9%

Performance score reliability data (1/1/2013 – 12/31/2013):
29,398 providers reporting via claims with an average of 167 cases per provider.

5,639 providers reporting via registry with an average of 197 cases per provider.

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*)

Data element reliability/validity sample (1/1/2014-12/31/2014):
Description of the population reporting the measure via claims:
Claims with Valid Denominator Criteria: 9,515,468/10,555,143 (90.2%)
3.6% were reported as performance exclusions with a total reported performance rate of 83.1%.

76.5% Urban
23.6% Rural

61.2% Female
38.8% Male

92.2% Non-underserved
7.8% Underserved (racial/ethnic minority)

0.8% Asian
5.6% Black
0.9% Hispanic
0.3% Native
90.5% White
0.9% Other
0.9% Unknown

4.8% Under 50
10.6% Aged 50-64
26.2% Aged 65 – 69
22.3% Aged 70 – 74
36.2% Aged 75

Performance score reliability data (1/1/2013-12/31/2013):
Total # of cases:
Claims: 5,004,383
Registry: 1,125,002

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.

Date element validity/reliability assessed with Part B Medicare claims with patient level detail from 1/1/2014 – 12/31/2014.

Performance score reliability was assessed using provider level performance data reported for PQRS for 2013.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).
n/a

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*)

Critical data element testing:
Quality Insights of Pennsylvania (Quality Insights) oversees the abstraction of 405 randomly generated Medicare Part B claims records for all 74 unique NPIs/eligible professionals who reported one of the G-codes for the measure during the 1/1/2014 – 12/31/2014 time period. Quality Insights requests the medical record documentation from the NPI/eligible professional for the randomly selected encounter date. The documentation is abstracted and a G-code is assigned by two registered nurse (RN) abstractors, one from Quality Insights and one from an independent reviewer contracted with Quality Insights, according to the measure specifications.

Agreement rates between independent reviewers were calculated (inter-rater reliability) as well as the rate of agreement between the numerator code submitted with the claim and an independent reviewer (critical data element validity. See 2b2. Validity testing). Crude agreement, prevalence adjusted kappa (PAK), Cohen´s kappa values and corresponding confidence intervals were calculated.

Cohen´s kappa represents chance-corrected proportional agreement. High prevalence of responses in a small number of cells is known to produce unexpected results known as the "kappa paradox." When the prevalence of a rating in the population is very high or low the value of kappa may indicate poor reliability even with a high observed proportion of agreement. In some cases, PAK is shown to provide an additional interpretation of agreement when the prevalence of responses is concentrated in a small number of cells.
See also 2b2. Validity testing

Performance measure score:
Reliability was calculated according to the methods outlined in a technical report prepared by J.L. Adams titled "The Reliability of Provider Profiling: A Tutorial" (RAND Corporation, TR-653-NCQA, 2009). In this context, reliability represents the ability of a measure to confidently distinguish the performance of one physician from another. As discussed in the report: "Conceptually, it is the ratio of signal to noise. The signal in this case is the proportion of variability in measured performance that can be explained by real differences in performance. There are 3 main drivers of reliability; sample size, differences between physicians, and measurement error."

According to this approach, reliability is estimated with a beta-binomial model. The beta-binomial model is appropriate for measuring the reliability of pass/fail measures such as those proposed.

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*)

Critical data element testing:
Inter-Rater Reliability:
Numerator crude agreement 95.0%
Prevalence adjusted kappa .90 (CI .86 – .94)
Kappa .87 (CI -.81 – .93)

See also 2b2. Validity testing.

Performance measure score (1/1/2013 – 12/31/2013):

Data source	N	Between-provider variance	Reliability mean	Reliability median	Reliability Std dev	Reliability min/max

Claims	29,398	.105	.994	1.0	.020	.457 - 1.0
Registry	5,639	.214	.996	1.0	.012	.817 – 1.0

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?*)

Critical data element testing:
Inter-rater reliability testing indicates high agreement.

Landis and Koch (1977) have proposed the following as standards for strength of agreement for the kappa coefficient: [less than or equal to] O=poor, .01 -.20=slight, .21 -.40=fair, .41.- 60=moderate, .61-.80=substantial and .81-1 =almost perfect (high). These categories are informal.
See also 2b2. Validity testing.

Performance measure score:
Provider-specific reliability demonstrates a sufficient level of reliability to detect real difference in performance scores.

In general, reliability scores vary from 0.0 to 1.0, with a score of zero indicating that all variation is attributable to measurement error (noise, or variation across patients within providers) whereas a reliability of 1.0 implies that all variation is caused by real difference in performance across accountable entities.

There is not a clear cut-off for minimum reliability level. Values above 0.7, however, are considered sufficient to see differences between some physicians (or clinics) and the mean, and values above 0.9 are considered sufficient to see differences between pairs of physicians (see RAND tutorial, 2009).

_____

2b2. VALIDITY TESTING
2b2.1. 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*)

2b2.2. 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)*

Critical data element testing:
Quality Insights of Pennsylvania (Quality Insights) oversees the abstraction of 405 randomly generated Medicare Part B claims records for all 74 unique NPIs/eligible professionals who reported one of the G-codes for the measure during the 1/1/2014 – 12/31/2014 time period. Quality Insights requests the medical record documentation from the NPI/eligible professional for the randomly selected encounter date. The documentation is abstracted and a G-code is assigned by two registered nurse (RN) abstractors, one from Quality Insights and one from an independent reviewer contracted with Quality Insights, according to the measure specifications.

Agreement rates between independent reviewers were calculated (inter-rater reliability) as well as the rate of agreement between the numerator code submitted with the claim and an independent reviewer (critical data element validity). Crude agreement, prevalence adjusted kappa (PAK), Cohen´s kappa values and corresponding confidence intervals were calculated.

Face validity:
Quality Insights of Pennsylvania conducts an Environmental Scan to evaluate the most current research and evidence-based guidelines. The TEP, composed of subject matter specialists and experts with technical measure expertise evaluates the results of the review and provides recommendations based on the scientific merits of the evidence using the Grading of Recommendations Assessment, Development and Evaluation (GRADE). The TEP also reviews and establishes the measure´s ability to capture what it is designed to capture using a consensus process.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Critical data element testing:
Overall Reliability of Claims vs. Independent Review:
Numerator crude agreement 85.9%
Prevalence adjusted kappa .72 (.66 - .79)
Kappa .55 (86% CI .45 - .65)

Face validity:
N/A

2b2.4. 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?*)

Critical data element testing:
There is substantial agreement between claims reporting and independent reviewer.

Landis and Koch (1977) have proposed the following as standards for strength of agreement for the kappa coefficient: [less than or equal to] O=poor, .01 -.20=slight, .21 -.40=fair, .41- .60=moderate, .61-.80=substantial and .81-1 =almost perfect (high). These categories are informal.

Face Validity:
Based on the process of multiple stakeholder input, expert panel discussion and public comment, face and content validity of CMS/Quality Insights measures can be assumed to be established.

2b3. EXCLUSIONS ANALYSIS
NA ☐ no exclusions — *skip to section 2b4*

2b3.1. 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*)

QIP analyzed 10,555,143 claims submitted for this measure. Of those 9,515,468 (90.2%) met the denominator criteria for patient age and relevant CPT codes as defined in the measure specifications. It was from that pool the sample for reliability testing was drawn. Two independent clinical reviewers abstracted 405 cases from 74 providers to assess validity of exclusion criteria in claims reporting for encounters from 1/1/2014 to 12/31/2014.

2b3.2. 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*)

3.6 % of the total number of valid claims were reported as exclusions.

Testing of exclusion criteria agreement demonstrated high reliability in measure reporting. Reliability between two independent clinical reviewers was almost perfect with a PAK = .98, (95% CI=.96 - 1.0) and crude agreement= 99.0%; similarly the "gold standard" clinical reviewer vs. claims agreement was almost perfect with a PAK = .98 (99% CI .97 -1.00), crude agreement=99.2%.

2b3.3. 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*)

Instances of reported exclusions were relatively small (3.6%) of the entire reported population and include:

- Severe mental and/or physical incapacity where the person is unable to express himself/herself in a manner understood by others. For example, cases where pain cannot be accurately assessed through use of nationally recognized standardized pain assessment tools
- Patient is in an urgent or emergent situation where time is of the essence and to delay treatment would jeopardize the patient's health status

Gold standard agreement with claims as well as agreement between two independent reviewers indicates almost perfect agreement.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. 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

2b4.2. If an outcome or resource use 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

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)
n/a

2b4.4a. What were the statistical results of the analyses used to select risk factors?
n/a

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)
n/a

2b4.5. 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 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:** n/a

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*): n/a

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves: n/a

2b4.9. Results of Risk Stratification Analysis: n/a

2b4.10. 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

2b4.11. 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

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE
2b5.1. 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)*

This PQRS measure is designed to encourage and improve the documentation and reporting of a pain assessment using a standardized tool and a follow-up plan if pain present. Performance rates are derived by dividing the number of claims with codes indicating that the recommended processes were followed (or that the patient was ineligible) by the total number of numerator reporting codes submitted.

Variation in performance rates were described by measures of central tendency, variation and percentile rankings. Chi-square was used to test for significant differences between expected and observed performance scores for various populations based on demographic traits.

2b5.2. 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)

Reported provider performance variation (2014):
N – 59,722
Mean – 81.9%
Min – 0.0%, Max – 100.0% Std Deviation .35
50th percentile – 100.0%
25th percentile – 90.6%
10th percentile – 0.0%
1st percentile – 0.0%

The overall performance rate reported via claims for the period 1/1/2014 to 12/31/2014 was 83.1%. The average provider performance rate was 81.9%.

Performance results by population groups:
Rural: 87.3% (n=2,156,781) Urban: 81.8% (n=7,002,960) (X^2 = 34753.94, p < .0001)
Female: 83.7% (n=5,613,407) Male: 82.2 % (n=3,560,902) (X^2 = 3424.87, p < .0001)
White: 84.2% (n=8,302,925) Non-white: 70.6% (n=699,165) (X^2 = 85850.38, p < .0001)
Asian: 76.2% (n=73,065) Black: 68.2% (n=513,909) Hispanic: 79.1% (n=82,542) Native: 73.6% (n=29,649)
White: 84.2% (n=8,302,925) Other: 79.6% (n=86,090) Unknown: 86.1% (n=86,129) (X^2 = 95002.59, p < .0001))
Age Under 50 years: 80.0% (n=436,357) 50-64 years: 80.9% (n=971,945) 65-69 years: 85.4% (n=2,404,142) 70-74 years: 84.6% (n=2,043,705) >=75: 81.7% (n=3,318,160) (X^2 = 23394.64, p < .0001)

2b5.3. 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?*)

Disparities in performance based on age, race/ethnicity, gender, urban/rural status, etc. can be identified if present.

Analysis of 2014 claims reveals a statistically significant difference in measure performance in relation to the provider´s rural/urban designation as well as patient gender, race and age group.

Average reported performance rates are above 80% however the need for improvement can be seen for the lowest 10% reporting (10th percentile 0.0%). It should also be noted that the measure is reported voluntarily and those eligible professionals who chose to report may not be representative of the total population of eligible providers.

2b6. 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 SDS 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 SDS 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.

2b6.1. 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

2b6.2. 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

2b6.3. 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

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*)

Analysis of performance was based on 100% of the cases reported for this measure via claims for the PQRS program from 1/1/2014 to 12/31/2014. Data element validity and inter-rater reliability testing was performed on a random sample of this population (see section 1.5 and 2b.2.).

Performance score reliability testing was performed on 100% of cases reported for the PQRS program via claims and registry from 1/1/2013 to 12/31/2013.

2b7.2. 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*)

The reporting of this measure is voluntary and total number of cases reported represents a small fraction of the total eligible population. Based on the 2014 PQRS Evaluation Report there were 26,978,892 eligible beneficiaries of which 2,212,704 (8.2%) were reported. The total number of eligible providers was 573,233 and 10.7% reported the measure.

2b7.3. 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 number of eligible providers reporting the measure is about 10.7% (3.6% in 2010, 4.5% in 2011, 1.8% in 2012, and 7.4% in 2013).

Because reporting is voluntary the reporting population cannot be said to be representative of the total eligible population. Generalizations to the overall eligible population should not be made.

Greater adoption of the measure, potentially via EHR reporting, will minimize potential bias caused by missing data from those who choose not to report.

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.**

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), 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)*

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.**

At the time of this submission, this measure is not currently being considered as eMeasure.

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.**

No feasibility assessment **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. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

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

In an effort to reduce future variability in measure specification interpretation, the following changes will be reviewed:

1. Simplifying Numerator Quality codes [G8442 or G8939] from two G codes to one G code to identify the "Not Eligible" population.
2. Identify locations in the measure specification to emphasize documentation of the standardized tool

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)*.

None

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.

Planned	Current Use (for current use provide URL)
	Public Reporting Physician Quality Reporting System http://www.cms.gov/PQRS Payment Program Physician Quality Reporting System http://www.cms.gov/PQRS

4a.1. For each CURRENT use, checked above, provide:

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

Public Use
Name: Physician Quality Reporting System (PQRS)
Sponsor: Centers for Medicare and Medicaid Services
Purpose and Geographical Area: PQRS is a national reporting program that uses a combination of incentive payments and payment adjustments to promote reporting of quality information by eligible professionals (EPs). EPs satisfactorily report data on quality measures for covered Physician Fee Schedule (PFS) services furnished to Medicare Part B Fee-for-Service (FFS) beneficiaries. Refer to the following link for additional information: http://www.cms.gov/PQRS

In 2014, there were 573,233 (10.7%) Eligible Professionals who could report NQF# 0420. In 2013, NQF #0420 was the 6th most reported measure within PQRS with 664,929 (7.4%) eligible professionals participating in reporting this measure.

Provider and Patients Statistics for program year 2014 (from "2014 Physician Quality Reporting System Program Monitoring and Evaluation Report"):
Providers
Eligible EPs in 2013-664,929
Eligible EPs in 2014=573,233

% of Eligible EPs who report in 2013=7.4%
% of Eligible EPs who report in 2014=10.7%

Beneficiaries
• Eligible Beneficiaries – 26,978,892
• Beneficiaries reported – 2,212,704
• % of Beneficiaries reported – 8.2%

Many types of providers/specialists report this measure as part of the PQRS as defined by the CPT codes in the measure specification.

Refer to section IV. Analysis of Claims Data in attached "NQF Endorsement Measurement Submission Summary Materials" document

4a.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?*)
n/a

4a.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.*)
n/a

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

Progress on Improvement:
Average Performance Rates by Year based on data from "2014 Physician Quality Reporting System Program Monitoring and Evaluation Report":
2010 – 97.3% (3.6% of eligible providers)
2011 – 94.8% (4.5% of eligible providers)
2012 – 86.9% (1.8% of eligible providers)
2013 – 85.7% (7.4% of eligible providers)
2014 – 88.5% (10.7% of eligible providers)

Eligible Professionals by Year based on data from "2014 Physician Quality Reporting System Program Monitoring and Evaluation Report":
2010 – 170,678
2011 – 177,520
2012 – 705,787
2013 – 664,929
2014 – 573,233

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
It is difficult to say with certainty the reason for the decrease after 2010. These performance rates are submitted voluntarily by providers and cannot be generalized to the total population of eligible providers. The smaller group of early adopters may have been biased towards better performers. As a larger percentage of providers opt to report the measure we would expect to see the aggregate performance rate more closely estimate the true rate for the population.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**

There were no unintended consequences

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)**
0383 : Oncology: Plan of Care for Pain – Medical Oncology and Radiation Oncology (paired with 0384)
0676 : Percent of Residents Who Self-Report Moderate to Severe Pain (Short-Stay)
0677 : Percent of Residents Who Self-Report Moderate to Severe Pain (Long-Stay)
1628 : Patients with Advanced Cancer Screened for Pain at Outpatient Visits
1634 : Hospice and Palliative Care -- Pain Screening
1637 : Hospice and Palliative Care -- Pain Assessment

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
0050 : Osteoarthritis: Function and Pain Assessment/ National Committee for Quality Assurance
0306 : Back Pain: Patient Reassessment/ National Committee for Quality Assurance
0322 : Back Pain: Initial Visit/ National Committee for Quality Assurance
0341 : PICU Pain Assessment on Admission/ National Association of Children´s Hospitals and Related Institutions
0342 : PICU Periodic Pain Assessment/ National Association of Children´s Hospitals and Related Institutions
0523 : Pain Assessment Conducted/ Centers for Medicare and Medicaid Services
0675 : The Percentage of Residents on a Scheduled Pain Medication Regimen on Admission Who Self-Report a Decrease in Pain Intensity or Frequency (Short-stay)/ Centers for Medicare and Medicaid

5a. Harmonization
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 completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Six related measures were identified that are not harmonized with NQF# 0420. The differences between these related measures and the submitted measure NQF# 0420 are listed below: 0383 - Oncology: Plan of Care for Pain – Medical Oncology and Radiation Oncology (paired with 0384 which is unrelated to and non-competing with 0420) - target population is specific to patients with a diagnosis of cancer currently receiving chemotherapy or radiation therapy who report having pain; 0383 does not include the use of a standardized pain assessment tool. Both measures are process measures. Both measures have outpatient care setting.
0676 - Percent of Residents Who Self-Report Moderate to Severe Pain (Short-Stay) – target population is specific to short - stay residents whereas 0420 has a broader outpatient population; 0420 is NOT a self-report measure, it is an eligible provider report; 0676 does not include the use of a standardized pain assessment tool; 0676 does not include documentation of a follow-up plan if pain is present; 0676 is an outcome measure whereas 0420 is a process measure. Care setting for 0676 is long term care/skilled nursing facilities whereas 0420 care setting is outpatient clinician office or outpatient rehabilitation. 0677 - Percent of Residents Who Self-Report Moderate to Severe Pain (Long-Stay) – target population is specific to long - stay residents whereas 0420 has a broader outpatient population; 0420 is NOT a self-report measure, it is an eligible provider report; 0677 does not

include the use of a standardized pain assessment tool; 0677 does not include documentation of a follow-up plan if pain is present; 0677 is an outcome measure whereas 0420 is a process measure. Care setting for 0677 is long term care/skilled nursing facilities whereas 0420 care setting is outpatient clinician office or outpatient rehabilitation. 1628 - Patients with Advanced Cancer Screened for Pain at Outpatient Visits - target population is specific to patients with a diagnosis of advanced cancer; 1628 does not include a follow-up plan if pain is present; Both 1628 and 0420 are process measures; Both measures have outpatient care setting. 1634 - Hospice and Palliative Care -- Pain Screening: target population has no age parameters whereas 0420 has an age range (> 18 yrs.); 1634 target population is specific to hospice and palliative care patients whereas 0420 is not diagnosis specific; 1634 does not include documentation of a follow-up plan if pain is present; Both 1634 and 0420 are process measures; Care setting for 1634 is restricted to Hospice/Hospital/Acute Care Facility, whereas 0420 care setting is outpatient clinician office or outpatient rehabilitation. 1637 – Hospice and Palliative Care—Pain Assessment- target population has no age parameters whereas 0420 has an age range (> 18 yrs.); 1637 target population is specific to hospice and palliative care patients whereas 0420 is not diagnosis specific; 1637 measure focus is clinical assessment within 24hrs of positive screening for pain; 0420 measure focus is performing a screening and a documented follow-up plan not just limited to a clinical assessment; Both are process measures; Care setting for 1637 is restricted to Hospice/Hospital/Acute Care Facility; whereas 0420 care setting is outpatient clinician office or outpatient rehabilitation.

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.)
There are no competing measures.

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:** NQF_Endorsement_Measurement_Submission_Summary_Materials.docx

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services
Co.2 **Point of Contact:** Sophia, Autrey, Sophia.autrey@cms.hhs.gov, 410-786-1158-
Co.3 **Measure Developer if different from Measure Steward:** Centers for Medicare & Medicaid Services
Co.4 **Point of Contact:** Sophia, Autrey, Sophia.autrey@cms.hhs.gov, 410-786-1158-

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.
Through a collaborative process, the Technical Expert Panel (TEP) reviewed the current 2016 measure specifications (description, numerator, denominator, definitions, clinical recommendation, and environmental scan).

Camielle Call, LCSW, MSW, Social Worker, University of Alaska Southeast

Jean Carter, PhD, Psychologist, Washington Psychological Center, P.C.

Ann Marie Feretti, Adv, MS, OTR/L, CHT, Occupational Therapist, PROACTIVE Physical & Hand Therapy

Craig S. Little, DC, FACO, Chiropractor, Independent Practice

Elisa Marks, OTR/L, CHT, Occupational Therapist, Center for Health Enhancement and Rehabilitation (CHEAR)

Gregory M. Martino, PhD, Clinical Psychologist, Independent Practice

William Glancey, Patient/Caregiver representative

Christine Goertz, DC. PhD, Chiropractor, Vice Chancellor for Research and Health Policy, Palmer College of Chiropractic

Deepthi Saxena, MD, Physiatrist, Medical Director, Affiliated Medical Rehabilitation

Donna M. Ulteig, LCSW, Licensed Clinical Social Worker, Psychiatric Services, SC

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2008
Ad.3 **Month and Year of most recent revision:** 09, 2015
Ad.4 **What is your frequency for review/update of this measure?** Annually
Ad.5 **When is the next scheduled review/update for this measure?** 09, 2016

Ad.7 **Disclaimers:** This measure and specifications are provided “as is” without warranty of any kind. This measure does not represent a practice guideline.

Ad.8 **Additional Information/Comments:**

NQF# 0420 Pain Assessment and Follow-Up

National Quality Forum

Person and Family Centered Care Project

Thank you for the opportunity to provide additional information for NQF #0420 Pain Assessment and Follow-Up presented at the Person and Family Centered Care Project meeting on 6/6/16. Because there was a lack of strong consensus among the Standing Committee regarding the evidence criteria, additional information was requested. We provide our response below.

Background Information: NQF #0420 is a process measure originally developed for non-MD/DO eligible professionals. Approximately 10% of eligible professionals reported the measure in 2014. While this represents an increase in total reporting, the performance rate decreased over the same period, demonstrating a gap.

The intent of the measure is for the eligible professional (EP) to screen each patient for the presence of pain using a standardized tool and develop a treatment plan based on the results of the assessment. The ICSI Guideline: Assessment and Management of Chronic Pain notes the "Failure to improve pain and function when a patient is following the plan of care should lead to changes of the plan."

While the measure does not prescribe the pain assessment tool to be used, the use of a standardized tool is a requirement to meet performance when reporting the measure. The choice of the tool should depend on the type of condition (Vianin, M., 2008). Standardized tools such as the Oswestry Disability Index (ODI), McGill Pain Questionnaire, and Brief Pain Inventory provide quantitative scoring which measures progress and function. The EP then re-evaluates the patient to determine progress and revise the treatment plan accordingly. This is in accordance with the guideline.

Evidence Review: The following guidelines were provided to support the re-endorsement of the measure:

- ICSI Guideline: Assessment and Management of Chronic Pain
- ICSI Guideline: Adult Acute and Subacute Low Back Pain
- Low Back Pain: Clinical Practice Guidelines

Pain is difficult to measure due to its subjective and multidimensional nature. Younger et al., 2009 conducted a review of pain instruments and techniques and concluded:

> "Despite the difficulty inherent to measuring pain, there are a number of accepted tools for tracking pain-related treatment outcomes. The proper use of these tools can allow clinicians and researchers to demonstrate both statistically and clinically significant treatment effects. These instruments range from quick, one-item assessments of pain intensity, to long surveys that tap into multiple dimensions of the pain experience and overall functioning. Until more objective physiologic/neurologic measurement techniques are perfected, clinicians who study pain will rely on the careful use of established self-report pain measures."

The area of concern raised by the Committee was whether the assessment of pain followed by a care plan leads to an improved outcome. To our knowledge, since the implementation of the measure, this question has not been studied specifically. However, there is evidence that suggests this may be the case, as described below.

1. Pain assessment followed by exercise is recommended. The ICSI Guideline: Assessment and Management of Chronic Pain which advises pain assessment as the critical first step and recommends exercise as a Level 1 Core Principle.

2. In subacute low-back pain, activity improves outcomes. The ICSI Guideline: Assessment and Management of Chronic Pain references a meta-analysis by Hayden et al. (2005) which concluded:

   "Exercise therapy appears to be slightly effective at decreasing pain and improving function in adults with chronic low-back pain, particularly in healthcare populations. In subacute low-back pain there is some evidence that a graded activity program improves absenteeism outcomes, though evidence for other types of exercise is unclear. In acute low-back pain, exercise therapy is as effective as either no treatment or other conservative treatments."

3. Collaborative intervention results in significant improvement. Dobscha , et al. 2009 conducted a randomized control trial, referenced in the ICSI guideline, with the following results:

   "In this cluster randomized controlled trial, a collaborative intervention resulted in significant improvements in pain disability and intensity and patient-rated global impression of change. Depression severity and pain disability and intensity improved among the patients with depression."

4. Other interventions outlined in care plans have been associated with improved outcomes. For example, adherence to a clinical practice guideline that includes physical therapy for low back pain can result in cost savings. Childs, et. al. 2015 concluded:

   "The potential for cost savings in the Military Health System (MHS) from early guideline adherent physical therapy may be substantial. These results also extend the findings from similar studies in civilian settings by demonstrating an association between early guideline adherent care and utilization and costs in a single payer health system. Future research is necessary to examine which patients with LBP benefit early physical therapy and determine strategies for providing early guideline adherent care."

References

Vianin, M. (2008). Psychometric properties and clinical usefulness of the Oswestry Disability Index. *Journal of Chiropractic Medicine*, *7*(4), 161–163. http://doi.org/10.1016/j.jcm.2008.07.001

Hooten, W.M., Timming, R., Belgrade, M., Gaul, .J, Goertz, M., Haake, B., … Walker, N.(2013). Assessment and management of chronic pain. *Institute for Clinical Systems Improvement* (6th ed.). Retrieved from https://www.icsi.org/_asset/bw798b/ChronicPain.pdf

Goertz, M., Thorson, D., Bonsell, J., Bonte, B., Campbell,.R, Haake B., …, Timming, R. (2012). Adult Acute and Subacute Low Back Pain. *Institute for Clinical Systems Improvement* (15th ed). Retrieved from https://www.icsi.org/_asset/bjvqrj/LBP.pdf

Delitto, A., George, S.Z., Van Dillen, L.R., Whitman, J.M., Sowa, G., Shekelle, P., & Denninger, T.R. (2012). Low back pain. Clinical Practice Guidelines Linked to the International Classification of Functioning, Disability, and Health from the Orthopaedic Section of the American Physical Therapy Association. *Journal of Orthopedic Sports Physical Therapy*, 42(4), A1-A57.

Hayden JA, van Tulder MW, Malmivaara AV, Koes BW. Meta-analysis: exercise therapy for nonspecific low back pain. *Ann Intern Med* 2005; 142:765-75.

Younger J, McCue R, Mackey S. Pain Outcomes: A Brief Review of Instruments and Techniques. Curr Pain Headache Rep. 2009; 13(1):39–43. doi: 10.1007/s11916-009-0009-x.

Dobscha SK, Corson K, Perrin NA, et al. Collaborative Care for Chronic Pain in Primary Care: A Cluster Randomized Trial. *JAMA.* 2009; 301(12):1242-1252. doi:10.1001/jama.2009.377.

Childs JD, Fritz JM, Wu SS, Flynn TW &…George SZ. (2015) Implications of early and guideline adherent physical therapy for low back pain on utilization and costs. BMC Health Services Research, 15:150. DOI 10.1186/s12913-015-0830-3.

Informed Patient Centered (IPC) Hip and Knee Replacement Surgery NQF#2958

The following provides a brief response to the main concerns raised in the committee meeting along with the revised text from the application.

- **There was some confusion regarding "exclusions" – suggest reviewing measure submission, specifically the specification to ensure clarity in this area;**

The denominator statement and denominator exclusions were edited to clarify the target population and the exclusions due to missing responses. These changes were carried through in several places where mentioned in the worksheet/application (De3 and S7, S9, S10, S11).

De.3. Brief Description of Measure: The measure is derived from patient responses to the Hip or Knee Decision Quality Instruments. Participants who have a passing knowledge score (60% or higher) and a clear preference for surgery are considered to have met the criteria for an informed, patient-centered decision.
The target population is adult patients who had a primary hip or knee replacement surgery for treatment of hip or knee osteoarthritis.

Numerator Statement: The numerator is the number of respondents who have an adequate knowledge score (60% or greater) and a clear preference for surgery.
Denominator Statement: The denominator includes the number of respondents from the target population of adults who have undergone primary knee or hip replacement surgery for treatment of knee or hip osteoarthritis.
Denominator Exclusions: Respondents who are missing 3 or more knowledge items do not get a total knowledge score and are excluded. Similarly, respondents who do not indicate a preferred treatment are excluded. No other exclusions as long as the respondent had the procedure for the designated condition.

- **2-year data collection timeframe – suggest reviewing/reconsider based on feedback from the Committee or providing additional data to support the timeframe**

We did not have any set time periods in the initial submission. We edited the description of sampling (S.5. and S.20) and added a clear recommendation for the timing of the survey with respect to the timing of the surgery. We also clarified the look back period for sites to collect responses.

S.5. Time Period for Data *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*

There are no set time periods. It is recommended that patients from the target population are sampled 1-6 months after the procedure, allowing some time for recovery. It would be reasonable for sites to survey patients on a rolling basis and report the measure annually, or when the site has reached a sufficient volume of responses (minimum recommended number is 150 per center).

S.20. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Patients of a particular surgeon or at a particular clinical site (which could be a group of providers or a hospital or other surgical site) who had a primary knee or hip replacement surgery are identified from medical records, claims or in some other way. Sampling should allow time for immediate recovery, while attempting to survey shortly after the procedure, and it is recommended that patients are sampled 1- 6 months after the procedure. Patients can be sampled sequentially, or a pool of such patients who had the procedure in a particular time period (e.g. in the last 3 months) can be created and sampled at a rate that produces the desired number of potential respondents.

- **Establishing reliability at the practice level**

We have edited the methods and results of the reliability analyses at the practice level to clarify the tests done and to include the correlation results.

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*)

1. At the item level, we measured test-retest reliability of the knowledge and preference items from same individuals 4-6 weeks apart. For the knowledge score we examined the intraclass correlation coefficient (ICC) of the knowledge score at time 1 and time 2. The ICC compares the variability of different ratings of the same subject to the total variation across all ratings and all subjects. For the preference item, we examined the kappa between the response at time 1 and response at time 2. The kappa statistic measures agreement for qualitative (categorical) items. It is generally thought to be a more robust measure than simple percent agreement calculation, since κ takes into account the agreement occurring by chance.
2. At the practice level, we randomly split patients at the same clinical site into groups of 25 or larger and correlated the scores; i.e. how well score from one sample's reports correlated with another sample's reports for same decision for same provider group.
3. At the practice level, we also divided data within each site to samples with a minimum size of 25. We then calculated the % with IPC within each sample. The reliability was calculated as variability from site divided by total variability. This is a valid measure of reliability similar to the traditional method of calculation intra-class or intra-rater correlation coefficient (in this case the rater is the site). [See for example, Fleiss J. The Design and Analysis of Clinical Experiments (Wiley Series in Probability and Statistics). Canada: Wiley and Sons, 1999.]

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*)

1. The test-retest reliability of the knowledge score was examined in sample 1 and found to be ICC=0.81 (95% CI 0.71 to 0.87). The test-retest reliability of the item assessing preferred treatment was (Kappa = 0.801).

2. At the practice level, the total sample size is 26 (site 1 has 1 combination, site 2 has 21 combinations, site 3 has 1 combination and site 4 has 3 combinations (sample 1 vs. 2, 2 vs. 3, 1 vs. 3)) and the results of the correlation analyses were 0.805.

3. At the practice level, we had 14 groups (site 1 had 2 samples, site 2 had 7, site 3 had 2 and site 4 had 3) and the reliability was 0.853.

- **Use/usability**

The measure itself is new, but it is based on a patient reported survey has been used by thousands of patients. These questions have been cognitively tested to ensure that they are consistently understood and that answers meaningfully assess patient knowledge and preferences for treatment. We have used the questions proposed in a variety of survey designs: cross-section surveys of adults 40 and older, Medicare beneficiaries known to have had procedures based on claims, and clinical settings in which patients were identified by office staff or via medical records, without any problems.

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 subcriterion 1b).

Brief Measure Information

NQF #: 2776
Corresponding Measures:
De.2. **Measure Title:** Functional Change: Change in Motor Score in Long Term Acute Care Facilities
Co.1.1. **Measure Steward:** Uniform Data System for Medical Rehabilitation, a division of UB Foundation Activities, Inc. and its successor in interest, UDSMR, LLC.
De.3. **Brief Description of Measure:** Change in rasch derived values of motor function from admission to discharge among adult long term acute care facility patients aged 18 years and older who were discharged alive. The timeframe for the measure is 12 months. The measure includes the following 12 items:Feeding, Grooming, Dressing Upper Body, Dressing Lower Body, Toileting, Bowel, Expression, Memory, Transfer Bed/Chair/Wheelchair, Transfer Toilet, Locomotion and Stairs.
1b.1. **Developer Rationale:** The current mandated quality measures for Long Term Acute Care facilities do not adequately address the rehabilitative objectives or functional status of patients. The measures do not allow facilities to substantiate the quality of their restorative care program to CMS or commercial payers. The emphasis on restoration or maintenance of function affected by the patient´s illness or injury is paramount in the episode of care. The primary aim of rehabilitation is to increase function to return the patient to living in the community or a less intensive setting of care. Yet the current measures don't adequately capture function or functional improvement. The motor measure is constructed by utilizing items which are presently collected across the post-acute care continuum. Measures of effectiveness, efficiency, timeliness, resource use and safety are an integral part of the items. There are LTACs that are currently collecting data on the items for outcomes purposes; therefore, it should not be difficult for all LTACs to collect this additional information. The change in motor measure has demonstrated both reliability and validity as results indicated a high overall internal consistency, the ability to capture significant functional gains during rehabilitation, has high discriminative capabilities for rehabilitation patients, and predictive of change in motor function outcomes and likelihood of patient discharge from inpatient rehabilitation to the community.
We feel it is imperative that any quality indicators used for the PAC setting take into account the overriding goal of rehabilitation outcomes, which is to restore and improve function and increase functional independence among individuals receiving rehabilitation, and by doing so allowing the patient the ability to return to a community setting or less intensive setting upon discharge.

S.4. **Numerator Statement:** Average change in rasch derived motor functional score from admission to discharge at the facility level for short term rehabilitation patients. Average is calculated as (sum of change at the patient level/total number of patients). Cases aged less than 18 years at admission to the LTAC or patients who died within the LTAC are excluded.
S.7. **Denominator Statement:** Facility adjusted expected change in rasch derived values, adjusted for CMG (Case Mix Group), based on impairment type, admission functional status, and age.
S.10. **Denominator Exclusions:** Patients age at admission less than 18 years old
Patients who died in the LTAC.

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Paper Medical Records
S.26. **Level of Analysis:** Facility

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?

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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

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

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., the benefits or improvements in quality envisioned by use of this measure)*
The current mandated quality measures for Long Term Acute Care facilities do not adequately address the rehabilitative objectives or functional status of patients. The measures do not allow facilities to substantiate the quality of their restorative care program to CMS or commercial payers. The emphasis on restoration or maintenance of function affected by the patient´s illness or injury is paramount in the episode of care. The primary aim of rehabilitation is to increase function to return the patient to living in the community or a less intensive setting of care. Yet the current measures don't adequately capture function or functional improvement. The motor measure is constructed by utilizing items which are presently collected across the post-acute care continuum. Measures of effectiveness, efficiency, timeliness, resource use and safety are an integral part of the items. There are LTACs that are currently collecting data on the items for outcomes purposes; therefore, it should not be difficult for all LTACs to collect this additional information. The change in motor measure has demonstrated both reliability and validity as results indicated a high overall internal consistency, the ability to capture significant functional gains during rehabilitation, has high discriminative capabilities for rehabilitation patients, and predictive of change in motor function outcomes and likelihood of patient discharge from inpatient rehabilitation to the community. In addition, this measure also can be used to measure maintenance or decline in functional status.
We feel it is imperative that any quality indicators used for the PAC setting take into account the overriding goal of rehabilitation outcomes, which is to restore and improve function and increase functional independence among individuals receiving rehabilitation, and by doing so allowing the patient the ability to return to a community setting or less intensive setting upon discharge.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
While this is a new measure, UDSMR has historical data on all 12items, and we are able to give information on the measure. See measure evaluation form for the trending data.

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.**
N/A

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 endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
See the measure evaluation sheet for disparity data overtime for the measure.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**

N/A

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**

Affects large numbers, Severity of illness

1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

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 subcriteria 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. **Cross Cutting Areas** *(check all the areas that apply):*

Functional Status, Health and Functional Status, Health and Functional Status : Development/Wellness, Health and Functional Status : Functional Status

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.)*

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:** NQF_Submission-635749865761904393.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
N/A

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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Average change in rasch derived motor functional score from admission to discharge at the facility level for short term rehabilitation patients. Average is calculated as (sum of change at the patient level/total number of patients). Cases aged less than 18 years at admission to the LTAC or patients who died within the LTAC are excluded.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
12 months

S.6. **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, 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.
The target population is all LTAC patients, at least 18 years old, who did not die in the LTAC. The numerator is the average change in rasch derived motor functional score from admission to discharge for each patient at the facility level, including items: Eating, Grooming, Dressing Upper Body, Dressing Lower Body, Toileting, Bowel,
Expression, Memory, Transfer Bed/Chair/Wheelchair, Transfer Toilet, Locomotion and Stairs. Average is calculated as: (sum of change at the patient level for all items (Eating, Grooming, Dressing Upper Body, Dressing Lower Body, Toileting, Bowel, Expression, Memory, Transfer Bed/Chair/Wheelchair, Transfer Toilet, Locomotion and Stairs) / total number of patients).

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Facility adjusted expected change in rasch derived values, adjusted for CMG (Case Mix Group), based on impairment type, admission functional status, and age.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Populations at Risk : Dual eligible beneficiaries, Populations at Risk : Individuals with multiple chronic conditions, Populations at Risk : Veterans, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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 target population is all LTAC patients, at least 18 years old, who did not die in the LTAC. Impairment type is defined as the primary medical reason for the LTAC stay (such as stroke, joint replacement, brain injury, etc.). Admission functional status is the expected value of the average of the sum 12 items (Eating, Grooming, Dressing Upper Body, Dressing Lower Body, Toileting, Bowel, Expression, Memory, Transfer Bed/Chair/Wheelchair, Transfer Toilet, Locomotion and Stairs) at the facility level. Age is the age of the patient at the time of admission to the LTAC. The denominator is meant to reflect the expected motor functional change score at the facility, if the facility had the same distribution of CMGs (based on impairment type, functional status at admission, and age at admission). This adjustment procedure is an indirect standardization procedure (observed facility average/expected facility average).

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Patients age at admission less than 18 years old
Patients who died in the LTAC.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)
Living at discharge and age at admission are collected through OASIS.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
See definition of the CMGs in the excel file provided.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Stratification by risk category/subgroup
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
This adjustment procedure is an indirect standarization procedure (observed facility average/expected facility average). The numerator is the facility´s average motor functional change score. The denominator is meant to reflect the expected motor functional change score at the facility, if the facility had the same distribution of CMGs (impairment, functional status at admission, and age at admission).

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**
Ratio
If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
1. Identify all patients during the assessment time frame (12 months).
2. Exclude any patients who died in the LTAC.
3. Exclude any patients who are less than 18 at the time of admission to the LTAC.
3. Calculate the total motor change score for each of the remaining patients (sum of change at the patient level for all items (Eating, Grooming, Dressing Upper Body, Dressing Lower Body, Toileting, Bowel, Expression, and Memory.)
4. Transform the patient level functional change scores to the rasch derived value (as stated in excel file).
5. Calculate the average rasch derived motor change score at the facility level.
6. Using national data and previously described adjustment procedure, calculate the facility´s expected rasch derived average motor change score for the time frame (12 months).
7. Calculate the ratio outcome by taking the observed facility average motor change score/facility´s national expected motor change score.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample*

size.)
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
This measure is not based on a sample, but rather is meant for all patients minus the exclusion criteria.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
This is not a survey/patient reported measure.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
There should not be missing data for this measure as all variables would be required, however, should data be missing, those cases will be deleted from the measure.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Functional Change Form, as seen in the appendix.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Post Acute/Long Term Care Facility : Long Term Acute Care Hospital
If other:

S.28. **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.)*

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
Measure_Testing_Motor_Total_LTAC.docx

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.**
Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)
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)*

ALL data elements are in defined fields in a combination of 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.**

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.**

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. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

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

There are LTACs that are currently using UDSMR and the 12 items in our proposed measure for quality benchmarking, both internally and as a national benchmarking system.

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 Functional Change: Change in Motor Score form (this form includes the items for the motor measure) submitted is copyrighted, however, it can be reproduced and distributed, without modification, for internal reporting of performance data or internal auditing that is for non-commercial purposes, e.g. use by health care providers in connection with their practices. Commercial use is defined as the sale, license, or distribution of the Functional Change: Change in Motor Score form for commercial gain, or incorporation of the Functional Change: Change in Motor Score form into a product or service that is sold, licensed or distributed for commercial gain. Commercial uses of the Functional Change: Change in Motor Score form requires a license agreement between the user and UDSMR. The fees charged for other uses or commercial uses shall be in the range of 0% – 15% per commercial sale.

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.

Planned	Current Use (for current use provide URL)
Public Reporting	Quality Improvement with Benchmarking (external benchmarking to multiple organizations) UDSMR www.udsmr.org Quality Improvement (Internal to the specific organization) UDSMR www.udsmr.org

4a.1. **For each CURRENT use, checked above, provide:**

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

Currently UDSMR provides both internal reporting and national benchmarking for LTACs who subscribe to the UDSMR software/outcomes reporting. The FIM System® is a an outcomes management program for skilled nursing facilities, subacute facilities, long-term care hospitals, Veterans Administration programs, international rehabilitation hospitals, and other related venues of care. The FIM System® enables providers and programs to document the severity of patient disability and the results of medical rehabilitation and establishes a common measure for the comparison of rehabilitation outcomes.

The 12 items in our proposed measure are in use in LTACs in the US. Outcomes based on the functional items are currently used for Quality Improvement with Benchmarking (external benchmarking to multiple organizations) and Quality Improvement (Internal to the specific organization).

4a.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?*)

4a.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.*)

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

N/A

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
This is a new measure.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
There were no unintended negative consequences to individuals or populations during the testing of this measure as previously collected data was used.

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

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 completely harmonized?

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:** Functional_Change_Appendix-635749866379372183.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Uniform Data System for Medical Rehabilitation, a division of UB Foundation Activities, Inc. and its successor in interest, UDSMR, LLC.
Co.2 **Point of Contact:** Paulette, Niewczyk, pniewczyk@udsmr.org, 716-817-7868-
Co.3 **Measure Developer if different from Measure Steward:** Uniform Data System for Medical Rehabilitation, a division of UB Foundation Activities, Inc. and its successor in interest, UDSMR, LLC.
Co.4 **Point of Contact:** Margaret, DiVita, mdivita@udsmr.org, 716-817-7800-

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.

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2016
Ad.3 **Month and Year of most recent revision:** 03, 2016
Ad.4 **What is your frequency for review/update of this measure?** Unknown, new measure
Ad.5 **When is the next scheduled review/update for this measure?** 03, 2017

Ad.6 **Copyright statement:**
Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b6)

Measure Title: Functional Change: Change in Motor Score in Long Term Acute Care Facilities

Date of Submission: 8/12/2015

Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☒ Outcome (*including PRO-PM*)
☐ Cost/resource	☐ Process
☐ 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, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** 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 (2b2-2b6) 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 20 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.

Note: The information provided in this form is intended to aid the Steering 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.

2b2. 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.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [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]

2b4. 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 that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

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.

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. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.

16. 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.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☒ clinical database/registry	☒ clinical database/registry
☒ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other:

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

FIM® instrument data from inpatient rehabilitation facilities, long term acute care facilities, and skilled nursing facilities from the Uniform Data System for Medical Rehabilitation. The UDSMR, a not-for-profit organization affiliated with the UB Foundation Activities, Inc. at the State University of New York at Buffalo, maintains the largest non-governmental database for medical rehabilitation outcomes.

1.3. What are the dates of the data used in testing? Years 2010-2012 were used for the motor measure development (reliability and validity testing, Rasch modeling for establishing psychometric properties of the measure). Years 2002-2013 were used in examining the data trends over time using the self-care measure and patient outcomes of inpatient rehabilitation

An updated analysis was completed as requested. For this update, we have used data from a wider date range to allow for a higher number of facilities to be included in the analysis. We used data from 2002-2007, which included 39 facilities. Included in those 39 facilities were 73.039 patients who were used in the analysis.

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.26*)	**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: patient level/aggregate

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*)

All three post-acute care hospital based venues are included, inpatient rehabilitation facilities (n = 746), long term acute care hospitals (n = 6), and skilled nursing facilities (n = 174). All facilities subscribed to UDSMR for outcomes reporting and severity adjusted benchmark analyses.

Of the 746 inpatient rehabilitation facilities included, 571 (76.5%) were units within an acute care hospital and 175 (23.5%) were free-standing IRFs. Every state in the U.S. were represented among the 746 facilities.

Of the 6 long term acute care hospitals (LTCHs), three were in Massachusetts, one was in Missouri, one was in Michigan, and one was in South Carolina.

Of the 174 skilled nursing facilities (SNFs), 141 (84.4%) were free-standing facilities, and 26 (15.6%) were located in an acute care hospital. Twenty-three of the 50 United States were represented.

An updated analysis was completed as requested. For this update, we have used data from a wider date range to allow for a higher number of facilities to be included in the analysis. We used data from 2002-2007, which included 39 facilities. Included in those 39 facilities were 73,039 patients who were used in the analysis.

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*)

We used a random sample of 11,525 patients for all three venues so that one venue was not over sampled in the analysis (to avoid overrepresentation of IRFs and underrepresentation of SNFs and LTCHs) and comparable case counts were included from each venue of care, IRFs (n = 3,619), LTACs (n = 3,922), and SNFs (n = 3,984). Below is a table displaying the demographic distribution.

	Total	IRFs	LTACs	SNFs
	n = 11,525	n = 3,619	n = 3,922	n = 3,984
Age, mean (SD)	70.2 (15.5)	69.2 (15.4)	76.1 (11.7)	65.2 (16.8)
Age Groups, count (%)				
44 years old or less	748 (6.5)	250 (6.9)	447 (11.4)	51 (1.3)
45 to 65 years old	2,782 (24.1)	961 (26.6)	1,229 (31.3)	592 (14.9)
65 to 74 years old	2,733 (23.7)	858 (23.7)	950 (24.2)	925 (23.2)
75 years and older	5,262 (45.7)	1,550 (42.8)	1,296 (33.0)	2,416 (60.6)
Rehabilitation Impairment Category, count (%)				
Stroke	1,547 (13.4)	784 (21.7)	553 (14.1)	210 (5.3)
Traumatic Brain Dysfunction	395 (3.4)	146 (4)	224 (5.7)	25 (0.6)
Non-traumatic Brain Dysfunction	344 (3)	195 (5.4)	103 (2.6)	46 (1.2)
Traumatic Spinal Cord Dysfunction	129 (1.1)	43 (1.2)	82 (2.1)	4 (0.1)
Non-traumatic Spinal Cord Dysfunction	219 (1.9)	152 (4.2)	54 (1.4)	13 (0.3)
Neurological Conditions	536 (4.7)	396 (10.9)	72 (1.8)	68 (1.7)
Lower Extremity Fracture	736 (6.4)	381 (10.5)	27 (0.7)	328 (8.2)
Lower Extremity Joint Replacement	1,084 (9.4)	363 (10)	46 (1.2)	675 (16.9)
Other Orthopaedic Conditions	670 (5.8)	222 (6.1)	92 (2.3)	356 (8.9)
Lower Extremity Amputation	180 (1.6)	111 (3.1)	40 (1)	29 (0.7)
Other Amputation	20 (0.2)	1 (0)	8 (0.2)	11 (0.3)
Osteoarthritis	39 (0.3)	9 (0.2)	3 (0.1)	27 (0.7)
Rheumatoid and Other Arthritis	50 (0.4)	25 (0.7)	8 (0.2)	17 (0.4)
Cardiac Conditions	601 (5.2)	147 (4.1)	124 (3.2)	330 (8.3)
Pulmonary Disorders	429 (3.7)	47 (1.3)	179 (4.6)	203 (5.1)
Pain Syndromes	114 (1)	29 (0.8)	18 (0.5)	67 (1.7)
Major Multiple Trauma w_o TBI, SCI	182 (1.6)	105 (2.9)	46 (1.2)	31 (0.8)
Major Multiple Trauma with TBI, SCI	110 (1)	58 (1.6)	49 (1.2)	3 (0.1)
Guillain-Barré Syndrome	28 (0.2)	15 (0.4)	12 (0.3)	1 (0)
Miscellaneous	4,102 (35.6)	384 (10.6)	2,181 (55.6)	1537 (38.6)
Burns	10 (0.1)	6 (0.2)	1 (0)	3 (0.1)
Gender, count (%)				
Missing	847 (7.3)	2 (0.1)	5 (0.1)	840 (21.1)
Male	4,991 (43.3)	1,663 (46.0)	2,195 (56)	1,133 (28.4)
Female	5,687 (49.3)	1,954 (54.0)	1,722 (43.9)	2,011 (50.5)

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.

_____

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*)

The validity and reliability of the FIM® instrument, the tool used for this measure, in total are well documented, including inter – and intra-rater reliability[1-7]. The measure proposed, however, uses only a subset of the FIM® instrument items. Therefore, Rasch analysis was conducted to test the psychometric properties of the subset of 12 items within the three venues of post-acute care, IRFs, LTACs, and SNFs. It is understood the proposed measure is intended for the inpatient rehabilitation setting, however we are aware that there has been a number of policy reports indicating the importance for a measure to be capable of use in all inpatient post-acute care venues. Additionally, it is well-recognized that policies such as site neutral payments and bundle payments have been proposed. Our self-care measure is appropriate for use in multiple post-acute care venues, which is a strength of the measure as it is advantageous to collect the exact same items which measure the same construct using the same risk adjustment methodology in all inpatient post-acute care to be able to compare outcomes, quality and value of care by setting and among patients that may have used several post-acute care venues for rehabilitation.

Rasch analysis was used to determine the measure reliability at both the person and item level, as well as internal consistency through the use of Cronbach's alpha. Rasch analysis was also used to determine the fit of each item within the measure (12 items: Eating, Grooming, Dressing Upper Body, Dressing Lower Body, Toileting, Bowel, Expression, Memory and Transfer Bed/Chair/Wheelchair, Transfer Toilet, Locomotion and Stairs.) through infit and outfit statistics and item specific correlations. We used Winsteps 3.73 for the analysis.

In addition, Rasch analysis allows for the conversion of ordinal-level data into interval-level data. Ordinal measures do not inherently act as interval measures, where the difference between one score is equidistance compared to the difference between another two scores, i.e. the difference between a 15 and a 16 in our measure may not reflect the same difference between a 56 and a 57, in terms of difficulty. If the data fit the Rasch model, a result of the analysis is the conversion of the raw ordinal scores to a Rasch derived interval score. This allows for a more precise estimation of differences in functional status both between patients and across facilities.

2a2.3. For each level 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 person-reliability correlation was 0.94. The Cronbach Alpha reliability statistic was 0.95. Item correlations within the measure ranged from 0.65 to 0.84. In addition, the infit and outfit statistics were acceptable for all items (less than 2.0).

In the updated analysis with 39 facilities, the new Cronbach Alpha reliability statistic was 0.965.

For the conversion of the ordinal level measure to an interval measure, we set the Rasch scale at 0 – 100 with a high value indicating more independence. The following figure displays the "ruler" or interval transformation scores for each item in the measure.

```
0   10   20   30   40   50   60   70   80   90  100
|-----+-----+-----+-----+-----+-----+-----+-----+-----+-----| NUM   Item
1              1  :  2 :3: 4: 5  :  6   :  7    7   10  Stairs
|                                    |
|                                    |
1            1  : 2 :3 :4 : 5  :  6   :  7      7   9  Locomotion
1           1  :  2 :3: 4 : 5 :  6   :   7      7   4  Dressing Lower
1           1  :  2 :3 :4 : 5 :   6  :   7      7   5  Toileting
1           1  :  2: 3: 4: 5  :  6   :  7       7   8  Transfer Toilet
1           1  : 2 : 3:4 : 5  :  6   :  7       7   7  Transfer Bed
|                                    |
1         1  :  2 :3 :4 : 5 :   6  :   7        7   3  Dressing Upper
1        1  :  2 :3: 4 : 5 :  6   :   7         7   6  Bowel
1       1  :  2 :3: 4 : 5 :  6   :   7         7   2  Grooming
|                                    |
1      1  :  2 :3 :4 : 5 :   6  :   7           7  12  Memory
1     1  :  2 :3: 4:  5 :  6   :   7           7   1  Eating
|                                    |
1    1  :  2: 3:4 : 5  :  6   :  7             7  11  Expression
|-----+-----+-----+-----+-----+-----+-----+-----+-----+-----| NUM   Item
0   10   20   30   40   50   60   70   80   90  100
```

The ruler shows that the easiest item is Expression, and the hardest Stairs and that the distances between a level 1 and 2 and 6 and 7 are greater than the distances between the remaining levels of each item. When calculated at the total level, the following table displays the Rasch-transformed values at each possible raw value.

TABLE OF MEASURES ON TEST OF 12 Item

SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.
12	.00	17.24	37	37.90	2.28	62	52.00	2.63
13	10.58	8.94	38	38.43	2.27	63	52.73	2.67
14	16.04	6.04	39	38.96	2.26	64	53.47	2.72
15	19.04	4.85	40	39.48	2.25	65	54.25	2.76
16	21.12	4.19	41	40.00	2.24	66	55.05	2.81
17	22.75	3.78	42	40.51	2.23	67	55.88	2.86
18	24.11	3.49	43	41.03	2.23	68	56.74	2.92
19	25.29	3.28	44	41.54	2.23	69	57.64	2.99
20	26.34	3.12	45	42.05	2.23	70	58.58	3.06
21	27.31	2.99	46	42.57	2.23	71	59.57	3.15
22	28.20	2.89	47	43.08	2.24	72	60.63	3.25
23	29.03	2.80	48	43.60	2.25	73	61.76	3.37
24	29.82	2.73	49	44.13	2.26	74	62.98	3.50
25	30.57	2.66	50	44.66	2.28	75	64.30	3.66
26	31.28	2.61	51	45.20	2.29	76	65.75	3.85
27	31.97	2.56	52	45.74	2.31	77	67.37	4.07
28	32.63	2.51	53	46.30	2.34	78	69.19	4.34
29	33.28	2.48	54	46.87	2.36	79	71.29	4.69
30	33.90	2.44	55	47.45	2.39	80	73.79	5.17
31	34.51	2.41	56	48.05	2.42	81	76.91	5.87
32	35.10	2.38	57	48.66	2.45	82	81.15	7.07
33	35.68	2.36	58	49.29	2.49	83	88.16	9.82
34	36.25	2.34	59	49.94	2.52	84	100.00E	17.75
35	36.81	2.32	60	50.61	2.56			
36	37.36	2.30	61	51.29	2.60			

In order to assess the score level reliability across facilities, we have completed an Intraclass Correlation Coefficient (ICC) using the split-half method, as suggested by the NQF PFCC committee staff. We used the updated data from the 39 LTAC facilities for 2002 - 2007. Each facility contained complete records meaning all items at admission and discharge for each patient. Each facility was randomly split into two datasets, and the averages at the facility level for each measure were calculated. We then compared across the facilities to get the ICC. A two-way random effects model was used to estimate the ICC for each measure. Using the definitions in McGraw, KO and Wong, SP (1996), "Forming Inferences About Some Intraclass Correlation Coefficients," *Psychological Methods*, 1(1): 30-46, a split-half ICC based on average measurements and using an agreement definition of the correlation. Thus, total score variance is the denominator of the measure. Thus, the high ICC value shown below shows that there is a high degree of absolute agreement in McGraw and Wong's terms – see their Table 5).

The ICC was 0.905, p <.001. This high ICC demonstrates that there is very high consistency for the motor measure.

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?*)

As mentioned before, the reliability of the FIM® instrument is well known. The results of the analysis for the measure proposed show the reliability holds even when looking at a subset of FIM® items. In addition, we also show the means of the measure by facility below. The sizable range of these (e.g., 8.8-25.6 for mobility, 11.1-20.9 for self-care and 5.6-14.6 for the motor measure) show both that (1) this high ICC is not due to a restricted range and (2) that there are important differences across facilities that can be reliably determined by these measures. Thus, distinctions among the facilities can be seen and the measure is reliable.

_____

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. For each level 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)*

Since the validity of the 18-item FIM® instrument has been well established, we examined the concurrent validity of the motor measure with the total FIM® score, both at admission and discharge. In particular, we used the total FIM score from all 18 items as our gold standard measure in which to test our new motor measure against. The two tests of validity we used were the Pearson correlation coefficient and linear regression to calculate an r-squared which represents the percent of variance of the dependent variable (FIM® total) explained by the independent variable (motor items). In this instance we examined the admission and discharge values separately.

We assessed the predictive validity of the motor measure to determine if the measure predicts outcomes such as: functional change (total functional gain as assessed with the 18 item FIM® instrument (the gold standard)), and likelihood of discharge to the community setting Linear regression was used to determine functional change, whereas the change in self-care was the independent variable, the r-squared value (proportion of change accounted for) and the Pearson correlation coefficient was examined. For discharge disposition, logistic regression was used, admission motor total was the independent variable and the dependent variable was dichotomized as discharge to the community (yes or no)t. We used the C-statistic derived from the area under the ROC curve to determine the discrimination of the model, or the ability of the model to discriminate between those patients s having the outcome of interest or not, as predicted by our measure. In SPSS this is completed by utilizing the

patient level probabilities created during the logistic regression in the ROC curve analysis. The C-statistic ranges from 0.5 (no predictive ability) to 1.0 (perfect discrimination).

We completed all testing for the total data set including all venues, and separately by venue of post-acute care. For all analyses, the Rasch derived values for the motor measure was used. SPSS version 21 was used in the analyses.

For the updated analysis we once again used the Pearson correlation coefficient and linear regression to calculate an r-squared which represents the percent of variance of the dependent variable (FIM® total) explained by the independent variable (motor items).

As before, we assessed the predictive validity of the motor measure to determine if the measure predicts outcomes such as: functional change (total functional gain as assessed with the 18 item FIM® instrument (the gold standard)), and likelihood of discharge to the community setting Linear regression was used to determine functional change, whereas the change in self-care was the independent variable, the r-squared value (proportion of change accounted for) and the Pearson correlation coefficient was examined. For discharge disposition, logistic regression was used, admission motor total was the independent variable and the dependent variable was dichotomized as discharge to the community (yes or no). We used the C-statistic derived from the area under the ROC curve to determine the discrimination of the model, or the ability of the model to discriminate between those patients s having the outcome of interest or not, as predicted by our measure. In SPSS this is completed by utilizing the patient level probabilities created during the logistic regression in the ROC curve analysis. The C-statistic ranges from 0.5 (no predictive ability) to 1.0 (perfect discrimination).

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Concurrent Validity

Correlations: For all venues, our measure at both admission and discharge was highly correlated with the FIM® total, 0.932 ($p < 0.001$) and 0.952 ($p < 0.001$), respectively. The correlations remained highly significantly within each venue of care; IRFs, 0.927 ($p < 0.001$) and 0.963 ($p < 0.001$); LTACs, 0.935 ($p < 0.001$) and 0.953 ($p < 0.001$); SNFs, .944 ($p < 0.001$) and .947 ($p < 0.001$).

Linear Regression: For all venues, when comparing our measure at admission and discharge to the respective FIM® totals, the r-square values were extremely high, 0.962 and 0.982, respectively. The values remained high at the venue specific level as well; IRFs, 0.945 and 0.974; LTACs, 0.968 and 0.985; SNFs, 0.960 and 0.980.

Updated Analysis:

Concurrent Validity:

Correlation: Our measure at both admission and discharge was highly correlated with the FIM® total, 0.876 ($p < .001$), and 0.905 ($p < .001$), respectively.

Linear Regression: : For all venues, when comparing our measure at admission and discharge to the respective FIM® totals, the r-square values were extremely high 0.936 ($p < .001$), and 0.951 ($p < .001$), respectively.

Predictive Validity

Functional Gain: For all venues, when comparing gain in our measure to overall FIM® gain including all items, the correlation was very high, 0.866 ($p < 0.001$). In addition, by venue, the correlations remained strong; IRFs, 0.868 ($p < 0.001$); LTACs, 0.887 ($p < 0.001$); SNFs, 0.837 ($p < 0.001$). The linear regression showed high r-squared values as well; all venues, 0.751; IRFs, 0.754; LTACs, 0.786; SNFs, 0.701.

Discharge Disposition – Community: For all venues, the logistic regression analysis shows that the gain in our measure has good predictive ability for discharge setting (community), with a C-statistic of 0.77. By venue, the results are similar; IRFs, 0.75; LTACs, 0.754.

Updated Analysis

Predictive Validity:

Functional Gain: When comparing gain in our measure to overall FIM® gain including all items, the correlation was very high, 0.985 ($p < 0.001$).

Discharge Disposition - Community: The logistic regression analysis shows that the gain in our measure has good predictive ability for discharge setting (community), with a C-statistic of 0.765 (95% CI .761 - .768), $p < .001$.

The final additional analysis completed (upon request), is the overall range in the motor measure averages at the facility level. The data are shown below.

Facility 38	5.58
Facility 24	6.43
Facility 28	6.58
Facility 39	6.84
Facility 6	6.95
Facility 36	7.53
Facility 5	8.08
Facility 31	8.19
Facility 29	8.21
Facility 30	8.56
Facility 27	8.72
Facility 23	8.81
Facility 7	9
Facility 9	9.44
Facility 12	9.51
Facility 15	9.59
Facility 16	9.93

Facility 2	10.17
Facility 14	10.23
Facility 25	10.39
Facility 34	10.54
Facility 17	10.66
Facility 4	10.68
Facility 3	10.97
Facility 1	11.1
Facility 11	11.23
Facility 21	11.5
Facility 10	11.53
Facility 37	11.56
Facility 19	11.67
Facility 32	11.79
Facility 22	12.11
Facility 20	12.77
Facility 8	12.86
Facility 33	13
Facility 35	13.33
Facility 13	13.76
Facility 26	14.27
Facility 18	14.64

2b2.4. 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 results show good validity across all analyses. The r-squared values were all above 0.8, meaning that the percent of variance explained in the dependent variables by our measure were all more than 80%. In addition, the predictive validity was also high.

The updated analysis continues to show good validity for the motor measure.

_____

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. 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*)

We excluded patients that had expired in the post acute care setting (an unanticipated outcome) and patient aged 18 years and older, both criteria consistent with published literature examining rehabilitation outcomes.

2b3.2. 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*)

2b3.3. 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*)

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?
☐ **No risk adjustment or stratification**
☒ **Statistical risk model with 1 risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use 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.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient 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 and not related to disparities*)

We used Skilled Nursing Facility Case Mix Group as our only adjustment variable through an indirect standardization method.

To calculate the facility's adjusted expected change in Rasch derived values, we use indirect standardization which weights national CMG-specific values by facility-specific CMG proportions. CMG-adjustment derives the expected value based on the case mix and severity mix of each facility. The case mix group classification system groups similarly impaired patients based on functional status at admission or patient severity. This is used for SNFs and IRFs, and the same procedure will be applied to the LTACs. Patients within the same CMG are expected to have similar resource utilization needs and similar outcomes. There are three steps to classifying a patient into a CMG at admission:

1. Identify the patient's impairment group code (IGC).
2. Calculate the patient's weighted motor index score, calculated from 12 of the 13 motor FIM® items.

3. Calculate the cognitive FIM® rating and the age at admission. (This step is not required for all CMGs.)

See file uploaded in S.15 for calculations.

2b4.4. What were the statistical results of the analyses used to select risk factors?

No statistical tests were calculated, CMG adjustment is a standard procedure.

2b4.5. 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 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. 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*)

***2b4.11. 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*)

_____

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE
2b5.1. 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)*

2b5.2. 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*)

2b5.3. 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?*)

_____

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

Note*: This criterion is directed 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).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different datasources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

2b6.2. 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*)

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of 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*)

References

1. Dodds TA, Martin DP, Stolov WC, Deyo RA. A validation of the functional independence measurement and its performance among rehabilitation inpatients. *Archives of physical medicine and rehabilitation.* May 1993;74(5):531-536.

2. Gerrard P, Goldstein R, Divita MA, et al. Validity and Reliability of the FIM(R) Instrument in the Inpatient Burn Rehabilitation Population. *Archives of physical medicine and rehabilitation.* Mar 5 2013.

3. Granger CV, Deutsch A, Russell C, Black T, Ottenbacher KJ. Modifications of the FIM instrument under the inpatient rehabilitation facility prospective payment system. *American journal of physical medicine & rehabilitation / Association of Academic Physiatrists.* Nov 2007;86(11):883-892.

4. Hall KM, Cohen ME, Wright J, Call M, Werner P. Characteristics of the Functional Independence Measure in traumatic spinal cord injury. *Archives of physical medicine and rehabilitation.* Nov 1999;80(11):1471-1476.

5. Keith RA, Granger CV, Hamilton BB, Sherwin FS. The functional independence measure: a new tool for rehabilitation. *Adv Clin Rehabil.* 1987;1:6-18.

6. Ottenbacher KJ, Hsu Y, Granger CV, Fiedler RC. The reliability of the functional independence measure: a quantitative review. *Archives of physical medicine and rehabilitation.* Dec 1996;77(12):1226-1232.

7. Stineman MG, Shea JA, Jette A, et al. The Functional Independence Measure: tests of scaling assumptions, structure, and reliability across 20 diverse impairment categories. *Archives of physical medicine and rehabilitation.* Nov 1996;77(11):1101-1108.

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 subcriterion 1b).

Brief Measure Information

NQF #: 2776
Corresponding Measures:
De.2. **Measure Title:** Functional Change: Change in Motor Score in Long Term Acute Care Facilities
Co.1.1. **Measure Steward:** Uniform Data System for Medical Rehabilitation, a division of UB Foundation Activities, Inc. and its successor in interest, UDSMR, LLC.
De.3. **Brief Description of Measure:** Change in rasch derived values of motor function from admission to discharge among adult long term acute care facility patients aged 18 years and older who were discharged alive. The timeframe for the measure is 12 months. The measure includes the following 12 items:Feeding, Grooming, Dressing Upper Body, Dressing Lower Body, Toileting, Bowel, Expression, Memory, Transfer Bed/Chair/Wheelchair, Transfer Toilet, Locomotion and Stairs.
1b.1. **Developer Rationale:** The current mandated quality measures for Long Term Acute Care facilities do not adequately address the rehabilitative objectives or functional status of patients. The measures do not allow facilities to substantiate the quality of their restorative care program to CMS or commercial payers. The emphasis on restoration or maintenance of function affected by the patient´s illness or injury is paramount in the episode of care. The primary aim of rehabilitation is to increase function to return the patient to living in the community or a less intensive setting of care. Yet the current measures don't adequately capture function or functional improvement. The motor measure is constructed by utilizing items which are presently collected across the post-acute care continuum. Measures of effectiveness, efficiency, timeliness, resource use and safety are an integral part of the items. There are LTACs that are currently collecting data on the items for outcomes purposes; therefore, it should not be difficult for all LTACs to collect this additional information. The change in motor measure has demonstrated both reliability and validity as results indicated a high overall internal consistency, the ability to capture significant functional gains during rehabilitation, has high discriminative capabilities for rehabilitation patients, and predictive of change in motor function outcomes and likelihood of patient discharge from inpatient rehabilitation to the community.
We feel it is imperative that any quality indicators used for the PAC setting take into account the overriding goal of rehabilitation outcomes, which is to restore and improve function and increase functional independence among individuals receiving rehabilitation, and by doing so allowing the patient the ability to return to a community setting or less intensive setting upon discharge.

S.4. **Numerator Statement:** Average change in rasch derived motor functional score from admission to discharge at the facility level for short term rehabilitation patients. Average is calculated as (sum of change at the patient level/total number of patients). Cases aged less than 18 years at admission to the LTAC or patients who died within the LTAC are excluded.
S.7. **Denominator Statement:** Facility adjusted expected change in rasch derived values, adjusted for CMG (Case Mix Group), based on impairment type, admission functional status, and age.
S.10. **Denominator Exclusions:** Patients age at admission less than 18 years old
Patients who died in the LTAC.

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Paper Medical Records
S.26. **Level of Analysis:** Facility

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?

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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

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

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., the benefits or improvements in quality envisioned by use of this measure)*
The current mandated quality measures for Long Term Acute Care facilities do not adequately address the rehabilitative objectives or functional status of patients. The measures do not allow facilities to substantiate the quality of their restorative care program to CMS or commercial payers. The emphasis on restoration or maintenance of function affected by the patient´s illness or injury is paramount in the episode of care. The primary aim of rehabilitation is to increase function to return the patient to living in the community or a less intensive setting of care. Yet the current measures don't adequately capture function or functional improvement. The motor measure is constructed by utilizing items which are presently collected across the post-acute care continuum. Measures of effectiveness, efficiency, timeliness, resource use and safety are an integral part of the items. There are LTACs that are currently collecting data on the items for outcomes purposes; therefore, it should not be difficult for all LTACs to collect this additional information. The change in motor measure has demonstrated both reliability and validity as results indicated a high overall internal consistency, the ability to capture significant functional gains during rehabilitation, has high discriminative capabilities for rehabilitation patients, and predictive of change in motor function outcomes and likelihood of patient discharge from inpatient rehabilitation to the community. In addition, this measure also can be used to measure maintenance or decline in functional status.
We feel it is imperative that any quality indicators used for the PAC setting take into account the overriding goal of rehabilitation outcomes, which is to restore and improve function and increase functional independence among individuals receiving rehabilitation, and by doing so allowing the patient the ability to return to a community setting or less intensive setting upon discharge.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
While this is a new measure, UDSMR has historical data on all 12items, and we are able to give information on the measure. See measure evaluation form for the trending data.

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.**
N/A

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 endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
See the measure evaluation sheet for disparity data overtime for the measure.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**

N/A

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**

Affects large numbers, Severity of illness

1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

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 subcriteria 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. **Cross Cutting Areas** *(check all the areas that apply):*

Functional Status, Health and Functional Status, Health and Functional Status : Development/Wellness, Health and Functional Status : Functional Status

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.)*

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:** NQF_Submission-635749865761904393.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
N/A

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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Average change in rasch derived motor functional score from admission to discharge at the facility level for short term rehabilitation patients. Average is calculated as (sum of change at the patient level/total number of patients). Cases aged less than 18 years at admission to the LTAC or patients who died within the LTAC are excluded.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
12 months

S.6. **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, 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.
The target population is all LTAC patients, at least 18 years old, who did not die in the LTAC. The numerator is the average change in rasch derived motor functional score from admission to discharge for each patient at the facility level, including items: Eating, Grooming, Dressing Upper Body, Dressing Lower Body, Toileting, Bowel,
Expression, Memory, Transfer Bed/Chair/Wheelchair, Transfer Toilet, Locomotion and Stairs. Average is calculated as: (sum of change at the patient level for all items (Eating, Grooming, Dressing Upper Body, Dressing Lower Body, Toileting, Bowel, Expression, Memory, Transfer Bed/Chair/Wheelchair, Transfer Toilet, Locomotion and Stairs) / total number of patients).

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Facility adjusted expected change in rasch derived values, adjusted for CMG (Case Mix Group), based on impairment type, admission functional status, and age.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Populations at Risk : Dual eligible beneficiaries, Populations at Risk : Individuals with multiple chronic conditions, Populations at Risk : Veterans, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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 target population is all LTAC patients, at least 18 years old, who did not die in the LTAC. Impairment type is defined as the primary medical reason for the LTAC stay (such as stroke, joint replacement, brain injury, etc.). Admission functional status is the expected value of the average of the sum 12 items (Eating, Grooming, Dressing Upper Body, Dressing Lower Body, Toileting, Bowel, Expression, Memory, Transfer Bed/Chair/Wheelchair, Transfer Toilet, Locomotion and Stairs) at the facility level. Age is the age of the patient at the time of admission to the LTAC. The denominator is meant to reflect the expected motor functional change score at the facility, if the facility had the same distribution of CMGs (based on impairment type, functional status at admission, and age at admission). This adjustment procedure is an indirect standardization procedure (observed facility average/expected facility average).

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Patients age at admission less than 18 years old
Patients who died in the LTAC.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)
Living at discharge and age at admission are collected through OASIS.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
See definition of the CMGs in the excel file provided.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Stratification by risk category/subgroup
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
This adjustment procedure is an indirect standarization procedure (observed facility average/expected facility average). The numerator is the facility´s average motor functional change score. The denominator is meant to reflect the expected motor functional change score at the facility, if the facility had the same distribution of CMGs (impairment, functional status at admission, and age at admission).

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**
Ratio
If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
1. Identify all patients during the assessment time frame (12 months).
2. Exclude any patients who died in the LTAC.
3. Exclude any patients who are less than 18 at the time of admission to the LTAC.
3. Calculate the total motor change score for each of the remaining patients (sum of change at the patient level for all items (Eating, Grooming, Dressing Upper Body, Dressing Lower Body, Toileting, Bowel, Expression, and Memory.)
4. Transform the patient level functional change scores to the rasch derived value (as stated in excel file).
5. Calculate the average rasch derived motor change score at the facility level.
6. Using national data and previously described adjustment procedure, calculate the facility´s expected rasch derived average motor change score for the time frame (12 months).
7. Calculate the ratio outcome by taking the observed facility average motor change score/facility´s national expected motor change score.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample*

size.)
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
This measure is not based on a sample, but rather is meant for all patients minus the exclusion criteria.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
This is not a survey/patient reported measure.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
There should not be missing data for this measure as all variables would be required, however, should data be missing, those cases will be deleted from the measure.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Functional Change Form, as seen in the appendix.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Post Acute/Long Term Care Facility : Long Term Acute Care Hospital
If other:

S.28. **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.)*

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
Measure_Testing_Motor_Total_LTAC.docx

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.**
Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)
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)*

ALL data elements are in defined fields in a combination of 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.**

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.**

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. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

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

There are LTACs that are currently using UDSMR and the 12 items in our proposed measure for quality benchmarking, both internally and as a national benchmarking system.

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 Functional Change: Change in Motor Score form (this form includes the items for the motor measure) submitted is copyrighted, however, it can be reproduced and distributed, without modification, for internal reporting of performance data or internal auditing that is for non-commercial purposes, e.g. use by health care providers in connection with their practices. Commercial use is defined as the sale, license, or distribution of the Functional Change: Change in Motor Score form for commercial gain, or incorporation of the Functional Change: Change in Motor Score form into a product or service that is sold, licensed or distributed for commercial gain. Commercial uses of the Functional Change: Change in Motor Score form requires a license agreement between the user and UDSMR. The fees charged for other uses or commercial uses shall be in the range of 0% – 15% per commercial sale.

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.

Planned	Current Use (for current use provide URL)
Public Reporting	Quality Improvement with Benchmarking (external benchmarking to multiple organizations) UDSMR www.udsmr.org Quality Improvement (Internal to the specific organization) UDSMR www.udsmr.org

4a.1. **For each CURRENT use, checked above, provide:**

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

Currently UDSMR provides both internal reporting and national benchmarking for LTACs who subscribe to the UDSMR software/outcomes reporting. The FIM System® is a an outcomes management program for skilled nursing facilities, subacute facilities, long-term care hospitals, Veterans Administration programs, international rehabilitation hospitals, and other related venues of care. The FIM System® enables providers and programs to document the severity of patient disability and the results of medical rehabilitation and establishes a common measure for the comparison of rehabilitation outcomes.

The 12 items in our proposed measure are in use in LTACs in the US. Outcomes based on the functional items are currently used for Quality Improvement with Benchmarking (external benchmarking to multiple organizations) and Quality Improvement (Internal to the specific organization).

4a.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?*)

4a.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.*)

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

N/A

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
This is a new measure.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
There were no unintended negative consequences to individuals or populations during the testing of this measure as previously collected data was used.

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

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 completely harmonized?

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:** Functional_Change_Appendix-635749866379372183.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Uniform Data System for Medical Rehabilitation, a division of UB Foundation Activities, Inc. and its successor in interest, UDSMR, LLC.
Co.2 **Point of Contact:** Paulette, Niewczyk, pniewczyk@udsmr.org, 716-817-7868-
Co.3 **Measure Developer if different from Measure Steward:** Uniform Data System for Medical Rehabilitation, a division of UB Foundation Activities, Inc. and its successor in interest, UDSMR, LLC.
Co.4 **Point of Contact:** Margaret, DiVita, mdivita@udsmr.org, 716-817-7800-

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.

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2016
Ad.3 **Month and Year of most recent revision:** 03, 2016
Ad.4 **What is your frequency for review/update of this measure?** Unknown, new measure
Ad.5 **When is the next scheduled review/update for this measure?** 03, 2017

Ad.6 **Copyright statement:**
Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b6)

Measure Title: Functional Change: Change in Self Care Score for Long Term Acute Care Facilities Click here to enter measure title

Date of Submission: 8/12/2015

Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☒ Outcome (*including PRO-PM*)
☐ Cost/resource	☐ Process
☐ 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, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** 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 (2b2-2b6) 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 20 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.

Note: The information provided in this form is intended to aid the Steering 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.

2b2. 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.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [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]

2b4. 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 that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration
OR
- rationale/data support no risk adjustment/ stratification.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

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.
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. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.
16. 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.23*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☒ clinical database/registry	☒ clinical database/registry
☒ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other:

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

Data was from inpatient rehabilitation facilities, long term acute care facilities, and skilled nursing facilities from the Uniform Data System for Medical Rehabilitation. The UDSMR, a not-for-profit organization affiliated with the UB Foundation Activities, Inc. at the State University of New York at Buffalo, maintains the largest non-governmental database for medical rehabilitation outcomes.

1.3. What are the dates of the data used in testing? Years 2010-2012 were used for the self-care measure development (reliability and validity testing, Rasch modeling for establishing psychometric properties of the measure). Years 2010 - 2014 were used in examining the data trends over time using the self-care measure and patient outcomes of skilled nursing facilities

An updated analysis was completed as requested. For this update, we have used data from a wider date range to allow for a higher number of facilities to be included in the analysis. We used data from 2002-2007, which included 39 facilities. Included in those 39 facilities were 73,039 patients who were used in the analysis.

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.26*)	**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: patient level, aggregate

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*)

All three post-acute care hospital based venues are included, inpatient rehabilitation facilities (n = 746), long term acute care hospitals (n = 6), and skilled nursing facilities (n = 174). All facilities subscribed to UDSMR for outcomes reporting and severity adjusted benchmark analyses.

Of the 746 inpatient rehabilitation facilities included, 571 (76.5%) were units within an acute care hospital and 175 (23.5%) were free-standing IRFs. Every state in the U.S. were represented among the 746 facilities.

Of the 6 long term acute care hospitals (LTCHs), three were in Massachusetts, one was in Missouri, one was in Michigan, and one was in South Carolina.

Of the 174 skilled nursing facilities (SNFs), 141 (84.4%) were free-standing facilities, and 26 (15.6%) were located in an acute care hospital. Twenty-three of the 50 United States were represented.

An updated analysis was completed as requested. For this update, we have used data from a wider date range to allow for a higher number of facilities to be included in the analysis. We used data from 2002-2007, which included 39 facilities. Included in those 39 facilities were 73,039 patients who were used in the analysis.

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*)

We used a random sample of 11,525 patients for all three venues so that one venue was not over sampled in the analysis (to avoid overrepresentation of IRFs and underrepresentation of SNFs and LTCHs) and comparable case counts were included from each venue of care, IRFs (n = 3,619), LTACs (n = 3,922), and SNFs (n = 3,984). Below is a table displaying the demographic distribution.

	Total	IRFs	LTACs	SNFs
	n = 11,525	n = 3,619	n = 3,922	n = 3,984
Age, mean (SD)	70.2 (15.5)	69.2 (15.4)	76.1 (11.7)	65.2 (16.8)
Age Groups, count (%)				
44 years old or less	748 (6.5)	250 (6.9)	447 (11.4)	51 (1.3)
45 to 65 years old	2,782 (24.1)	961 (26.6)	1,229 (31.3)	592 (14.9)
65 to 74 years old	2,733 (23.7)	858 (23.7)	950 (24.2)	925 (23.2)
75 years and older	5,262 (45.7)	1,550 (42.8)	1,296 (33.0)	2,416 (60.6)
Rehabilitation Impairment Category, count (%)				
Stroke	1,547 (13.4)	784 (21.7)	553 (14.1)	210 (5.3)
Traumatic Brain Dysfunction	395 (3.4)	146 (4)	224 (5.7)	25 (0.6)
Non-traumatic Brain Dysfunction	344 (3)	195 (5.4)	103 (2.6)	46 (1.2)
Traumatic Spinal Cord Dysfunction	129 (1.1)	43 (1.2)	82 (2.1)	4 (0.1)
Non-traumatic Spinal Cord Dysfunction	219 (1.9)	152 (4.2)	54 (1.4)	13 (0.3)
Neurological Conditions	536 (4.7)	396 (10.9)	72 (1.8)	68 (1.7)
Lower Extremity Fracture	736 (6.4)	381 (10.5)	27 (0.7)	328 (8.2)
Lower Extremity Joint Replacement	1,084 (9.4)	363 (10)	46 (1.2)	675 (16.9)
Other Orthopaedic Conditions	670 (5.8)	222 (6.1)	92 (2.3)	356 (8.9)
Lower Extremity Amputation	180 (1.6)	111 (3.1)	40 (1)	29 (0.7)
Other Amputation	20 (0.2)	1 (0)	8 (0.2)	11 (0.3)
Osteoarthritis	39 (0.3)	9 (0.2)	3 (0.1)	27 (0.7)
Rheumatoid and Other Arthritis	50 (0.4)	25 (0.7)	8 (0.2)	17 (0.4)
Cardiac Conditions	601 (5.2)	147 (4.1)	124 (3.2)	330 (8.3)
Pulmonary Disorders	429 (3.7)	47 (1.3)	179 (4.6)	203 (5.1)
Pain Syndromes	114 (1)	29 (0.8)	18 (0.5)	67 (1.7)
Major Multiple Trauma w_o TBI, SCI	182 (1.6)	105 (2.9)	46 (1.2)	31 (0.8)
Major Multiple Trauma with TBI, SCI	110 (1)	58 (1.6)	49 (1.2)	3 (0.1)
Guillain-Barré Syndrome	28 (0.2)	15 (0.4)	12 (0.3)	1 (0)
Miscellaneous	4,102 (35.6)	384 (10.6)	2,181 (55.6)	1537 (38.6)
Burns	10 (0.1)	6 (0.2)	1 (0)	3 (0.1)
Gender, count (%)				
Missing	847 (7.3)	2 (0.1)	5 (0.1)	840 (21.1)
Male	4,991 (43.3)	1,663 (46.0)	2,195 (56)	1,133 (28.4)
Female	5,687 (49.3)	1,954 (54.0)	1,722 (43.9)	2,011 (50.5)

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.

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*)

The validity and reliability of the FIM® instrument, the tool used for this measure, in total are well documented, including inter – and intra-rater reliability[1-7]. The measure proposed, however, uses only a subset of the FIM® instrument items. Therefore, Rasch analysis was conducted to test the psychometric properties of the subset of 8 items within the three venues of post-acute care, IRFs, LTACs, and SNFs. It is understood the proposed measure is intended for the inpatient rehabilitation setting, however we are aware that there has been a number of policy reports indicating the importance for a measure to be capable of use in all inpatient post-acute care venues. Additionally, it is well-recognized that policies such as site neutral payments and bundle payments have been proposed. Our self-care measure is appropriate for use in multiple post-acute care venues, which is a strength of the measure as it is advantageous to collect the exact same items which measure the same construct using the same risk adjustment methodology in all inpatient post-acute care to be able to compare outcomes, quality and value of care by setting and among patients that may have used several post-acute care venues for rehabilitation.

Rasch analysis was used to determine the measure reliability at both the person and item level, as well as internal consistency through the use of Cronbach's alpha. Rasch analysis was also used to determine the fit of each item within the measure (8 items: Eating, Grooming, Dressing Upper Body, Dressing Lower Body, Toileting, Bowel, Expression, and Memory) through infit and outfit statistics and item specific correlations. We used Winsteps 3.73 for the analysis.

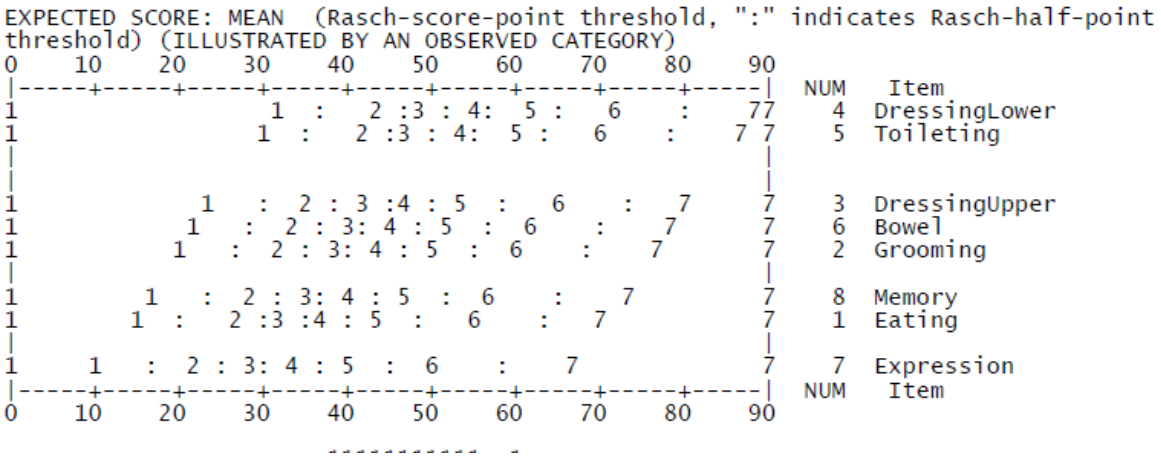
In addition, Rasch analysis allows for the conversion of ordinal-level data into interval-level data. Ordinal measures do not inherently act as interval measures, where the difference between one score is equidistance compared to the difference between another two scores, i.e. the difference between a 15 and a 16 in our measure may not reflect the same difference between a 56 and a 57, in terms of difficulty. If the data fit the Rasch model, a result of the analysis is the conversion of the raw ordinal scores to a Rasch derived interval score. This allows for a more precise estimation of differences in functional status both between patients and across facilities.

2a2.3. For each level 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 person-reliability correlation was 0.89. The Cronbach Alpha reliability statistic was 0.92. Item correlations within the measure ranged from 0.70 to 0.84. In addition, the infit and outfit statistics were acceptable for all items (less than 2.0).

In the updated analysis with 39 facilities, the new Cronbach Alpha reliability statistic was 0.956

For the conversion of the ordinal level measure to an interval measure the Rasch scale was set to 0 – 100 with a high value indicating more independence. The following figure displays the "ruler" or interval transformation scores for each item in the measure.



The ruler shows that the easiest functional item is Expression, and the most challenging functional item is Dressing Lower, additionally, the distances between a level 1 and 2 and 5, 6 and 7 are greater than the distances between the remaining levels of each item. When calculated at the total level, the following table displays the Rasch-transformed values at each possible raw value.

TABLE OF MEASURES ON TEST OF 8 Item

SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.
8	.00E	19.37	25	38.91	3.03	42	54.11	3.50
9	11.92	10.11	26	39.75	3.01	43	55.26	3.61
10	18.17	6.88	27	40.58	3.00	44	56.50	3.72
11	21.64	5.55	28	41.40	3.00	45	57.81	3.86
12	24.08	4.81	29	42.22	3.00	46	59.23	4.01
13	25.99	4.35	30	43.05	3.00	47	60.77	4.18
14	27.60	4.03	31	43.88	3.01	48	62.44	4.37
15	29.00	3.80	32	44.71	3.02	49	64.28	4.60
16	30.26	3.62	33	45.56	3.04	50	66.33	4.88
17	31.42	3.49	34	46.41	3.06	51	68.67	5.24
18	32.50	3.38	35	47.28	3.09	52	71.42	5.73
19	33.52	3.29	36	48.17	3.12	53	74.81	6.48
20	34.49	3.22	37	49.07	3.16	54	79.39	7.78
21	35.42	3.16	38	50.00	3.21	55	86.98	10.87
22	36.33	3.12	39	50.97	3.27	56	100.00E	19.83
23	37.20	3.08	40	51.97	3.34			
24	38.07	3.05	41	53.01	3.41			

In order to assess the score level reliability across facilities, we have completed an Intraclass Correlation Coefficient (ICC) using the split-half method, as suggested by the NQF PFCC committee staff. We used the updated data from the 39 LTAC facilities for 2002 - 2007. Each facility contained complete records meaning all items at admission and discharge for each patient. Each facility was randomly split into two datasets, and the averages at the facility level for each measure were calculated. We then compared across the facilities to get the ICC. A two-way random effects model was used to estimate the ICC for each measure. Using the definitions in McGraw, KO and Wong, SP (1996), "Forming Inferences About Some Intraclass Correlation Coefficients," *Psychological Methods*, 1(1): 30-46, a split-half ICC based on average measurements and using an agreement definition of the correlation. Thus, total score variance is the denominator of the measure. Thus, the high ICC value shown below shows that there is a high degree of absolute agreement in McGraw and Wong's terms – see their Table 5).

The ICC was 0.951, p <.001. This high ICC demonstrates that there is very high consistency for the self-care measure.

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 results of the analysis for the self-care measure were statistically significant, the Cronbach's alpha indicated very high internal consistency, thus a very stable measure.

The updated analyses support our previous interpretation that our measure maintains high reliability. In addition, we also show the means of the measure by facility below. The sizable range of these (e.g., 8.8-25.6 for mobility, 11.1-20.9 for self-care and 5.6-14.6 for the motor measure) show both that (1) this high ICC is not due to a restricted range and (2) that there are important differences across facilities that can be reliably determined by these measures. Thus, distinctions among the facilities can be seen and the measure is reliable.

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. For each level 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)*

Since the validity of the 18-item FIM® instrument has been well established, we examined the concurrent validity of the self-care measure with the total FIM® score, both at admission and discharge. In particular, we used the total FIM score from all 18 items as our gold standard measure in which to

test our new self-care measure against. The two tests of validity we used were the Pearson correlation coefficient and linear regression to calculate an r-squared which represents the percent of variance of the dependent variable (FIM® total) explained by the independent variable (self-care items). In this instance we examined the admission and discharge values separately.

We assessed the predictive validity of the self-care measure to determine if the measure predicts outcomes such as: functional change (total functional gain as assessed with the 18 item FIM® instrument (the gold standard)), and likelihood of discharge to the community setting Linear regression was used to determine functional change, whereas the change in self-care was the independent variable, the r-squared value (proportion of change accounted for) and the Pearson correlation coefficient was examined. For discharge disposition, logistic regression was used, admission self-care total was the independent variable and the dependent variable was dichotomized as discharge to the community (yes or no). We used the C-statistic derived from the area under the ROC curve to determine the discrimination of the model, or the ability of the model to discriminate between those patients having the outcome of interest or not, as predicted by our measure. In SPSS this is completed by utilizing the patient level probabilities created during the logistic regression in the ROC curve analysis. The C-statistic ranges from 0.5 (no predictive ability) to 1.0 (perfect discrimination).

We completed all testing for the total data set including all venues, and separately by venue of post-acute care. For all analyses, the Rasch derived values for the self-care measure was used. SPSS version 21 was used in the analyses.

For the updated analysis we once again used the Pearson correlation coefficient and linear regression to calculate an r-squared which represents the percent of variance of the dependent variable (FIM® total) explained by the independent variable (mobility items).

As before, we assessed the predictive validity of the motor measure to determine if the measure predicts outcomes such as: functional change (total functional gain as assessed with the 18 item FIM® instrument (the gold standard)), and likelihood of discharge to the community setting Linear regression was used to determine functional change, whereas the change in self-care was the independent variable, the r-squared value (proportion of change accounted for) and the Pearson correlation coefficient was examined. For discharge disposition, logistic regression was used, admission motor total was the independent variable and the dependent variable was dichotomized as discharge to the community (yes or no). We used the C-statistic derived from the area under the ROC curve to determine the discrimination of the model, or the ability of the model to discriminate between those patients s having the outcome of interest or not, as predicted by our measure. In SPSS this is completed by utilizing the patient level probabilities created during the logistic regression in the ROC curve analysis. The C-statistic ranges from 0.5 (no predictive ability) to 1.0 (perfect discrimination).

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Concurrent Validity

Correlations: For all venues, our measure at both admission and discharge was highly correlated with the FIM® total, 0.929 ($p < 0.001$) and 0.881 ($p < 0.001$), respectively. The correlations remained significant within each venue of care; IRFs, 0.933 ($p < 0.001$) and 0.896 ($p < 0.001$); LTACs, 0.928 ($p < 0.001$) and 0.888 ($p < 0.001$); SNFs, 0.937 ($p < 0.001$) and 0.871 ($p < 0.001$).

Linear Regression: For all venues, when comparing our measure at admission and discharge to the respective FIM® totals, the r-square values were very high for admission FIM® total and discharge FIM® total, 0.864 and 0.775, respectively. The values remained similar at the venue specific level as well; IRFs, 0.870 and 0.804; LTACs, 0.861 and 0.788; SNFs, 0.877 and 0.758.

Updated Analysis:

Concurrent Validity:

Correlation: Our measure at both admission and discharge was highly correlated with the FIM® total, 0.937 ($p < .001$), and 0.939 ($p < .001$), respectively.

Linear Regression: : For all venues, when comparing our measure at admission and discharge to the respective FIM® totals, the r-square values were extremely high 0.878 ($p < .001$), and 0.882 ($p < .001$), respectively.

Predictive Validity

Functional Gain: For all venues, when comparing gain in our measure to overall FIM® gain including all items, the correlation was strong, 0.721 ($p < 0.001$). In addition, by venue, the correlations remained strong; IRFs, 0.780 ($p < 0.001$); LTACs, 0.757 ($p < 0.001$); SNFs, 0.681 ($p < 0.001$). The linear regression showed significant, high r-squared values as well; all venues, 0.519; IRFs, 0.608; LTACs, 0.574; SNFs, 0.464.

Discharge Disposition – Community: For all venues, the logistic regression analysis shows that the gain in self-care has good predictive ability for discharge setting (community), with a C-statistic of 0.76. By venue, the results are similar; IRFs, 0.74; LTACs, 0.73; SNFs, 0.80.

Updated Analysis

Predictive Validity:

Functional Gain: When comparing gain in our measure to overall FIM® gain including all items, the correlation was moderate at 0.326 ($p < 0.001$).

Discharge Disposition - Community: The logistic regression analysis shows that the gain in our measure has good predictive ability for discharge setting (community), with a C-statistic of 0.729 (95% CI .726 - .733), $p < .001$.

The final additional analysis completed (upon request), is the overall range in the self-care measure averages at the facility level, from smallest to largest. The data are shown below.

Facility 6	11.08
Facility 36	11.77
Facility 28	12.11
Facility 32	12.53
Facility 20	12.62
Facility 16	12.7
Facility 9	12.78
Facility 37	12.9
Facility 30	13.15

Facility 2	13.19
Facility 4	13.31
Facility 25	13.48
Facility 14	13.51
Facility 23	13.61
Facility 29	13.67
Facility 22	13.81
Facility 17	13.91
Facility 26	14.17
Facility 11	14.25
Facility 39	14.36
Facility 1	14.48
Facility 27	14.65
Facility 7	15.07
Facility 38	15.13
Facility 8	15.18
Facility 15	15.57
Facility 5	15.63
Facility 10	16.01
Facility 33	16.03
Facility 19	16.12
Facility 3	16.16
Facility 21	16.33
Facility 34	16.67
Facility 31	17.86
Facility 35	17.87
Facility 12	17.91
Facility 13	18.26
Facility 24	18.4
Facility 18	20.89

2b2.4. 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 results show the self-care measure is valid; the measure demonstrated construct, concurrent, discriminant and predictive validity in all analyses. The r-square values were all consistent, 0.6 or higher, meaning that the percent of variance explained in the dependent variables by our measure were all more than 60%. The predictive validity was also high.

The updated analyses support our previous interpretation that our measure maintains high validity.

2b3. EXCLUSIONS ANALYSIS
NA ☐ no exclusions — *skip to section 2b4*

2b3.1. 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*)

We excluded patients that had expired in the post acute care setting (an unanticipated outcome) and patient aged 18 years and older, both criteria consistent with published literature examining rehabilitation outcomes.

2b3.2. 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*)
No statistical tests completed.

2b3.3. 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*)

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?
☐ **No risk adjustment or stratification**
☒ **Statistical risk model with 1 risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use 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.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient 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 and not related to disparities*)

We used Case Mix Group as our only adjustment variable through an indirect standardization method.

To calculate the facility's adjusted expected change in Rasch derived values, we use indirect standardization which weights national CMG-specific values by facility-specific CMG proportions. CMG-adjustment derives the expected value based on the case mix and severity mix of each facility. The case mix group classification system groups similarly impaired patients based on functional status at admission or patient severity. This is used for SNFs and IRFs, and the same procedure will be applied to the LTACs. Patients within the same CMG are expected to have similar resource utilization needs and similar outcomes. There are three steps to classifying a patient into a CMG at admission:

1. Identify the patient's impairment group code (IGC).
2. Calculate the patient's weighted motor index score, calculated from 12 of the 13 motor FIM® items.
3. Calculate the cognitive FIM® rating and the age at admission. (This step is not required for all CMGs.)

See file uploaded in S.15 for calculations.

2b4.4. What were the statistical results of the analyses used to select risk factors?

No statistical tests were calculated, CMG adjustment is a standard procedure.

2b4.5. 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 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. 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*)

***2b4.11. 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*)

_____

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. 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)*

2b5.2. 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*)

2b5.3. 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?*)

_____

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

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

Note*: This criterion is directed 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).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different datasources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

2b6.2. 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*)

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of 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*)

References

1. Dodds TA, Martin DP, Stolov WC, Deyo RA. A validation of the functional independence measurement and its performance among rehabilitation inpatients. *Archives of physical medicine and rehabilitation.* May 1993;74(5):531-536.

2. Gerrard P, Goldstein R, Divita MA, et al. Validity and Reliability of the FIM(R) Instrument in the Inpatient Burn Rehabilitation Population. *Archives of physical medicine and rehabilitation.* Mar 5 2013.

3. Granger CV, Deutsch A, Russell C, Black T, Ottenbacher KJ. Modifications of the FIM instrument under the inpatient rehabilitation facility prospective payment system. *American journal of physical medicine & rehabilitation / Association of Academic Physiatrists.* Nov 2007;86(11):883-892.

4. Hall KM, Cohen ME, Wright J, Call M, Werner P. Characteristics of the Functional Independence Measure in traumatic spinal cord injury. *Archives of physical medicine and rehabilitation.* Nov 1999;80(11):1471-1476.

5. Keith RA, Granger CV, Hamilton BB, Sherwin FS. The functional independence measure: a new tool for rehabilitation. *Adv Clin Rehabil.* 1987;1:6-18.

6. Ottenbacher KJ, Hsu Y, Granger CV, Fiedler RC. The reliability of the functional independence measure: a quantitative review. *Archives of physical medicine and rehabilitation.* Dec 1996;77(12):1226-1232.

7. Stineman MG, Shea JA, Jette A, et al. The Functional Independence Measure: tests of scaling assumptions, structure, and reliability across 20 diverse impairment categories. *Archives of physical medicine and rehabilitation.* Nov 1996;77(11):1101-1108.

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 subcriterion 1b).

Brief Measure Information

NQF #: 2778
Corresponding Measures:
De.2. **Measure Title:** Functional Change: Change in Mobility Score for Long Term Acute Care Facilities
Co.1.1. **Measure Steward:** Uniform Data System for Medical Rehabilitation, a
De.3. **Brief Description of Measure:** Change in rasch derived values of mobility function from admission to discharge among adult LTAC patients aged 18 years and older who were discharged alive. The time frame for the measure is 12 months. The measure includes the following 4 mobility items:Transfer Bed/Chair/Wheelchair, Transfer Toilet, Locomotion and Stairs.
1b.1. **Developer Rationale:** The current mandated quality measures for Long Term Acute Care facilities do not adequately address the rehabilitative objectives or functional status of patients. The measures do not allow facilities to substantiate the quality of their restorative care program to CMS or commercial payers. The emphasis on restoration or maintenance of function affected by the patient´s illness or injury is paramount in the episode of care. The primary aim of rehabilitation is to increase function to return the patient to living in the community or to another less intensive venue of care. Yet the current measures don't adequately capture function or functional improvement. There are LTACs that are currently collect data on the items in the proposed measure for outcomes purposes; therefore, it should not be difficult for all LTACs to collect this additional information. The change in mobility measure has demonstrated both reliability and validity as results indicated a high overall internal consistency, the ability to capture significant functional gains during rehabilitation, has high discriminative capabilities for rehabilitation patients, and predictive of change in mobility function outcomes and likelihood of patient discharge from inpatient rehabilitation to the community. The current mandated quality measures for LTACs do not adequately address the rehabilitative objectives or functional status of patients. The measures do not allow facilities to substantiate the quality of their restorative care program to CMS or payers. The emphasis on restoration or maintenance of function affected by the patient´s illness or injury is paramount in the episode of care.
We feel it is imperative that any quality indicators used for the PAC setting take into account the overriding goal of rehabilitation outcomes, which is to restore and improve function and increase functional independence among individuals receiving rehabilitation, and by doing so allowing the patient the ability to return to a community setting upon discharge or other less intensive venue of care after their LTAC stay.

S.4. **Numerator Statement:** Average change in rasch derived mobility functional score (Items Transfer Bed/Chair/Wheelchair, Transfer Toilet, Locomotion and Stairs) from admission to discharge at the facility level. Average is calculated as (sum of change at the patient level/total number of patients). Cases aged less than 18 years at admission to the facility or patients who died within the facility are excluded.
S.7. **Denominator Statement:** Facility adjusted adjusted expected change in rasch derived values, adjusted at the Case Mix Group level.
S.10. **Denominator Exclusions:** Excluded in the measure are patients who died in the LTAC or patients less than 18 years old.

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Paper Medical Records
S.26. **Level of Analysis:** Facility

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?

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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
Measure_Evaluation_Mobility_LTAC-635950314051745274.docx

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., the benefits or improvements in quality envisioned by use of this measure)*
The current mandated quality measures for Long Term Acute Care facilities do not adequately address the rehabilitative objectives or functional status of patients. The measures do not allow facilities to substantiate the quality of their restorative care program to CMS or commercial payers. The emphasis on restoration or maintenance of function affected by the patient´s illness or injury is paramount in the episode of care. The primary aim of rehabilitation is to increase function to return the patient to living in the community or to another less intensive venue of care. Yet the current measures don't adequately capture function or functional improvement. There are LTACs that are currently collect data on the items in the proposed measure for outcomes purposes; therefore, it should not be difficult for all LTACs to collect this additional information. The change in mobility measure has demonstrated both reliability and validity as results indicated a high overall internal consistency, the ability to capture significant functional gains during rehabilitation, has high discriminative capabilities for rehabilitation patients, and predictive of change in mobility function outcomes and likelihood of patient discharge from inpatient rehabilitation to the community. The current mandated quality measures for LTACs do not adequately address the rehabilitative objectives or functional status of patients. The measures do not allow facilities to substantiate the quality of their restorative care program to CMS or payers. The emphasis on restoration or maintenance of function affected by the patient´s illness or injury is paramount in the episode of care. In addition, this measure also can be used to measure maintenance or decline in functional status.
We feel it is imperative that any quality indicators used for the PAC setting take into account the overriding goal of rehabilitation outcomes, which is to restore and improve function and increase functional independence among individuals receiving rehabilitation, and by doing so allowing the patient the ability to return to a community setting upon discharge or other less intensive venue of care after their LTAC stay.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Please see measure evaluation form.

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.**
N/A

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 endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Please see measure evaluation form.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from**

the literature that addresses disparities in care on the specific focus of measurement. Include citations.
N/A

1c. High Priority (previously referred to as High Impact)
The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, Severity of illness
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

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 subcriteria 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. **Cross Cutting Areas** *(check all the areas that apply):*
Functional Status, Health and Functional Status, Health and Functional Status : Development/Wellness, Health and Functional Status : Functional Status

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.)*

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:** NQF_Submission_Mobility-635749871757956568.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date 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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Average change in rasch derived mobility functional score (Items Transfer Bed/Chair/Wheelchair, Transfer Toilet, Locomotion and Stairs) from admission to discharge at the facility level. Average is calculated as (sum of change at the patient level/total number of patients). Cases aged less than 18 years at admission to the facility or patients who died within the facility are excluded.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
12 months

S.6. **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, 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.
The target population is all LTAC patients, at least 18 years old, who did not die in the LTAC. The numerator is the average change in rasch derived mobility functional score from admission to discharge for each patient at the facility level, including items: Transfer Bed/Chair/Wheelchair, Transfer Toilet, Locomotion and Stairs. Average is calculated as: (sum of change at the patient level for all items (Transfer Bed/Chair/Wheelchair, Transfer Toilet, Locomotion and Stairs) / total number of patients).

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Facility adjusted adjusted expected change in rasch derived values, adjusted at the Case Mix Group level.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Populations at Risk : Dual eligible beneficiaries, Populations at Risk : Individuals with multiple chronic conditions, Populations at Risk : Veterans, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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 target population is all LTAC patients, at least 18 years old, who did not die in the LTAC. Impairment type is defined as the primary medical reason for the LTAC stay (such as stroke, joint replacement, brain injury, etc.). Admission functional status is the expected value of the average of the sum 4 items (Transfer Bed/Chair/Wheelchair, Transfer Toilet, Locomotion and Stairs) at the facility level. Age is the age of the patient at the time of admission to the LTAC. The denominator is meant to reflect the expected Mobility functional change score at the facility, if the facility had the same distribution of CMGs (based on impairment type, functional status at admission,and age at admission). This adjustment procedure is an indirect standardization procedure (observed facility average/expected
facility average).

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Excluded in the measure are patients who died in the LTAC or patients less than 18 years old.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
Living at discharge and age at admission are collected through OASIS

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables,*

definitions, 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 with at S.2b)
See definition of the CMGs in the excel file provided.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Stratification by risk category/subgroup
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
This adjustment procedure is an indirect standarization procedure (observed facility average/expected facility average). The numerator is the facility´s average mobility functional change score. The denominator is meant to reflect the expected Mobility functional change score at the facility, if the facility had the same distribution of CMGs(impairment, functional status at admission, and age at admission).

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**
Ratio
If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
1. Identify all patients during the assessment time frame (12 months).
2. Exclude any patients who died in the LTAC.
3. Exclude any patients who are less than 18 at the time of admission to the LTAC.
3. Calculate the total mobility change score for each of the remaining patients (sum of change at the patient level for all items (Transfer Bed/Chair/Wheelchair, Transfer Toilet, Locomotion and Stairs.)
4. Transform the patient level functional change scores to the rasch derived value (as stated in excel file).
5. Calculate the average rasch derived mobility change score at the facility level.
6. Using national data and previously described adjustment procedure, calculate the facility´s expected rasch derived average mobility change score for the time frame (12 months).
7. Calculate the ratio outcome by taking the observed facility average mobility change score/facility´s national expected mobility change score.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
This measure is not based on a sample, but rather is meant for all patients minus the exclusion criteria.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
This is not a survey/patient reported measure.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
There should not be missing data for this measure as all variables would be required, however, should data be missing, those cases will be deleted from the measure.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Functional Change Form, as seen in the appendix.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Post Acute/Long Term Care Facility : Long Term Acute Care Hospital
If other:

S.28. **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.)*

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
Measure_Testing_Mobility_LTAC.docx

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.**
Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)
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)*
ALL data elements are in defined fields in a combination of 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.**

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.**
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. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
While this is a new measure, the data collection procedure is in place for LTACs utilizing UDSMR software.

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 Functional Change: Change in Motor Score form (this form includes the items for the mobility measure) submitted is copyrighted, however, it can be reproduced and distributed, without modification, for internal reporting of performance data or internal auditing that is for non-commercial purposes, e.g. use by health care providers in connection with their practices. Commercial use is defined as the sale, license, or distribution of the Functional Change: Change in Motor Score form for commercial gain, or incorporation of the Functional Change: Change in Motor Score form into a product or service that is sold, licensed or distributed for commercial gain. Commercial uses of the Functional Change: Change in Motor Score form requires a license agreement between the user and UDSMR. The fees charged for other uses or commercial uses shall be in the range of 0% – 15% per commercial sale.

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.

Planned	Current Use (for current use provide URL)

Public Reporting	Quality Improvement with Benchmarking (external benchmarking to multiple organizations) UDSMR www.udsmr.org Quality Improvement (Internal to the specific organization) UDSMR www.udsmr.org

4a.1. **For each CURRENT use, checked above, provide:**

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

Currently UDSMR provides both internal reporting and national benchmarking for LTACs who subscribe to the UDSMR software/outcomes reporting. The FIM System® is a an outcomes management program for skilled nursing facilities, subacute facilities, long-term care hospitals, Veterans Administration programs, international rehabilitation hospitals, and other related venues of care. The FIM System® enables providers and programs to document the severity of patient disability and the results of medical rehabilitation and establishes a common measure for the comparison of rehabilitation outcomes.

The FIM System® provides an established means of collecting rehabilitation data in a consistent manner. It allows clinicians to follow changes in the functional status of their patients from the start of rehabilitative care through discharge and follow-up.

4a.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?*)

4a.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.*)

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

N/A

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
N/A

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of**

unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.
As we used existing data that has already been collected, there were no unintended negative consequences to individuals or populations identified during our testing

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

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 completely harmonized?

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:** Functional_Change_Appendix-635749878241675737.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Uniform Data System for Medical Rehabilitation, a
Co.2 **Point of Contact:** Paulette, Niewczyk, pniewczyk@udsmr.org, 716-817-7868-
Co.3 **Measure Developer if different from Measure Steward:** Uniform Data System for Medical Rehabilitation, a
Co.4 **Point of Contact:** Margaret, DiVita, mdivita@udsmr.org, 716-817-7800-

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.

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2016
Ad.3 **Month and Year of most recent revision:** 03, 2016
Ad.4 **What is your frequency for review/update of this measure?** Unknown, new measure
Ad.5 **When is the next scheduled review/update for this measure?** 03, 2017

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b6)

Measure Title: Functional Change: Change in Mobility Score for Long Term Acute Care Facilities
Date of Submission: 8/15/2015
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☒ Outcome (*including PRO-PM*)
☐ Cost/resource	☐ Process
☐ 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, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** 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 (2b2-2b6) 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 20 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.

Note: The information provided in this form is intended to aid the Steering 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.

2b2. 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.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [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]

2b4. 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 that influence the measured outcome (but not factors related to disparities in care or the quality of care)

and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration
OR
- rationale/data support no risk adjustment/ stratification.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

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.
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. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.
16. 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.23*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☒ clinical database/registry	☒ clinical database/registry
☒ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other:

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

Data was from inpatient rehabilitation facilities, long term acute care facilities, and skilled nursing facilities from the Uniform Data System for Medical Rehabilitation. The UDSMR, a not-for-profit organization affiliated with the UB Foundation Activities, Inc. at the State University of New York at Buffalo, maintains the largest non-governmental database for medical rehabilitation outcomes.

1.3. What are the dates of the data used in testing? Years 2010-2012 were used for the mobility measure development (reliability and validity testing, Rasch modeling for establishing psychometric properties of the measure). Years 2002-2013 were used in examining the data trends over time using the mobility measure and patient outcomes of inpatient rehabilitation.

An updated analysis was completed as requested. For this update, we have used data from a wider date range to allow for a higher number of facilities to be included in the analysis. We used data from 2002-2007, which included 39 facilities. Included in those 39 facilities were 73,039 patients who were used in the analysis.

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.26*)	**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: patient level/aggregate

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*)

All three post-acute care hospital based venues are included, inpatient rehabilitation facilities (n = 746), long term acute care hospitals (n = 6), and skilled nursing facilities (n = 174). All facilities subscribed to UDSMR for outcomes reporting and severity adjusted benchmark analyses.

Of the 746 inpatient rehabilitation facilities included, 571 (76.5%) were units within an acute care hospital and 175 (23.5%) were free-standing IRFs. Every state in the U.S. were represented among the 746 facilities.

Of the 6 long term acute care hospitals (LTCHs), three were in Massachusetts, one was in Missouri, one was in Michigan, and one was in South Carolina.

Of the 174 skilled nursing facilities (SNFs), 141 (84.4%) were free-standing facilities, and 26 (15.6%) were located in an acute care hospital. Twenty-three of the 50 United States were represented.

An updated analysis was completed as requested. For this update, we have used data from a wider date range to allow for a higher number of facilities to be included in the analysis. We used data from 2002-2007, which included 39 facilities. Included in those 39 facilities were 73,039 patients who were used in the analysis.

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*)

We used a random sample of 11,525 patients for all three venues so that one venue was not over sampled in the analysis (to avoid overrepresentation of IRFs and underrepresentation of SNFs and LTCHs) and comparable case counts were included from each venue of care, IRFs (n = 3,619), LTACs (n = 3,922), and SNFs (n = 3,984). Below is a table displaying the demographic distribution.

	Total	IRFs	LTACs	SNFs
	n = 11,525	n = 3,619	n = 3,922	n = 3,984
Age, mean (SD)	70.2 (15.5)	69.2 (15.4)	76.1 (11.7)	65.2 (16.8)
Age Groups, count (%)				
44 years old or less	748 (6.5)	250 (6.9)	447 (11.4)	51 (1.3)
45 to 65 years old	2,782 (24.1)	961 (26.6)	1,229 (31.3)	592 (14.9)
65 to 74 years old	2,733 (23.7)	858 (23.7)	950 (24.2)	925 (23.2)
75 years and older	5,262 (45.7)	1,550 (42.8)	1,296 (33.0)	2,416 (60.6)
Rehabilitation Impairment Category, count (%)				
Stroke	1,547 (13.4)	784 (21.7)	553 (14.1)	210 (5.3)
Traumatic Brain Dysfunction	395 (3.4)	146 (4)	224 (5.7)	25 (0.6)
Non-traumatic Brain Dysfunction	344 (3)	195 (5.4)	103 (2.6)	46 (1.2)
Traumatic Spinal Cord Dysfunction	129 (1.1)	43 (1.2)	82 (2.1)	4 (0.1)
Non-traumatic Spinal Cord Dysfunction	219 (1.9)	152 (4.2)	54 (1.4)	13 (0.3)
Neurological Conditions	536 (4.7)	396 (10.9)	72 (1.8)	68 (1.7)
Lower Extremity Fracture	736 (6.4)	381 (10.5)	27 (0.7)	328 (8.2)
Lower Extremity Joint Replacement	1,084 (9.4)	363 (10)	46 (1.2)	675 (16.9)
Other Orthopaedic Conditions	670 (5.8)	222 (6.1)	92 (2.3)	356 (8.9)
Lower Extremity Amputation	180 (1.6)	111 (3.1)	40 (1)	29 (0.7)
Other Amputation	20 (0.2)	1 (0)	8 (0.2)	11 (0.3)
Osteoarthritis	39 (0.3)	9 (0.2)	3 (0.1)	27 (0.7)
Rheumatoid and Other Arthritis	50 (0.4)	25 (0.7)	8 (0.2)	17 (0.4)
Cardiac Conditions	601 (5.2)	147 (4.1)	124 (3.2)	330 (8.3)
Pulmonary Disorders	429 (3.7)	47 (1.3)	179 (4.6)	203 (5.1)
Pain Syndromes	114 (1)	29 (0.8)	18 (0.5)	67 (1.7)
Major Multiple Trauma w_o TBI, SCI	182 (1.6)	105 (2.9)	46 (1.2)	31 (0.8)
Major Multiple Trauma with TBI, SCI	110 (1)	58 (1.6)	49 (1.2)	3 (0.1)
Guillain-Barré Syndrome	28 (0.2)	15 (0.4)	12 (0.3)	1 (0)
Miscellaneous	4,102 (35.6)	384 (10.6)	2,181 (55.6)	1537 (38.6)
Burns	10 (0.1)	6 (0.2)	1 (0)	3 (0.1)
Gender, count (%)				
Missing	847 (7.3)	2 (0.1)	5 (0.1)	840 (21.1)
Male	4,991 (43.3)	1,663 (46.0)	2,195 (56)	1,133 (28.4)
Female	5,687 (49.3)	1,954 (54.0)	1,722 (43.9)	2,011 (50.5)

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.

_____

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*)

The validity and reliability of the FIM® instrument, the tool used for this measure, in total are well documented, including inter – and intra-rater reliability[1-7]. The measure proposed, however, uses only a subset of the FIM® instrument items. Therefore, Rasch analysis was conducted to test the psychometric properties of the subset of 4 items within the three venues of post-acute care, IRFs, LTACs, and SNFs. It is understood the proposed measure is intended for long term acute care facilities, however we are aware that there has been a number of policy reports indicating the importance for a measure to be capable of use in all inpatient post-acute care venues. Additionally, it is well-recognized that policies such as site neutral payments and bundle payments have been proposed. Our mobility measure is appropriate for use in multiple post-acute care venues, which is a strength of the measure as it is advantageous to collect the exact same items which measure the same construct using the same risk adjustment methodology in all inpatient post-acute care to be able to compare outcomes, quality and value of care by setting and among patients that may have used several post-acute care venues for rehabilitation.

Rasch analysis was used to determine the measure reliability at both the person and item level, as well as internal consistency through the use of Cronbach's alpha. Rasch analysis was also used to determine the fit of each item within the measure (4 items: Transfer Bed/Chair/Wheelchair, Transfer Toilet, Locomotion and Stairs.) through infit and outfit statistics and item specific correlations. We used Winsteps 3.73 for the analysis.

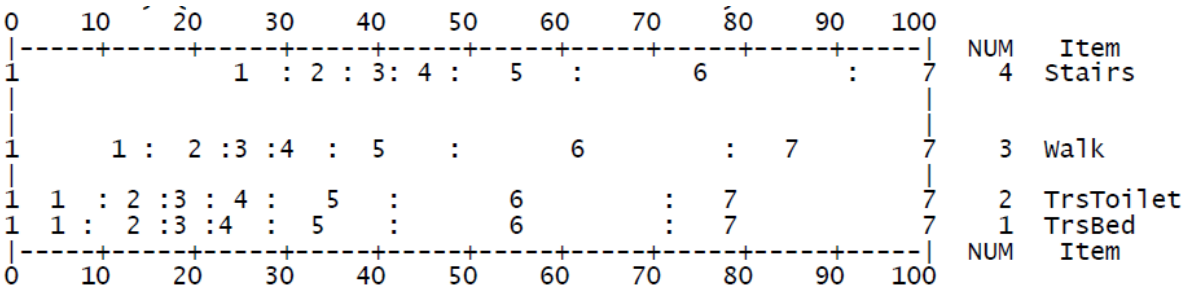
In addition, Rasch analysis allows for the conversion of ordinal-level data into interval-level data. Ordinal measures do not inherently act as interval measures, where the difference between one score is equidistance compared to the difference between another two scores, i.e. the difference between a 15 and a 16 in our measure may not reflect the same difference between a 56 and a 57, in terms of difficulty. If the data fit the Rasch model, a result of the analysis is the conversion of the raw ordinal scores to a Rasch derived interval score. This allows for a more precise estimation of differences in functional status both between patients and across facilities.

2a2.3. For each level 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 person-reliability correlation was 0.89. The Cronbach Alpha reliability statistic was 0.92. Item correlations within the measure ranged from 0.82 to 0.90. In addition, the infit and outfit statistics were acceptable for all items (less than 2.0).

In the updated analysis with 39 facilities, the new Cronbach Alpha reliability statistic was 0.903.

For the conversion of the ordinal level measure to an interval measure, we set the Rasch scale at 0 – 100 with a high value indicating more independence. The following figure displays the "ruler" or interval transformation scores for each item in the measure.



The ruler shows that the easiest item is Transfers: Bed/Chair/Wheelchair, and the hardest Stairs and that the distances between a level 1 and 2 and 5, 6 and 7 are greater than the distances between the remaining levels of each item. When calculated at the total level, the following table displays the Rasch-transformed values at each possible raw value.

TABLE OF MEASURES ON TEST OF 4 Item

SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.
4	.00E	12.48	13	26.23	3.76	22	50.27	5.85
5	8.05	6.65	14	28.41	3.94	23	55.99	6.63
6	12.45	4.65	15	30.76	4.06	24	62.97	7.09
7	15.07	3.91	16	33.17	4.04	25	70.32	7.08
8	17.10	3.59	17	35.50	3.95	26	77.95	7.52
9	18.91	3.46	18	37.76	3.93	27	87.92	9.24
10	20.65	3.45	19	40.08	4.07	28	100.00E	13.82
11	22.41	3.50	20	42.69	4.42			
12	24.25	3.61	21	45.94	5.04			

In order to assess the score level reliability across facilities, we have completed an Intraclass Correlation Coefficient (ICC) using the split-half method, as suggested by the NQF PFCC committee staff. We used the updated data from the 39 LTAC facilities for 2002 - 2007. Each facility contained complete records meaning all items at admission and discharge for each patient. Each facility was randomly split into two datasets, and the averages at the facility level for each measure were calculated. We then compared across the facilities to get the ICC. A two-way random effects model was used to estimate the ICC for each measure. Using the definitions in McGraw, KO and Wong, SP (1996), "Forming Inferences About Some Intraclass Correlation Coefficients," *Psychological Methods*, 1(1): 30-46, a split-half ICC based on average measurements and using an agreement definition of the correlation. Thus, total score variance is the denominator of the measure. Thus, the high ICC value shown below shows that there is a high degree of absolute agreement in McGraw and Wong's terms – see their Table 5).

The ICC was 0.938, p <.001. This high ICC demonstrates that there is very high consistency for the mobility measure.

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?*)

As mentioned before, the reliability of the FIM® instrument is well known. The results of the analysis for the measure proposed show the reliability holds even when looking at a subset of FIM® items. In addition, we also show the means of the measure by facility below. The sizable range of these (e.g., 8.8-25.6 for mobility, 11.1-20.9 for self-care and 5.6-14.6 for the motor measure) show both that (1) this high ICC is not due to a restricted range and (2) that there are important differences across facilities that can be reliably determined by these measures. Thus, distinctions among the facilities can be seen and the measure is reliable.

The updated analyses support our previous interpretation that our measure maintains high reliability.

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. For each level 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)*

Since the validity of the 18-item FIM® instrument has been well established, we examined the concurrent validity of the mobility measure with the total FIM® score, both at admission and discharge. In particular, we used the total FIM score from all 18 items as our gold standard measure in which to test our new mobility measure against. The two tests of validity we used were the Pearson correlation coefficient and linear regression to calculate an r-squared which represents the percent of variance of the dependent variable (FIM® total) explained by the independent variable (mobility items). In this instance we examined the admission and discharge values separately.

We assessed the predictive validity of the mobility measure to determine if the measure predicts outcomes such as: functional change (total functional gain as assessed with the 18 item FIM® instrument (the gold standard)), and likelihood of discharge to the community setting. Linear regression was used to determine functional change, whereas the change in mobility was the independent variable, the r-squared value (proportion of change accounted for) and the Pearson correlation coefficient was examined. For discharge disposition, logistic regression was used, admission mobility total was the independent variable and the dependent variable was dichotomized as discharge to the community (yes or no). We used the C-statistic derived from the area under the ROC curve to determine the discrimination of the model, or the ability of the model to discriminate between those patients having the outcome of interest or not, as predicted by our measure. In SPSS this is completed by utilizing the patient level probabilities created during the logistic regression in the ROC curve analysis. The C-statistic ranges from 0.5 (no predictive ability) to 1.0 (perfect discrimination).

We completed all testing for the total data set including all venues, and separately by venue of post-acute care. For all analyses, the Rasch derived values for the mobility measure was used. SPSS version 21 was used in the analyses.

For the updated analysis we once again used the Pearson correlation coefficient and linear regression to calculate an r-squared which represents the percent of variance of the dependent variable (FIM® total) explained by the independent variable (mobility items).

As before, we assessed the predictive validity of the motor measure to determine if the measure predicts outcomes such as: functional change (total functional gain as assessed with the 18 item FIM® instrument (the gold standard)), and likelihood of discharge to the community setting Linear regression was used to determine functional change, whereas the change in self-care was the independent variable, the r-squared value (proportion of change accounted for) and the Pearson correlation coefficient was examined. For discharge disposition, logistic regression was used, admission motor total was the independent variable and the dependent variable was dichotomized as discharge to the community (yes or no). We used the C-statistic derived from the area under the ROC curve to determine the discrimination of the model, or the ability of the model to discriminate between those patients s having the outcome of interest or not, as predicted by our measure. In SPSS this is completed by utilizing the patient level probabilities created during the logistic regression in the ROC curve analysis. The C-statistic ranges from 0.5 (no predictive ability) to 1.0 (perfect discrimination).

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Concurrent Validity

Correlations: For all venues, our measure at both admission and discharge was correlated with the FIM® total, 0.671 ($p < 0.001$) and 0.768 ($p < 0.001$), respectively. The correlations remained significant within each venue of care; IRFs, 0.605 ($p < 0.001$) and 0.847 ($p < 0.001$); LTACs, 0.711 ($p < 0.001$) and 0.764 ($p < 0.001$); SNFs, 0.659 ($p < 0.001$) and 0.787 ($p < 0.001$).
Linear Regression: For all venues, when comparing our measure at admission and discharge to the respective FIM® totals, the r-square values ranged from respectable for admission FIM® total, to high for discharge FIM® total, 0.512 and 0.706, respectively. The values remained similar at the venue specific level as well; IRFs, 0.400 and 0.676; LTACs, 0.540 and 0.707; SNFs, 0.454 and 0.707.

Updated Analysis:

Concurrent Validity:

Correlation: Our measure at both admission and discharge was highly correlated with the FIM® total, 0.761 ($p < .001$), and 0.847 ($p < .001$), respectively.

Linear Regression: : For all venues, when comparing our measure at admission and discharge to the respective FIM® totals, the r-square values were extremely high 0.936 ($p < .001$), and 0.951 ($p < .001$), respectively.

Predictive Validity

Functional Gain: For all venues, when comparing gain in our measure to overall FIM® gain including all items, the correlation was acceptable, 0.615 ($p < 0.001$). In addition, by venue, the correlations remained acceptable; IRFs, 0.598 ($p < 0.001$); LTACs, 0.665 ($p < 0.001$); SNFs, 0.611 ($p < 0.001$). The linear regression showed acceptable r-squared values as well; all venues, 0.506; IRFs, 0.438; LTACs, 0.559; SNFs, 0.486.
Discharge Disposition – Community: For all venues, the logistic regression analysis shows that the gain in our measure has good predictive ability for discharge setting (community), with a C-statistic of 0.79. By venue, the results are similar; IRFs, 0.78; LTACs, 0.77; SNFs, 0.77.

Updated Analysis

Predictive Validity:

Functional Gain: When comparing gain in our measure to overall FIM® gain including all items, the correlation was high, 0.867 ($p < 0.001$).

Discharge Disposition - Community: The logistic regression analysis shows that the gain in our measure has good predictive ability for discharge setting (community), with a C-statistic of 0.783 (95% CI .780 - .787), $p < .001$.

The final additional analysis completed (upon request), is the overall range in the mobility measure averages at the facility level , smallest to largest average. The data are shown below.

Facility 38	8.81
Facility 24	8.95
Facility 6	9.13
Facility 28	11.12
Facility 31	11.22
Facility 36	12.89
Facility 5	13.24
Facility 39	13.39
Facility 4	13.91
Facility 30	14.33
Facility 29	15.03
Facility 14	15.81
Facility 7	15.94
Facility 12	16.36
Facility 27	16.51
Facility 23	16.99
Facility 9	17.2
Facility 16	17.45
Facility 10	17.95
Facility 11	18
Facility 34	18.4
Facility 15	18.67
Facility 17	18.8
Facility 25	18.84
Facility 8	19.37
Facility 37	19.43
Facility 33	19.81
Facility 1	19.83
Facility 2	19.98
Facility 3	20.2

Facility 26	20.35
Facility 19	20.48
Facility 22	20.59
Facility 20	21.22
Facility 32	21.96
Facility 21	23.47
Facility 35	24.71
Facility 18	25.13
Facility 13	25.6

2b2.4. 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 results show good validity across all analyses. The r-square values were all consistent around 0.5 – 0.6, meaning that the percent of variance explained in the dependent variables by our measure were all more than 50%. Considering we are testing the correlation between 4 items of an 18 item scale, these r-squared values are quite good. In addition, the predictive validity was also high.

The updated analyses support our previous interpretation that our measure maintains high validity.

_____

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. 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*)

We excluded patients that had expired in the post acute care setting (an unanticipated outcome) and patient less than age 18, both criteria consistent with published literature examining rehabilitation outcomes.

2b3.2. 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*)

2b3.3. 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*)

_____

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?
☐ **No risk adjustment or stratification**
☒ **Statistical risk model with 1 risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use 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.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient 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 and not related to disparities*)

We used Case Mix Group as our only adjustment variable through an indirect standardization method.

To calculate the facility's adjusted expected change in Rasch derived values, we use indirect standardization which weights national CMG-specific values by facility-specific CMG proportions. CMG-adjustment derives the expected value based on the case mix and severity mix of each facility. The case mix group classification system groups similarly impaired patients based on functional status at admission or patient severity. This is used for SNFs and IRFs, and the same procedure will be applied to the LTACs. Patients within the same CMG are expected to have similar resource utilization needs and similar outcomes. There are three steps to classifying a patient into a CMG at admission:

1. Identify the patient's impairment group code (IGC).
2. Calculate the patient's weighted motor index score, calculated from 12 of the 13 motor FIM® items.
3. Calculate the cognitive FIM® rating and the age at admission. (This step is not required for all CMGs.)

See file uploaded in S.15 for calculations.

2b4.4. What were the statistical results of the analyses used to select risk factors?

No statistical tests were calculated, CMG adjustment is a standard procedure.

2b4.5. 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 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. 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*)

***2b4.11. 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*)

_____

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE
2b5.1. 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)*

2b5.2. 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*)

2b5.3. 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?*)

_____

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

Note: *This criterion is directed 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).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different datasources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

2b6.2. 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*)

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of 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*)

References

1. Dodds TA, Martin DP, Stolov WC, Deyo RA. A validation of the functional independence measurement and its performance among rehabilitation inpatients. *Archives of physical medicine and rehabilitation.* May 1993;74(5):531-536.

2. Gerrard P, Goldstein R, Divita MA, et al. Validity and Reliability of the FIM(R) Instrument in the Inpatient Burn Rehabilitation Population. *Archives of physical medicine and rehabilitation.* Mar 5 2013.

3. Granger CV, Deutsch A, Russell C, Black T, Ottenbacher KJ. Modifications of the FIM instrument under the inpatient rehabilitation facility prospective payment system. *American journal of physical medicine & rehabilitation / Association of Academic Physiatrists.* Nov 2007;86(11):883-892.

4. Hall KM, Cohen ME, Wright J, Call M, Werner P. Characteristics of the Functional Independence Measure in traumatic spinal cord injury. *Archives of physical medicine and rehabilitation.* Nov 1999;80(11):1471-1476.

5. Keith RA, Granger CV, Hamilton BB, Sherwin FS. The functional independence measure: a new tool for rehabilitation. *Adv Clin Rehabil.* 1987;1:6-18.

6. Ottenbacher KJ, Hsu Y, Granger CV, Fiedler RC. The reliability of the functional independence measure: a quantitative review. *Archives of physical medicine and rehabilitation.* Dec 1996;77(12):1226-1232.

7. Stineman MG, Shea JA, Jette A, et al. The Functional Independence Measure: tests of scaling assumptions, structure, and reliability across 20 diverse impairment categories. *Archives of physical medicine and rehabilitation.* Nov 1996;77(11):1101-1108.

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 #: 2967
Measure Title: CAHPS® Home- and Community-Based Services Measures
Measure Steward: Centers for Medicare and Medicaid Services
Brief Description of Measure: CAHPS Home- and Community-Based Services measures derive from a cross disability survey to elicit feedback from adult Medicaid beneficiaries receiving home and community based services (HCBS) about the quality of the long-term services and supports they receive in the community and delivered to them under the auspices of a state Medicaid HCBS program. The unit of analysis is the Medicaid HCBS program, and the accountable entity is the operating entity responsible for managing and overseeing a specific HCBS program within a given state. (For additional information on the accountable entity, see Measures Testing form item #1.5 below.)

The measures consist of seven scale measures, 6 global rating and recommendation measures, and 6 individual measures:

Scale Measures
1. Staff are reliable and helpful –top-box score composed of 6 survey items
2. Staff listen and communicate well –top-box score composed of 11 survey items
3. Case manager is helpful - top-box score composed of 3 survey items
4. Choosing the services that matter to you - top-box score composed of 2 survey items
5. Transportation to medical appointments - top-box score composed of 3 survey items
6. Personal safety and respect - top-box score composed of 3 survey items
7. Planning your time and activities top-box score composed of 6 survey items

Global Ratings Measures
8. Global rating of personal assistance and behavioral health staff- top-box score on a 0-10 scale
9. Global rating of homemaker- top-box score on a 0-10 scale
10. Global rating of case manager- top-box score on a 0-10 scale

Recommendations Measures
11. Would recommend personal assistance/behavioral health staff to family and friends – top-box score on a 1-4 scale (Definitely no, Probably no, Probably yes, Definitely yes)
12. Would recommend homemaker to family and friends –– top-box score on a 1-4 scale (Definitely no, Probably no, Probably yes, Definitely yes)
13. Would recommend case manager to family and friends– top-box score on a 1-4 scale (Definitely no, Probably no, Probably yes, Definitely yes)

Unmet Needs Measures
14. Unmet need in dressing/bathing due to lack of help–top-box score on a Yes, No scale
15. Unmet need in meal preparation/eating due to lack of help– top-box score on a Yes, No scale
16. Unmet need in medication administration due to lack of help– top-box score on a Yes, No scale
17. Unmet need in toileting due to lack of help– top-box score on a Yes, No scale

18. Unmet need with household tasks due to lack of help– top-box score on a Yes, No scale

Physical Safety Measure
19. Hit or hurt by staff – top-box score on a Yes, No scale

Developer Rationale:
All measures were developed based on formative research to reflect the constructs most salient to the beneficiaries and other HCBS stakeholders. The research team ran a confirmatory factor analysis to test the hypothesized factor structure for the scale measures. The model was an acceptable fit (RMSEA=0.044, CFI=0.954, TLI=0.950).

Scale Measures
Staff are Reliable and Helpful. Assessing the performance of Medicaid direct care providers (i.e., personal assistants, behavioral health staff, homemakers) from the perspective of the beneficiary is important in evaluating the quality of services they render. This measure is based on beneficiary assessment of direct care staff reliability (showing up on time, stay as long as supposed to, communicate absences) and sensitivity to their privacy needs during the provision of personal care.

Staff Listen and Communicate Well. This measure is based on beneficiary assessment of direct care staff's communication skills and responsiveness to the person's needs. Specifically communication in a way that is understood by the beneficiary, respectful, and staff who listen carefully to what the beneficiary needs/wants and who, therefore, understand what the beneficiary needs. This is essential to the delivery of person-centered care and support. Person-centered care and support is required in Medicaid HCBS programs (Federal Register: https://federalregister.gov/a/2014-00487).

Case Manager Is Helpful. In HCBS programs, the case manager is responsible for monitoring the beneficiary's receipt of services and supports to ensure the service plan is being implemented as specified and that the person's needs are being adequately met. In order to meet these requirements, the case manager must be available to the beneficiary when s/he contacts him/her, and responsive to their changing/emerging needs. This measure is based on the beneficiary's assessment of case manager accessibility and responsiveness.

Choosing Services That Matter to You. A basic tenet of Medicaid HCBS services is that the beneficiary is involved in choosing their services/supports so that the service plan is truly person-centered, and that direct care staff implement the service plan in a person-centered manner. This measure is based on the beneficiary's assessment of the extent to which their service plan and direct care workers are person-centered.

Transportation to Medical Appointments. The health and welfare of beneficiaries must be ensured in the delivery of Medicaid HCBS (42 CFR §441: 302). Integral to assuring the health of beneficiaries is getting to medical appointments. This composite is based on the beneficiary's assessment of the extent to which they have transportation to medical appointments, whether the transportation provider is reliable, and whether the transportation is sufficiently accessible.

Planning your time and activities. Medicaid home and community-based services and supports should facilitate outcomes that are consistent with allowing beneficiaries to live the lives they choose – both in terms of daily routine as well as socializing with family and friends, and engaging in community activities. This measure is based on the beneficiary's assessment of the extent to which they have choice and control over these aspects of their lives.

Personal Safety and Respect. Beneficiaries of Medicaid HCBS should be assured that HCBS providers treat them with respect, that they will not be financially exploited by providers coming into their homes, and that they have someone to go to if they are treated badly. This measure will help HCBS programs assess this aspect of program quality. This measure has very high scores and thus very low variance so there is not much of a performance gap and reliability estimates are low. However, these activities should never occur and are critical to assess.

Individual Item Measures
Global Ratings of Staff (i.e., Personal Assistance/Behavioral Health Staff, Homemaker, Case Manager) – separate measures per staff type. In concert with more specific measures and scale measures, global ratings provide additional information for assessing program quality and can be used as a metric in evaluating quality improvement.

Would Recommend Staff (i.e., Personal Assistance/Behavioral Health Staff, Homemaker, Case Manager) to Family and Friends –separate measures per staff type. Beneficiaries' recommendation are yet another aspect of global experience with a program, and can be used for evaluating program quality and in quality improvement initiatives. While they are measuring similar topics as the global ratings items, these items measure a slightly different aspect of the care experience. The correlations between the related measures ranged from 0.44-0.52 suggesting that while related, they are measuring slightly different constructs. These measures are frequently requested by CAHPS survey end users for quality improvement initiatives.

Individual Unmet Need Measures:
- Unmet Need in Dressing/bathing Due to Lack of Help
- Unmet Need in Meal Preparation/Eating Due to Lack of Help
- Unmet Need in Medication Administration Due to Lack of Help
- Unmet Need in Toileting Due to Lack of Help
- Unmet Need with Household Tasks Due to Lack of Help

None of the Unmet Need items were captured in a scale measure because they did not correlate with each other in factor analysis. But the advisory panel for the measures development strongly recommended all unmet need stand-alone items be treated as individual measures as the evaluation of unmet need in HCBS is critically important for determining program quality. One of the most basic reasons for the existence of HCBS programs is to meet activities of daily living needs (bathing, dressing, toileting, medication administration) and instrumental activities of daily living (meal preparation/eating, cleaning/laundry) needs that, if not met, both jeopardize beneficiary health and make successful community living untenable. That is, having unmet needs related to these activities places individuals at risk of institutionalization rather than remaining at home and in their communities. Therefore, by definition, the need for assistance among HCBS populations will be high but, when HCBS programs are effective, the *unmet need* experienced by beneficiaries will be low. These measures are intended for use in assessing program quality and for quality improvement initiatives.

Hit or Hurt by Staff. This item was not retained in the Personal Safety and Respect scale measure due to low variation within responses and is thus presented as an individual item measure. This measure has very high scores and thus very low variance so there is not much of a performance gap and reliability estimates are low. However, the advisory panel for the measures development felt this measure is important for establishing the personal safety of program beneficiaries, as physical abuse by staff is a "never event" that should be tracked in any HCBS quality management system.

Numerator Statement: The CAHPS Home- and Community-Based Services measures are created using top-box scoring. This refers to the percentage of respondents that give the most positive response. Details regarding the definition of the most positive response are noted below. HCBS service experience is measured in the following areas. Attached Excel Table S.2b includes the specific item wording for each measure and the response options that go into the numerator.

Scale Measures
1. Staff are reliable and helpful – average proportion of respondents that gave the most positive response on 6 survey items
2. Staff listen and communicate well – average proportion of respondents that gave the most positive response on 11 survey items
3. Case manager is helpful - average proportion of respondents that gave the most positive response on 3 survey items

4. Choosing the services that matter to you - average proportion of respondents that gave the most positive response on 2 survey items
5. Transportation to medical appointments - average proportion of respondents that gave the most positive response on 3 survey items
6. Personal safety and respect - average proportion of respondents that gave the most positive response on 3 survey items
7. Planning your time and activities - average proportion of respondents that gave the most positive response on 6 survey items

Global Rating Measures
8. Global rating of personal assistance and behavioral health staff- average proportion of respondents that gave the most positive response of 9 or 10 on a 0-10 scale
9. Global rating of homemaker- average proportion of respondents that gave the most positive response of 9 or 10 on a 0-10 scale
10. Global rating of case manager- average proportion of respondents that gave the most positive response of 9 or 10 on a 0-10 scale

Recommendation Measures
11. Would recommend personal assistance/behavioral health staff to family and friends – average proportion of respondents that gave the most positive response of "Definitely Yes" on a 1-4 scale (Definitely no, Probably no, Probably yes, Definitely yes)
12. Would recommend homemaker to family and friends –– average proportion of respondents that gave the most positive response of "Definitely Yes" on a 1-4 scale (Definitely no, Probably no, Probably yes, Definitely yes)
13. Would recommend case manager to family and friends– average proportion of respondents that gave the most positive response of "Definitely Yes" on a 1-4 scale (Definitely no, Probably no, Probably yes, Definitely yes)

Unmet Needs Measures
14. Unmet need in dressing/bathing due to lack of help–average proportion of respondents that gave the most positive response of "No" on a 1-2 scale (Yes, No)
15. Unmet need in meal preparation/eating due to lack of help–average proportion of respondents that gave the most positive response of "No" on a 1-2 scale (Yes, No)
16. Unmet need in medication administration due to lack of help–average proportion of respondents that gave the most positive response of "No" on a 1-2 scale (Yes, No)
17. Unmet need in toileting due to lack of help–average proportion of respondents that gave the most positive response of "No" on a 1-2 scale (Yes, No)
18. Unmet need with household tasks due to lack of help–average proportion of respondents that gave the most positive response of "No" on a 1-2 scale (Yes, No)

Physical Safety Measure
19. Hit or hurt by staff –average proportion of respondents that gave the most positive response of "No" on a 1-2 scale (Yes, No)

Denominator Statement: The denominator for all measures is the number of survey respondents. Individuals eligible for the CAHPS Home- and Community-Based Services survey include Medicaid beneficiaries who are at least 18 years of age in the sample period, and have received HCBS services for 3 months or longer and their proxies.[1] Eligibility is further determined using three cognitive screening items, administered during the interview:

[1] According to guidance produced under the CMS TEFT Technical Assistance contract, individuals who are more likely to be good proxy respondents during the CAHPS Home- and Community-Based Services survey data collection are: (a) those who are willing to respond on behalf of the beneficiary; (b) unpaid caregivers, family members, friends, and neighbors; and (c) those who know the beneficiary well enough that s/he is familiar with the services/supports they are receiving, and has regular, ongoing contact with

Q1. Does someone come into your home to help you? (Yes, No)
Q2. How do they help you?
Q3. What do you call them?

Individuals who are unable to answer these cognitive screening items are excluded. Some measures also have topic-specific screening items as well. Additional detail is provided in S.9.

Denominator Exclusions: Individuals less than 18 years of age and individuals that have not received HCBS services for at least 3 months should be excluded. During survey administration, additional exclusions include individuals that failed any of the cognitive screening items mentioned in the denominator statement below. There were 227 beneficiaries excluded due to not passing the cognitive screener (53 Aged/Disabled, 59 ID/DD, 25 TBI, and 90 SMI). Allowing proxy respondents in future administrations has the potential to further reduce these numbers.

Measure Type: PRO
Data Source: Patient Reported Data/Survey
Level of Analysis: HCBS Program

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

New Measure Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if th relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

This submission contains information for 19 Patient Reported Outcome Performance Measures (PRO-PMs) derived from the Home and Community Based Services (HCBS) Experience of Care (EoC) survey. The measures consist of seven scale measures, 6 global rating and recommendation measures, and 6 individual measures:

Scale Measures
1. Staff are reliable and helpful
2. Staff listen and communicate well
3. Case manager is helpful
4. Choosing the services that matter to you
5. Transportation to medical appointments
6. Personal safety and respect
7. Planning your time and activities

Global Ratings Measures
8. Global rating of personal assistance and behavioral health staff
9. Global rating of homemaker
10. Global rating of case manager

them. Examples of circumstances that increase the likelihood that someone has knowledge about the beneficiary and their care situation include living with the beneficiary, managing the beneficiary's in-home care for a majority of the day, having regular conversations with the beneficiary about the services they receive, in-person visits with the beneficiary, and being present when services/supports are delivered. Individuals who are less likely to be good proxy respondents are (a) those with paid responsibilities for providing services/supports to the beneficiary, including family members and friends who are paid to help the beneficiary and (b) guardians or conservators whose only responsibility is to oversee the beneficiary's finances.

Recommendations Measures
11. Would recommend personal assistance/behavioral health staff to family and friends
12. Would recommend homemaker to family and friends
13. Would recommend case manager to family and friends

Unmet Needs Measures
14. Unmet need in dressing/bathing due to lack of help
15. Unmet need in meal preparation/eating due to lack of help
16. Unmet need in medication administration due to lack of help
17. Unmet need in toileting due to lack of help
18. Unmet need with household tasks due to lack of help

Physical Safety Measure
19. Hit or hurt by staff

Summary of evidence:

- The developer provides a diagram that illustrates the path to potential beneficiary outcomes starting with the key processes (i.e., person-centered assessment and service planning) and resulting services (i.e., HCBS services and supports) that are expected to influence the beneficiary assessment of services/supports as well as beneficiary outcome Although not stated explicitly, these activities likely also would affect overall ratings of the care provided and willingness to recommend the HCBS services and supports.
- To assess if the target population values the measured PROs and find them useful, the developer utilized input from the HCBS beneficiary audience as well as stakeholders in the broader HCBS community. They state that the audiences have consistently supported the proposed measures as necessary and important.
- This input included focus groups and interviews, public comment via the Federal Register, and a Federal Advisory Panel.

Guidance from the Evidence Algorithm
Pro-based measure (Box 1) →Relationship between the outcome and at least one healthcare action is identified and supported by the rationale (Box 2) → PASS

Question for the Committee:
- Is there at least one thing that the provider can do to achieve a change in the measure results?
- Does the Committee agree that HCBS patients value queries about the various domains included in the HCBS Experience of Care survey?

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

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.

- See tables 1b.2a and 1b.2b in attached tables.
  Performance data were calculated through the testing of the measure and were provided at both the summary score (measure) and item (question) level. The data provided was collected from March – October, 2013 and consist of data from 26 Medicaid HCBS programs across 10 states. Performance data on the individual items used for the various measures are included in the supplementary materials.

Disparities
- The developer indicates the measures in the submission focus on people who are elderly with disabilities, individual with physical disabilities, persons with intellectual/developmental disability, individuals with brain injury, and those with serious mental illness. who receive Medicaid-funded home and community-based services. As such, the target population mirrors those in a typical Medicaid population with evidence of disparities due to lower income, race an ethnicity.
- Tables 1b.4a, 1b.4b, 1b.4c, and 1b.4d provide summary statistics for the measure groupings for these populations.

Questions for the Committee:

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

Preliminary rating for opportunity for improvement:

Scale Measures

1. Staff are reliable and helpful ☐ **High** ☐ **Moderate** X **Low** ☐ **Insufficient**
2. Staff listen and communicate well ☐ **High** ☐ **Moderate** X **Low** ☐ **Insufficient**
3. Case manager is helpful ☐ **High** X **Moderate** ☐ **Low** ☐ **Insufficient**
4. Choosing the services that matter to you ☐ **High** X **Moderate** ☐ **Low** ☐ **Insufficient**
5. Transportation to medical appointments ☐ **High** X **Moderate** ☐ **Low** ☐ **Insufficient**
6. Personal safety and respect ☐ **High** ☐ **Moderate** X **Low** ☐ **Insufficient**
7. Planning your time and activities ☐ **High** X **Moderate** ☐ **Low** ☐ **Insufficient**

Global Ratings Measures

8. Global rating of personal assistance and behavioral health staff ☐ **High** X **Moderate** ☐ **Low** ☐ **Insufficient**
9. Global rating of homemaker ☐ **High** X **Moderate** ☐ **Low** ☐ **Insufficient**
10. Global rating of case manager ☐ **High** X **Moderate** ☐ **Low** ☐ **Insufficient**

Recommendations Measures

11. Would recommend personal assistance/behavioral health staff to family and friends ☐ **High** X **Moderate** ☐ **Low** ☐ **Insufficient**
12. Would recommend homemaker to family and friends ☐ **High** X **Moderate** ☐ **Low** ☐ **Insufficient**
13. Would recommend case manager to family and friends ☐ **High** X **Moderate** ☐ **Low** ☐ **Insufficient**

Unmet Needs Measures

14. Unmet need in dressing/bathing due to lack of help X **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**
15. Unmet need in meal preparation/eating due to lack of help X **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**
16. Unmet need in medication administration due to lack of help X **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**
17. Unmet need in toileting due to lack of help ☐ **High** ☐ **Moderate** X **Low** ☐ **Insufficient**
18. Unmet need with household tasks due to lack of help X **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Physical Safety Measure

19. Hit or hurt by staff ☐ **High** ☐ **Moderate** X **Low** ☐ **Insufficient**

Committee pre-evaluation comments

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

1a. Evidence to Support Measure Focus

Comments:

**I would rate as moderately important. If I understand the specifications correctly, this measure is intended to assess an HCB "program", which I believe the developer means to be an entity, probably a company like the VNA, providing HCB services to the state's Medicaid population. The assessment of the quality of care being provided from the perspective of the patient could be valuable to the state in monitoring the quality of the service.

**Measurement at the global, scale and individual level to demonstrate qualitative and quantitative evidence related directly and tangentially to HCBS. Many of these quality and safety measures are required as part of a HCBS under Medicaid regulations-For example, transportation falls under federal regulation for Medicaid beneficiaries as transport to medical apts is part of their covered benefit. Measuring the quality if the service and availability of service is a tangential outcome of access and utilization of the service.

For Scale Measure: Is it possible to have a metric that shows patient activation and/or engagement in HCBS? For example, how empowered to patients feel that they can participate in and/or codesign their home care plan

Scale measure: Is there a way to add cultural competency and/or language? Ease of interpretation on interpreter services

1b. Performance Gap

Comments:

**Here I would rate the measure as barely moderate. Except for the unmet needs measures the performance gap is pretty narrow. I have some concerns about the appropriateness of the unmet needs measure since, generally, the volume and kind of HCB services are usually dictated by a plan of care that is determined by the state, not the program. The program probably cannot increase either without state approval. We should explore this issue with the developer when we discuss this component of 2967.

**Yes. Disparities due to lower income, race and ethnicity (the target population focused on Medicaid, people who are elderly with disabilities and individuals with physical disabilite4s and server mental illness)

Criteria 2: Scientific Acceptability of Measure Properties
2a. Reliability
2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Self-reports of Medicaid beneficiaries of home and community based services

Specifications:

- The measures is specified for the program level of analysis for home and community based services; higher scores are an indicator of better quality
- The measures that comprise this submission include scale measures (7), global ratings (3), recommendation ratings (3), unmet needs (5), and a physical safety measure (1). The attached spreadsheet contains the individual survey items and item mapping for each measure grouping
- The frequency of data collection/aggregation is at the discretion of state users. The developer notes that CMS has determined the survey from which the measures are derived will be conducted on a voluntary basis by states. It is anticipated that states would field the survey no more frequently than annually per HCBS program.
- The denominator is Medicaid beneficiaries who are at least 18 years of age in the sample period, and have received HCBS services for 3 months or longer.
- Eligibility is further determined using three cognitive screening items, administered during the interview (Individuals who are unable to answer these cognitive screening items are excluded):
  Q1. Does someone come into your home to help you? (Yes, No)
  Q2. How do they help you?
  Q3. What do you call them?
- The proposed provider-related measures in this submission focus on the most common provider types for adults receiving Medicaid HCBS. These include personal assistance providers, behavioral health staff, homemakers and case managers.
- Case-mix adjustment is done via regression methodology or a covariance adjustment. Case-mix adjustment is used to adjust scores for various patient and survey mode characteristics.
- Scoring specifications for the measures follow the same general scoring approach as used by other CAHPS surveys that use the CAHPS analysis program. The measures are based on case-mix adjusted
- Sampling should be stratified by HCBS program within each state, in order to allow comparisons of measure results for each HCBS program to the state mean. The source of the sample frame is the state Medicaid agency or an entity delegated by the state Medicaid agency (e.g., state agency other than the Medicaid agency that operates the program, a MCO, a case management agency, state county, etc.).

- Results suggest that the effective sample size should be 400 people per stratum (with smaller programs including the census).
- Due to the impairments (i.e., cognitive, hearing) prevalent among individuals served by HCBS programs, stakeholders recommend that the survey be conducted through in-person interviews. Based on field test results, administering the survey by phone was found appropriate if a statistical adjustment for survey mode is made for mixed-mode administrations.

Questions for the Committee :

- *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?*

2a2. Reliability Testing Testing attachment

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.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☐ **Data element** ☒ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing

- The developers conducted a pilot test and a field test of the survey with 26 Medicaid HCBS programs across ten states. The 10 states were geographically dispersed and included AZ, CO, CT, GA, KY, LA, MD, MN, NH, and TN; these states (with the exception of TN) were CMS Testing Experience and Functional Tools (TEFT) Demonstration grantees
- There were 2,336 completed HCBS EoC surveys from 26 Medicaid HCBS programs included in the analysis of the survey data. The testing was conducted from October 2013 – March 2015
- Reference Exhibit 1. States, Populations, Programs, Authorities, and Total Returned Surveys
- Data element reliability was assessed using Cronbach's Alpha values which assess internal consistency of the survey items used in the scale measures.
- HCBS program-level reliability was assessed by determining inter-unit reliability (IUR). Unit-level reliability indicates the extent to which the experiences of respondents within a unit (e.g., HCBS program) correlate with one another compared to the amount that reported experiences differ among units. The developers indicate that one of the primary purposes of these measures is to be able to detect difference among HCBS programs, and thus, this ratio is a good indicator of the extent to which the scale measures and other survey items accomplish this goal.

Results of reliability testing

- Tab 1.b.2a in the supplementary tables file for item-level IUR statistics for survey items used in the scale measures

HCBS Inter-unit reliability (IUR) Statistics

- For Cronbach's alpha, 0.70 or higher is a widely-accepted rule of thumb for a set of items to be considered a scale.
  - The Cronbach's Alpha scores range from 0.84 to 0.17, with three measures falling below the recommended 0.70 threshold. While these values are below the recommended threshold, the developer indicated these measures were all deemed critical by the technical expert panel for assessing the quality of a HCBS program.
- If the IUR is higher, the ability of the item or scale measure to discriminate across programs is greater. Scales with reliability coefficients above 0.70 provide adequate precision for use in statistical analysis of unit-level comparisons. As the IUR gets smaller, a larger sample is needed in order to reliably discriminate across programs.

- The IUR values at the program level for the scale measures, global measures and recommendation measures range from 0.77 to 0.32, with the majority of measures (10/13) falling below the 0.70 threshold. This indicates that these measures will need a larger sample size to effectively discriminate among programs.

The IUR values at the program level for the unmet needs and physical safety measures range from -0.28 - 0.63.

Guidance from the Reliability Algorithm

Precise specifications (Box 1) → Empirical testing conducted with measure as specified (Box 2) → Score-level testing conducted (Box 4) → Method of testing appropriate (Box 5) → Moderate certainty that the scores are reliable for 8 measures; lower certainty for 11 measures, although reliability will likely be higher if number of respondents is higher (than 200).

Questions for the Committee:

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

Preliminary rating for reliability:

Scale Measures

1. Staff are reliable and helpful ☐ **High** X **Moderate** ☐ **Low** ☐ **Insufficient**
2. Staff listen and communicate well ☐ **High** X **Moderate** ☐ **Low** ☐ **Insufficient**
3. Case manager is helpful ☐ **High** X **Moderate** ☐ **Low** ☐ **Insufficient**
4. Choosing the services that matter to you ☐ **High** X **Moderate** ☐ **Low** ☐ **Insufficient**
5. Transportation to medical appointments ☐ **High** X **Moderate** ☐ **Low** ☐ **Insufficient**
6. Personal safety and respect ☐ **High** ☐ **Moderate** X **Low** ☐ **Insufficient**
7. Planning your time and activities ☐ **High** ☐ **Moderate** X **Low** ☐ **Insufficient**

Global Ratings Measures

8. Global rating of personal assistance and behavioral health staff ☐ **High** X **Moderate** ☐ **Low** ☐ **Insufficient**
9. Global rating of homemaker ☐ **High** ☐ **Moderate** X **Low** ☐ **Insufficient**
10. Global rating of case manager ☐ **High** ☐ **Moderate** X **Low** ☐ **Insufficient**

Recommendations Measures

11. Would recommend personal assistance/behavioral health staff to family and friends ☐ **High** ☐ **Moderate** X **Low** ☐ **Insufficient**
12. Would recommend homemaker to family and friends ☐ **High** X **Moderate** ☐ **Low** ☐ **Insufficient**
13. Would recommend case manager to family and friends ☐ **High** ☐ **Moderate** X **Low** ☐ **Insufficient**

Unmet Needs Measures

14. Unmet need in dressing/bathing due to lack of help ☐ **High** ☐ **Moderate** X **Low** ☐ **Insufficient**
15. Unmet need in meal preparation/eating due to lack of help ☐ **High** ☐ **Moderate** X **Low** ☐ **Insufficient**
16. Unmet need in medication administration due to lack of help ☐ **High** X **Moderate** ☐ **Low** ☐ **Insufficient**
17. Unmet need in toileting due to lack of help ☐ **High** ☐ **Moderate** X **Low** ☐ **Insufficient**
18. Unmet need with household tasks due to lack of help ☐ **High** ☐ **Moderate** X **Low** ☐ **Insufficient**

Physical Safety Measure

19. Hit or hurt by staff ☐ **High** ☐ **Moderate** X **Low** ☐ **Insufficient**

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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.

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

☐ **Face validity only**

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

Validity testing method:

- Criterion validity refers to the extent to which the HCBS scale measures agree with some criterion of the "true" value of the measure, and can be predictive or concurrent. The developers estimated correlation coefficients between each global rating measure and each scale measure.
- The developers examined correlations among the scale measures to determine if they measure different constructs. As these are all measures of beneficiary experience with HCB services, the factors are expected to be related; however, all inter-scale measure correlations should be below 0.80 to indicate that these 7 factors, while related, do not overlap to the point of being redundant.

Validity testing results:

- If the scale measures have good concurrent validity, then they should have a moderate to strong correlation ($r > 0.30$) with a conceptually related global rating measure.

Correlation of Scale Measures and Related Global Rating Measures

- For most measures, the correlations between the scale measures and the related global rating measures were moderate, suggesting that the scale measures are valid measures of beneficiary experience with these providers. The correlation for Personal Safety and Respect was low; however, it should be noted that there was not much variance in the items for this measure.

Inter-Scale Correlations

- The scale measures were somewhat correlated with each other as they are all measures of beneficiary experience. However, no values were above 0.80, suggesting that these scales are measuring unique concepts.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

N/A – there are no "true" exclusions to these measures

Questions for the Committee:

- *Do you have any reasons/evidence to believe there should be exclusions to these measures?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included ? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☒ **Yes** ☐ **No**

Risk adjustment summary [Risk adjustment summary

- The developers tested the beneficiary characteristics of age, health status (both general health and emotional/mental health), gender, and whether the respondent lived alone as case-mix adjusters. These characteristics typically have the strongest and most consistent associations with patient-reported problems in other CAHPS surveys. In addition, they tested several survey design characteristics – survey mode.
- The research team used stepwise regression to select a subset of the potential case-mix adjusters for further analysis. Stepwise regression analyses evaluated the strength of the relationship of each potential adjuster to ten global rating and scale measures in separate models in which each measure was regressed on all of the potential adjusters.
- The research team then estimated the heterogeneity factor, predictive power, explanatory power, and impact factor for each potential case-mix variable selected in the regression models.
- Variables that had an impact factor >1.0, and were eligible to be considered as case- mix adjusters, included general health rating, mental health rating, age, gender, whether respondent lives alone, survey administration mode, and response option.

Questions for the Committee:

- *Is an appropriate risk-adjustment strategy included in the measure?*
- *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, describe the rationale provided.*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- The developer *used t-tests to compare the case-mix adjusted mean scores of each item, scale score, and global rating for each HCBS program within a state to the mean score of all programs combined within the state. A p-value of <0.05 was used to determine whether the scores were statistically significantly different from each other.*
- Exhibit 9 in the testing form shows counts of programs that were statistically significantly different above or below their state mean for each measure. The exhibit also reports the percentage of programs that were statistically significant in either direction from their state mean.
- The developer summarizes that the findings demonstrate that the measures produce results that adequately discriminate between service recipients' experience of care in their program compared to all programs within a state.

Question for the Committee:

Does this measure identify meaningful differences about quality?

2b6. Comparability of data sources/methods:

N/A

2b7. Missing Data

- The developers conducted a nonresponse bias analysis to evaluate whether respondents and nonrespondents differed significantly.

Exhibit 10. Sample Frame Demographic Characteristics

Characteristics	Nonrespondents n=13,940	Respondents n=1,624	Total (Nonrespondents and Respondents Combined) N=15,564
HCBS Population*			
Aged (65+)	34.0	31.0	33.7
Disabled (<65)	36.4	41.8	36.9
ID/DD	19.0	11.3	18.2
TBI	4.2	6.3	4.4
SMI	6.4	9.6	6.8
Primary Language			
English	97.1	97.7	97.2
Spanish	2.0	1.9	2.0
Other	0.9	0.4	0.8
Metropolitan Statistical Area*			
Yes	74.3	76.5	74.5
No	25.7	23.5	25.5
Gender			
Male	41.9	43.0	42.0
Female	58.2	57.0	58.0
Assigned Survey Response			
Alternate	50.1	49.0	49.9
Standard CAHPS	50.0	51.1	50.1
Assigned Survey Mode			
In-person	80.6	79.2	80.4
Phone	19.4	20.8	19.6
State†*			
AZ	9.4	11.4	9.6
CO	17.7	15.0	17.4
GA	14.1	16.2	14.3

MD	19.2	7.1	18.0
MN	14.5	23.7	15.4
NH	25.2	26.6	25.3
Guardian*			
Yes	10.3	4.0	9.7
No	89.7	96.0	90.4

*Nonrespondents and respondents significantly differ by this characteristics at p <0.05

Guidance from the Validity Algorithm

Specifications consistent with evidence (Box 1) →Threats to validity assessed (Box 2) → Empirical testing conducted for the measure as specified (Box 3) → Testing at the score-level conducted (Box 6) →High certainly that the scores are valid indicators of quality

Preliminary rating for validity:

Scale Measures

1. Staff are reliable and helpful ☐ High X Moderate ☐ Low ☐ Insufficient
2. Staff listen and communicate well ☐ High X Moderate ☐ Low ☐ Insufficient
3. Case manager is helpful ☐ High X Moderate ☐ Low ☐ Insufficient
4. Choosing the services that matter to you ☐ High X Moderate ☐ Low ☐ Insufficient
5. Transportation to medical appointments ☐ High X Moderate ☐ Low ☐ Insufficient
6. Personal safety and respect ☐ High X Moderate ☐ Low ☐ Insufficient
7. Planning your time and activities ☐ High X Moderate ☐ Low ☐ Insufficient

Global Ratings Measures

8. Global rating of personal assistance and behavioral health staff ☐ High X Moderate ☐ Low ☐ Insufficient
9. Global rating of homemaker ☐ High X Moderate ☐ Low ☐ Insufficient
10. Global rating of case manager ☐ High X Moderate ☐ Low ☐ Insufficient

Recommendations Measures

11. Would recommend personal assistance/behavioral health staff to family and friends ☐ High X Moderate ☐ Low ☐ Insufficient
12. Would recommend homemaker to family and friends ☐ High X Moderate ☐ Low ☐ Insufficient
13. Would recommend case manager to family and friends ☐ High X Moderate ☐ Low ☐ Insufficient

Unmet Needs Measures

14. Unmet need in dressing/bathing due to lack of help ☐ High X Moderate ☐ Low ☐ Insufficient
15. Unmet need in meal preparation/eating due to lack of help ☐ High X Moderate ☐ Low ☐ Insufficient
16. Unmet need in medication administration due to lack of help ☐ High X Moderate ☐ Low ☐ Insufficient
17. Unmet need in toileting due to lack of help ☐ High X Moderate ☐ Low ☐ Insufficient
18. Unmet need with household tasks due to lack of help ☐ High X Moderate ☐ Low ☐ Insufficient

Physical Safety Measure

19. Hit or hurt by staff ☐ High X Moderate ☐ Low ☐ Insufficient

Committee pre-evaluation comments

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

2a.1 and 2b.1 Specifications:

Comments:

**Specification around the cognitive screening questions- if this is done at the hospital (where people are assessed often for home service or in the post acute care setting) how might you control for false positives on the cognitive impairment screening? Knowing, this is a higher risk when people are in acute care settings??

For the denominator of Medicaid beneficiaries who are 18 or older and have had HCBS services for three month or longer, does this county resumption of care or is it aggregate three months in a certain amount of time?

Case-mix adjustment is done via regression methodology or a covariance adjustment.- How do we ensure quality of coding for risk adjusted revenue or is this out of scope of this measure?

Concern for implementation: how do we spread and scale, especially of the recommendations are for in person ? How easy will it be to regularly implement this survey? What about considerations of Medicaid churn?

2a2. Reliability Testing

Comments:

**Developer reports the survey was used for 26 different programs. Total respondents were 2336, an average of less than 100 per program. Is this sufficient to determine reliability?

We also need an explanation of the recommendation that results should be "stratified" in order to compare a program's score with the state mean. Stratified how: size of program? composition of caseload?

**Both measure score and data element 2336 completed surveys across 10 Medicaid SCHS service sites

Data element reliability was assessed using Cronbach's Alpha values. There needs to be a larger sample size to effectively discriminate among programs because the scale measures, global measures and recommendation measures range from .77-.32

2b2. Validity–Testing

Comments:

**For most measures, the correlations between the scale measures and the related global rating measures were moderate, suggesting that the scale measures are valid measures of beneficiary experience with these providers. The correlation for Personal Safety and Respect was low; however, it should be noted that there was not much variance in the items for this measure.

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.

- It is recommended that the HCBS EoC Survey be administered in-person or by phone. CATI or CAPI data collection is recommended which allow for the creation of electronic databases post data collection.
- The developers include notes on opportunities to improve survey data collection learned from the field-test and recommendations on sampling and seasonality timing.
- The final HCBS EoC survey will be available to state Medicaid Agencies for use free of charge. In addition to the survey instrument, users will have access to comprehensive materials supporting fielding, analysis, and reporting as well as CAHPS Analysis Program that performs analysis and significance testing.

Questions for the Committee:

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

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3 Feasibility

Comments:

**I am very doubtful that many states will want to use this measure. The data source is the responses from a 95 question survey which the developer recommends be administered in person, or possibly by phone. The suggested sample size is 400. Administering a survey of that length even by phone is expensive; it is unlikely that states will require the program to do and pay for (HCB programs are generally not well funded) nor that the state will be able to pay for. I believe the states in the test received federal grants to cover the cost.

**My concern is spread and scale- how do large systems do interviews in person?. I am also concerned with interview responses creating bias. What is the plan for Medicaid churn and re-surveying patients? What about training and oversight of contracted agencies of whom there may be little power or influence to improve performance? Would this be a way to vet these agencies?

Criterion 4: Usability and Use

4. Usability and 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.

Current uses of the measure

The measure is new and not currently in use, but public reporting and quality improvement uses are planned.

Publicly reported? ☐ **Yes** ☒ **No**

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

OR

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

Accountability program details The HCBS EOC survey is new, and so are the measures described in this submission. The survey is under review by the CAHPS Consortium for evaluation of use of the CAHPS trademark. Upon receipt, it is anticipated this survey and measures will be put into voluntary use by state programs for QI initiatives and service planning.

Improvement results New Measure

Unexpected findings (positive or negative) during implementation None identified

Potential harms None identified

Feedback:

Due to the newness of the measures in this submission and the recently completion of survey testing and analysis, the submission has not been viewed by other NQF bodies. However, both the MAP Duals Workgroup and the Home and Community Based Services Committee have been following the development and have expressed interest in the measurement set. They cite a paucity of measures for the HCBS care setting and the broader targeted populations that comprise this denominator.

***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**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4 Usability and Use

Comments:

**I believe the scores are not meant to be publicly reported. The developer talks of their use to "compare with state mean", which suggest intended uses are for QI improvement on the part of the program and for QI oversight by the state. From consumer perspective, it would desirable for the results to be public in order to guide consumer selection of HCB program (if there is a choice in her region).

We should have a separate discussion of the physical safety measure: this veers close to the "never event" category and public reporting is a sensitive issue. We might wish to recommend some cautionary language if we decide to recommend endorsement.

**Would the public reporting be on consumer report

Criterion 5: Related and Competing Measures

Related or competing measures

None

Harmonization

N/A

Pre-meeting public and member comments

Identifying person- and family-centered (PFCC) quality measures for home and community-based services (HCBS) is important, especially in developing accountability for the person-centered care requirements in the Centers for Medicare & Medicaid Services HCBS regulations. PFCC quality measures for HCBS are also becoming increasingly important as health care and long-term services and supports become integrated. The HCBS Experience of Care measures collect information from the perspective of the individual, and as such have a person-centered focus. After reviewing the survey questions to be included for the HCBS measure, The SCAN Foundation (Foundation) recommends adjusting or removing the following questions.

Staff listen and communicate well

Survey items 29 and 42 identified as part of the outcome measure for staff listening and communicating well is phrased, “How often are the explanations [personal assistance/behavioral health staff] or [homemaker] gives you hard to understand because of an accent or the way he or she speaks English?” While it is important to identify whether communication between the personal assistance/behavioral health staff/homemaker and the individual receiving services is clearly understood, the way this question is phrased does not effectively address cultural competencies and potential language barriers as it assumes the person receiving care is a native English speaker. The Foundation suggests reframing or removing survey items 29 and 42 to capture whether someone is generally able to understand the provider, spoken to in a language they understand, and can effectively communicate instructions, wishes, and concerns with staff. We acknowledge that survey item 31, “How often do [personal assistance/behavioral health staff] explain things in a way that is easy to understand?” may already addresses the communication concern effectively.

Physical safety measure

The Foundation applauds the inclusion of measures addressing physical safety. However, the proposed measure, “Do any staff that you have now hit you or hurt you?” included in isolation raises concerns. The survey question does not clearly identify new accounts of abuse as opposed to reports that have been addressed and does not appear to include follow up questions for to help with addressing any current concerns. If this measure is to be included, we recommend including additional questions to better understand the current situation in the event of an affirmative response and a clear protocol outlining how to the surveyor should respond to ensure the individual’s safety.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: CAHPS® Home- and Community-Based Services Measures
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: Click here to enter a date

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.
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

☐ Health outcome: Click here to name the health outcome

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

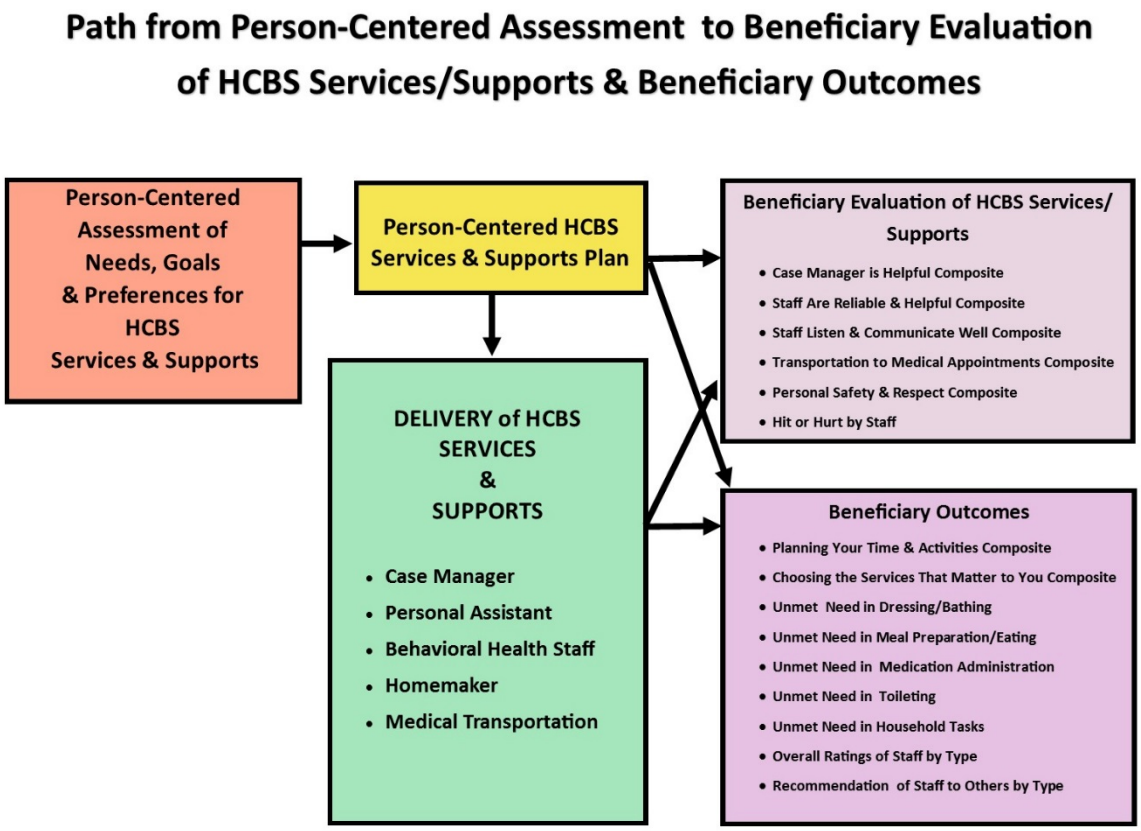
☐ Other: Click here to name what is being measured

_____

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

The following diagram illustrates the path to the beneficiary outcomes proposed in this submission, starting with the key processes (i.e., person-centered assessment and service planning) and resulting services (i.e., HCBS services and supports) that are expected to influence the beneficiary assessment of services/supports as well as beneficiary outcomes.



1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

The person-centered approach to beneficiary assessment (of need, goals and preferences), together with person-centered service planning, is expected to influence -- either directly or indirectly via service delivery -- both beneficiary evaluation of services/supports as well as beneficiary outcomes.

The person-centered approach that drives and shapes the beneficiary experience is a fundamental tenet of Medicaid HCBS programs. In 2014, CMS issued new regulations that require Medicaid HCBS programs to work with beneficiaries to develop a person-centered service plan that (a) has individually identified goals and preferences to assist the person in achieving personally-identified outcomes and (b) insures the delivery of services/support in a manner that reflect personal preferences and choice.[1, 2, 3]

The person-centered service planning process is expected to directly influence three composite outcome measures in the following ways:

- A primary case manager responsibility is working with the beneficiary to develop a services/supports plan which in turn will determine the services/supports that the beneficiary receives. Once the services/supports plan has been developed, the case manager also has responsibility for monitoring the plan's implementation to insure it meets the beneficiary's needs/preferences and supports the person in achieving their goals. Thus, it is expected that the case manager's role in both the service planning process and service monitoring will affect the beneficiary's evaluation of the case manager as captured in the composite measure "*Case Manager is Helpful."*
- The purpose of Medicaid HCBS programs is not merely to provide a service(s) but to support beneficiaries' ability to live as they want in the community. Thus, the person-centered planning process is intended to identify the assistance that the beneficiary requires to direct their own lives, as represented in the outcome measure *"Planning Your Time and Activities."*
- The service planning process is expected to directly affect the composite *"Choosing the Services That Matter To You"* because a fundamental principle of that process is to work with the beneficiary to identify the services of their choosing.

The person-centered service planning process is expected to indirectly affect beneficiary evaluation of services/supports as a result of whether HCBS providers deliver services and supports in accordance with the plan. These impacts are captured by nearly all beneficiary outcomes (except the composite *"Choosing Services That Matter To You"*).

The delivery of HCBS services/supports by providers is expected to directly impact both beneficiary evaluation of service provision as well as beneficiary outcomes. While there are many types of HCBS services and supports, beneficiary experience with those most commonly delivered to people in Medicaid HCBS programs is the focus of the beneficiary evaluation of service/support-related measures. These most common services and supports include:

- Personal Attendant and Behavioral Health Staff who provide assistance with personal care activities.
- Homemakers who assist beneficiaries in activities such as housekeeping, meal preparation and laundry.
- Case Managers who assess the beneficiary's need for services/supports; work with them to develop a service plan responsive to the person's needs, goals and person preferences; monitor service delivery; and assist the person in arranging more/different services as their needs and circumstances change.
- Medical Transportation which provides transportation to medical appointments.

The delivery of these HCBS services/supports is expected to mitigate beneficiary unmet needs as well as influence how beneficiaries assess their experience with the provision of services/supports. The delivery of services/supports in a person-centered manner and responsive to beneficiary preferences is also expected to impact the person's assessment of the degree to which they have control over planning their daily activities (as measured by the composite *"Planning Your Time and Activities"*).

References

Guidance to HHS Agencies for Implementing Principles of Section 2402(a) of the Affordable Care Act: Standards for Person-Centered Planning and Self-Direction in Home and Community-Based Services Programs: http://www.acl.gov/Programs/CIP/OCASD/docs/2402-a-Guidance.pdf

[2] 2016 Medicaid HCBS Rule in Federal Register: https://federalregister.gov/a/2014-00487

[3] CMS Fact Sheet on 2014 Medicaid HCBS Rule: https://www.medicaid.gov/medicaid-chip-program-information/by-topics/long-term-services-and-supports/home-and-community-based-services/downloads/final-rule-fact-sheet.pdf

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

_____

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

_____

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

_____

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

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 subcriterion 1b).

Brief Measure Information

NQF #: 2967
De.2. **Measure Title:** CAHPS® Home- and Community-Based Services Measures
Co.1.1. **Measure Steward:** Centers for Medicare and Medicaid Services
De.3. **Brief Description of Measure:** CAHPS Home- and Community-Based Services measures derive from a cross disability survey to elicit feedback from adult Medicaid beneficiaries receiving home and community based services (HCBS) about the quality of the long-term services and supports they receive in the community and delivered to them under the auspices of a state Medicaid HCBS program. The unit of analysis is the Medicaid HCBS program, and the accountable entity is the operating entity responsible for managing and overseeing a specific HCBS program within a given state. (For additional information on the accountable entity, see Measures Testing form item #1.5 below.)

The measures consist of seven scale measures, 6 global rating and recommendation measures, and 6 individual measures:
Scale Measures
1. Staff are reliable and helpful –top-box score composed of 6 survey items
2. Staff listen and communicate well –top-box score composed of 11 survey items
3. Case manager is helpful - top-box score composed of 3 survey items
4. Choosing the services that matter to you - top-box score composed of 2 survey items
5. Transportation to medical appointments - top-box score composed of 3 survey items
6. Personal safety and respect - top-box score composed of 3 survey items
7. Planning your time and activities top-box score composed of 6 survey items

Global Ratings Measures
8. Global rating of personal assistance and behavioral health staff- top-box score on a 0-10 scale
9. Global rating of homemaker- top-box score on a 0-10 scale
10. Global rating of case manager- top-box score on a 0-10 scale

Recommendations Measures
11. Would recommend personal assistance/behavioral health staff to family and friends – top-box score on a 1-4 scale (Definitely no, Probably no, Probably yes, Definitely yes)
12. Would recommend homemaker to family and friends –– top-box score on a 1-4 scale (Definitely no, Probably no, Probably yes, Definitely yes)
13. Would recommend case manager to family and friends– top-box score on a 1-4 scale (Definitely no, Probably no, Probably yes, Definitely yes)

Unmet Needs Measures
14. Unmet need in dressing/bathing due to lack of help–top-box score on a Yes, No scale
15. Unmet need in meal preparation/eating due to lack of help– top-box score on a Yes, No scale
16. Unmet need in medication administration due to lack of help– top-box score on a Yes, No scale
17. Unmet need in toileting due to lack of help– top-box score on a Yes, No scale
18. Unmet need with household tasks due to lack of help– top-box score on a Yes, No scale

Physical Safety Measure
19. Hit or hurt by staff – top-box score on a Yes, No scale

1b.1. Developer Rationale:

All measures were developed based on formative research to reflect the constructs most salient to the beneficiaries and other HCBS stakeholders. The research team ran a confirmatory factor analysis to test the hypothesized factor structure for the scale measures. The model was an acceptable fit (RMSEA=0.044, CFI=0.954, TLI=0.950).

Scale Measures

Staff are Reliable and Helpful. Assessing the performance of Medicaid direct care providers (i.e., personal assistants, behavioral health staff, homemakers) from the perspective of the beneficiary is important in evaluating the quality of services they render. This measure is based on beneficiary assessment of direct care staff reliability (showing up on time, stay as long as supposed to, communicate absences) and sensitivity to their privacy needs during the provision of personal care.

Staff Listen and Communicate Well. This measure is based on beneficiary assessment of direct care staff's communication skills and responsiveness to the person's needs. Specifically communication in a way that is understood by the beneficiary, respectful, and staff who listen carefully to what the beneficiary needs/wants and who, therefore, understand what the beneficiary needs. This is essential to the delivery of person-centered care and support. Person-centered care and support is required in Medicaid HCBS programs (Federal Register: https://federalregister.gov/a/2014-00487).

Case Manager Is Helpful. In HCBS programs, the case manager is responsible for monitoring the beneficiary's receipt of services and supports to ensure the service plan is being implemented as specified and that the person's needs are being adequately met. In order to meet these requirements, the case manager must be available to the beneficiary when s/he contacts him/her, and responsive to their changing/emerging needs. This measure is based on the beneficiary's assessment of case manager accessibility and responsiveness.

Choosing Services That Matter to You. A basic tenet of Medicaid HCBS services is that the beneficiary is involved in choosing their services/supports so that the service plan is truly person-centered, and that direct care staff implement the service plan in a person-centered manner. This measure is based on the beneficiary's assessment of the extent to which their service plan and direct care workers are person-centered.

Transportation to Medical Appointments. The health and welfare of beneficiaries must be ensured in the delivery of Medicaid HCBS (42 CFR §441: 302). Integral to assuring the health of beneficiaries is getting to medical appointments. This composite is based on the beneficiary's assessment of the extent to which they have transportation to medical appointments, whether the transportation provider is reliable, and whether the transportation is sufficiently accessible.

Planning your time and activities. Medicaid home and community-based services and supports should facilitate outcomes that are consistent with allowing beneficiaries to live the lives they choose – both in terms of daily routine as well as socializing with family and friends, and engaging in community activities. This measure is based on the beneficiary's assessment of the extent to which they have choice and control over these aspects of their lives.

Personal Safety and Respect. Beneficiaries of Medicaid HCBS should be assured that HCBS providers treat them with respect, that they will not be financially exploited by providers coming into their homes, and that they have someone to go to if they are treated badly. This measure will help HCBS programs assess this aspect of program quality. This measure has very high scores and thus very low variance so there is not much of a performance gap and reliability estimates are low. However, these activities should never occur and are critical to assess.

Individual Item Measures

Global Ratings of Staff (i.e., Personal Assistance/Behavioral Health Staff, Homemaker, Case Manager) – separate measures per staff type. In concert with more specific measures and scale measures, global ratings provide additional information for assessing program quality and can be used as a metric in evaluating quality improvement.

Would Recommend Staff (i.e., Personal Assistance/Behavioral Health Staff, Homemaker, Case Manager) to Family and Friends – separate measures per staff type. Beneficiaries' recommendation are yet another aspect of global experience with a program, and can be used for evaluating program quality and in quality improvement initiatives. While they are measuring similar topics as the global ratings items, these items measure a slightly different aspect of the care experience. The correlations between the related

measures ranged from 0.44-0.52 suggesting that while related, they are measuring slightly different constructs. These measures are frequently requested by CAHPS survey end users for quality improvement initiatives.

Individual Unmet Need Measures:

- Unmet Need in Dressing/bathing Due to Lack of Help
- Unmet Need in Meal Preparation/Eating Due to Lack of Help
- Unmet Need in Medication Administration Due to Lack of Help
- Unmet Need in Toileting Due to Lack of Help
- Unmet Need with Household Tasks Due to Lack of Help

None of the Unmet Need items were captured in a scale measure because they did not correlate with each other in factor analysis. But the advisory panel for the measures development strongly recommended all unmet need stand-alone items be treated as individual measures as the evaluation of unmet need in HCBS is critically important for determining program quality. One of the most basic reasons for the existence of HCBS programs is to meet activities of daily living needs (bathing, dressing, toileting, medication administration) and instrumental activities of daily living (meal preparation/eating, cleaning/laundry) needs that, if not met, both jeopardize beneficiary health and make successful community living untenable. That is, having unmet needs related to these activities places individuals at risk of institutionalization rather than remaining at home and in their communities. Therefore, by definition, the need for assistance among HCBS populations will be high but, when HCBS programs are effective, the *unmet need* experienced by beneficiaries will be low. These measures are intended for use in assessing program quality and for quality improvement initiatives.

Hit or Hurt by Staff. This item was not retained in the Personal Safety and Respect scale measure due to low variation within responses and is thus presented as an individual item measure. This measure has very high scores and thus very low variance so there is not much of a performance gap and reliability estimates are low. However, the advisory panel for the measures development felt this measure is important for establishing the personal safety of program beneficiaries, as physical abuse by staff is a "never event" that should be tracked in any HCBS quality management system.

S.4. **Numerator Statement:** The CAHPS Home- and Community-Based Services measures are created using top-box scoring. This refers to the percentage of respondents that give the most positive response. Details regarding the definition of the most positive response are noted below. HCBS service experience is measured in the following areas. Attached Excel Table S.2b includes the specific item wording for each measure and the response options that go into the numerator.

Scale Measures

1. Staff are reliable and helpful – average proportion of respondents that gave the most positive response on 6 survey items
2. Staff listen and communicate well – average proportion of respondents that gave the most positive response on 11 survey items
3. Case manager is helpful - average proportion of respondents that gave the most positive response on 3 survey items
4. Choosing the services that matter to you - average proportion of respondents that gave the most positive response on 2 survey items
5. Transportation to medical appointments - average proportion of respondents that gave the most positive response on 3 survey items
6. Personal safety and respect - average proportion of respondents that gave the most positive response on 3 survey items
7. Planning your time and activities - average proportion of respondents that gave the most positive response on 6 survey items

Global Rating Measures

8. Global rating of personal assistance and behavioral health staff- average proportion of respondents that gave the most positive response of 9 or 10 on a 0-10 scale
9. Global rating of homemaker- average proportion of respondents that gave the most positive response of 9 or 10 on a 0-10 scale
10. Global rating of case manager- average proportion of respondents that gave the most positive response of 9 or 10 on a 0-10 scale

Recommendation Measures

11. Would recommend personal assistance/behavioral health staff to family and friends – average proportion of respondents that gave the most positive response of "Definitely Yes" on a 1-4 scale (Definitely no, Probably no, Probably yes, Definitely yes)
12. Would recommend homemaker to family and friends — average proportion of respondents that gave the most positive response of "Definitely Yes" on a 1-4 scale (Definitely no, Probably no, Probably yes, Definitely yes)

13. Would recommend case manager to family and friends– average proportion of respondents that gave the most positive response of "Definitely Yes" on a 1-4 scale (Definitely no, Probably no, Probably yes, Definitely yes)

Unmet Needs Measures
14. Unmet need in dressing/bathing due to lack of help–average proportion of respondents that gave the most positive response of "No" on a 1-2 scale (Yes, No)
15. Unmet need in meal preparation/eating due to lack of help–average proportion of respondents that gave the most positive response of "No" on a 1-2 scale (Yes, No)
16. Unmet need in medication administration due to lack of help–average proportion of respondents that gave the most positive response of "No" on a 1-2 scale (Yes, No)
17. Unmet need in toileting due to lack of help–average proportion of respondents that gave the most positive response of "No" on a 1-2 scale (Yes, No)
18. Unmet need with household tasks due to lack of help–average proportion of respondents that gave the most positive response of "No" on a 1-2 scale (Yes, No)

Physical Safety Measure
19. Hit or hurt by staff –average proportion of respondents that gave the most positive response of "No" on a 1-2 scale (Yes, No)

S.7. **Denominator Statement:** The denominator for all measures is the number of survey respondents. Individuals eligible for the CAHPS Home- and Community-Based Services survey include Medicaid beneficiaries who are at least 18 years of age in the sample period, and have received HCBS services for 3 months or longer and their proxies.[2] Eligibility is further determined using three cognitive screening items, administered during the interview:

Q1. Does someone come into your home to help you? (Yes, No)
Q2. How do they help you?
Q3. What do you call them?

Individuals who are unable to answer these cognitive screening items are excluded. Some measures also have topic-specific screening items as well. Additional detail is provided in S.9.

S.10. **Denominator Exclusions:** Individuals less than 18 years of age and individuals that have not received HCBS services for at least 3 months should be excluded. During survey administration, additional exclusions include individuals that failed any of the cognitive screening items mentioned in the denominator statement below. There were 227 beneficiaries excluded due to not passing the cognitive screener (53 Aged/Disabled, 59 ID/DD, 25 TBI, and 90 SMI). Allowing proxy respondents in future administrations has the potential to further reduce these numbers.

De.1. **Measure Type:** PRO
S.23. **Data Source:** Patient Reported Data/Survey
S.26. **Level of Analysis:** HCBS Program

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:

[2] According to guidance produced under the CMS TEFT Technical Assistance contract, individuals who are more likely to be good proxy respondents during the CAHPS Home- and Community-Based Services survey data collection are: (a) those who are willing to respond on behalf of the beneficiary; (b) unpaid caregivers, family members, friends, and neighbors; and (c) those who know the beneficiary well enough that s/he is familiar with the services/supports they are receiving, and has regular, ongoing contact with them. Examples of circumstances that increase the likelihood that someone has knowledge about the beneficiary and their care situation include living with the beneficiary, managing the beneficiary's in-home care for a majority of the day, having regular conversations with the beneficiary about the services they receive, in-person visits with the beneficiary, and being present when services/supports are delivered. Individuals who are less likely to be good proxy respondents are (a) those with paid responsibilities for providing services/supports to the beneficiary, including family members and friends who are paid to help the beneficiary and (b) guardians or conservators whose only responsibility is to oversee the beneficiary's finances.

De.4. **IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** Not applicable.

1. Evidence, 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
HCBS_EoC_NQF_Measures_evidence-attachment_3-29-2016.docx

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., the benefits or improvements in quality envisioned by use of this measure)*

Scale Measures

Staff are Reliable and Helpful. Assessing the performance of Medicaid direct care providers (i.e., personal assistants, behavioral health staff, homemakers) from the perspective of the beneficiary is important in evaluating the quality of services they render. This measure is based on beneficiary assessment of direct care staff reliability (showing up on time, stay as long as supposed to, communicate absences) and sensitivity to their privacy needs during the provision of personal care.

Staff Listen and Communicate Well. This measure is based on beneficiary assessment of direct care staff's communication skills and responsiveness to the person's needs. Specifically communication in a way that is understood by the beneficiary, respectful, and staff who listen carefully to what the beneficiary needs/wants and who, therefore, understand what the beneficiary needs. This is essential to the delivery of person-centered care and support. Person-centered care and support is required in Medicaid HCBS programs (Federal Register: https://federalregister.gov/a/2014-00487).

Case Manager Is Helpful. In HCBS programs, the case manager is responsible for monitoring the beneficiary's receipt of services and supports to ensure the service plan is being implemented as specified and that the person's needs are being adequately met. In order to meet these requirements, the case manager must be available to the beneficiary when s/he contacts him/her, and responsive to their changing/emerging needs. This measure is based on the beneficiary's assessment of case manager accessibility and responsiveness.

Choosing Services That Matter to You. A basic tenet of Medicaid HCBS services is that the beneficiary is involved in choosing their services/supports so that the service plan is truly person-centered, and that direct care staff implement the service plan in a person-centered manner. This measure is based on the beneficiary's assessment of the extent to which their service plan and direct care workers are person-centered.

Transportation to Medical Appointments. The health and welfare of beneficiaries must be ensured in the delivery of Medicaid HCBS (42 CFR §441: 302). Integral to assuring the health of beneficiaries is getting to medical appointments. This composite is based on the beneficiary's assessment of the extent to which they have transportation to medical appointments, whether the transportation provider is reliable, and whether the transportation is sufficiently accessible.

Planning your time and activities. Medicaid home and community-based services and supports should facilitate outcomes that are consistent with allowing beneficiaries to live the lives they choose – both in terms of daily routine as well as socializing with family and friends, and engaging in community activities. This measure is based on the beneficiary's assessment of the extent to which they have choice and control over these aspects of their lives.

Personal Safety and Respect. Beneficiaries of Medicaid HCBS should be assured that HCBS providers treat them with respect, that they will not be financially exploited by providers coming into their homes, and that they have someone to go to if they are treated badly. This measure will help HCBS programs assess this aspect of program quality. This measure has very high scores and thus very low variance so there is not much of a performance gap and reliability estimates are low. However, these activities should never occur and are critical to assess.

Individual Item Measures
Global Ratings of Staff (i.e., Personal Assistance/Behavioral Health Staff, Homemaker, Case Manager) – separate measures per staff type. In concert with more specific measures and scale measures, global ratings provide additional information for assessing program quality and can be used as a metric in evaluating quality improvement.

Would Recommend Staff (i.e., Personal Assistance/Behavioral Health Staff, Homemaker, Case Manager) to Family and Friends – separate measures per staff type. Beneficiaries' recommendation are yet another aspect of global experience with a program, and can be used for evaluating program quality and in quality improvement initiatives. While they are measuring similar topics as the global ratings items, these items measure a slightly different aspect of the care experience. The correlations between the related measures ranged from 0.44-0.52 suggesting that while related, they are measuring slightly different constructs. These measures are frequently requested by CAHPS survey end users for quality improvement initiatives.

Individual Unmet Need Measures:
- Unmet Need in Dressing/bathing Due to Lack of Help
- Unmet Need in Meal Preparation/Eating Due to Lack of Help
- Unmet Need in Medication Administration Due to Lack of Help
- Unmet Need in Toileting Due to Lack of Help
- Unmet Need with Household Tasks Due to Lack of Help

None of the Unmet Need items were captured in a scale measure because they did not correlate with each other in factor analysis. But the advisory panel for the measures development strongly recommended all unmet need stand-alone items be treated as individual measures as the evaluation of unmet need in HCBS is critically important for determining program quality. One of the most basic reasons for the existence of HCBS programs is to meet activities of daily living needs (bathing, dressing, toileting, medication administration) and instrumental activities of daily living (meal preparation/eating, cleaning/laundry) needs that, if not met, both jeopardize beneficiary health and make successful community living untenable. That is, having unmet needs related to these activities places individuals at risk of institutionalization rather than remaining at home and in their communities. Therefore, by definition, the need for assistance among HCBS populations will be high but, when HCBS programs are effective, the *unmet need* experienced by beneficiaries will be low. These measures are intended for use in assessing program quality and for quality improvement initiatives.

Hit or Hurt by Staff. This item was not retained in the Personal Safety and Respect scale measure due to low variation within responses and is thus presented as an individual item measure. This measure has very high scores and thus very low variance so there is not much of a performance gap and reliability estimates are low. However, the advisory panel for the measures development felt this measure is important for establishing the personal safety of program beneficiaries, as physical abuse by staff is a "never event" that should be tracked in any HCBS quality management system.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. Include mean, 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

Measure	Number of Programs	Top-box Mean	Standard Deviation	Minimum	25th Percentile	50th Percentile	75th Percentile	Maximum
Scale Measures								
Staff are reliable and helpful	25	89.5	3.77	78.65	87.36	90.02	92.19	94.07
Staff listen and communicate well	25	89.89	3.53	79.89	88.08	90.77	92.29	94.49

Case manager is helpful	25	91.27	3.94	84.28	88.61	91.59	93.86	97.94
Choosing the services that matter to you	25	75.71	7.29	60.71	71.88	75.71	78.88	89.54
Transportation to medical appointments	25	85.41	5.48	73.86	82.24	85.5	88.94	95.72
Personal safety and respect	25	97.56	1.09	95.42	96.51	97.73	98.42	99.69
Planning your time and activities	25	75.68	3.04	70.34	74.42	75.25	77.69	81.3
Global Ratings Measures								
Global Rating of Personal Assistance/Behavioral Health Staff	25	65.34	9.71	31.99	60.22	67.57	71.77	78.42
Global Rating of Homemaker	18	63.63	13.26	39.66	57.9	65.34	75.34	81.29
Global Rating of Case Manager	25	59.99	8.84	40.41	54.74	59.99	66.82	73.18
Recommendation Measures								
Recommendation of Personal Assistance/Behavioral Health Staff	25	76.58	9.08	52.09	70.9	78.67	83.58	89.6
Recommendation of Homemaker	18	74.59	14.08	49.02	63.41	75.29	81.93	96.24
Recommendation of Case Manager	25	71.84	7.87	56.22	66.83	72.11	77.52	89.69
Unmet Needs Measures								
There are no staff to help dress, shower, or bathe	16	43.76	23.3	0	31.18	51.34	56.11	88.9
Sufficient staff to help you with meals	17	38.95	26.95	0	23.34	39.67	46.25	100
Sufficient staff to help you with medications	19	70.67	19.83	28.51	61.91	74.01	84.41	100
Sufficient staff to help you with toileting	23	96.19	4.83	80.79	94.43	97.91	100	100
Sufficient homemakers to help you with household tasks	19	50.38	23.07	0	35.15	50.8	72.99	80.81
Physical Safety Measure								
Not hit or hurt by staff*	25	99.73	0.53	97.76	99.76	99.97	100	100
* The original question was written negatively, but we are reporting it positively to support comparing the questions.								

See table 1b.2 in attached tables for more detailed results.

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 endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
See tables 1b.4a, 1b.4b, 1b.4c, 1b.4d, and 1b.4e for disparities data from the measures by population group (disability type).

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.
The measures in this submission focus on people who are elderly with disabilities, individuals with physical disabilities, persons with intellectual/developmental disability, individuals with brain injury, and those with serious mental illness. who receive Medicaid-funded home and community-based services. The Medicaid population with disabilities is, by definition, a population with substantially limited economic resources. Consistent with Medicaid status, adults with disability have a higher poverty rate than those without disability [age 18-64: 28.2% vs. 13.9%, respectively; age 65+: 13.0% vs 7.5% respectively (U.S. Census Bureau, 2014a)]. In addition, U.S. working age adults (Age 18-64) with disability have a lower employment rate than their non-disabled peers [34.4% vs. 75.4% (U.S. Census Bureau, 2014b)].

In terms of racial/ethnic disparities, Blacks, Hispanics and American Indians/Alaskan Natives (AIAN) have higher prevalence of disabilities in self-care and independent living than does the total U.S. adult population with these types of disabilities. These types of disabilities mirror those that beneficiaries in Medicaid HCBS programs tend to exhibit. In the US, 2.1% of the adult population has self-care disabilities and 6.1% have independent living disabilities, respectively. This contrasts to Blacks with respective prevalence of 5.7% and 9.2%; Hispanics at 4.8% and 7.7%; and AIAN at 6.6% and 11.4% (CDC, 2013).

Safety is a major concern for programs serving people with disabilities, who experience higher rates of violent crime victimization. The rate of victimization from violent crime for the U.S. population without disabilities is 14 per 1,000 population. For people with disabilities of the type served in HCBS programs (i.e., disabilities in self-care and independent living), the rates are 26.0/1,000 and 32.4/1,000, respectively. Of most relevance to the safety-related measures in this submission would be statistics on victimization from abuse by paid caregivers; however, the Department of Justice's estimates do not identify paid caregivers as a category of perpetrator (Harrell, 2015).

U.S. Census Bureau. (2014). 2014 American Community Survey, 1-Year Estimates, American FactFinder, Table B18130; http://factfinder.census.gov .

U.S. Census Bureau. (2014). 2014 American Community Survey, 1-Year Estimates, American FactFinder, Table B18120; http://factfinder.census.gov .

Center for Disease Control, Online Disability and Health Data System. (2013). http://www.cdc.gov/ncbddd/disabilityandhealth/dhds.html. Data from the 2013 Behavioral Risk Factor Surveillance System (BRFSS).

Harrell, E. (2015). Crime Against Persons with Disabilities, Statistical Tables. U.S. Department of Justice, May 2015, http://www.bjs.gov/content/pub/pdf/capd0913st.pdf.

1c. High Priority (previously referred to as High Impact)
The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR

- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
High resource use
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
The development and testing of the measures included in this submission are in direct response to the dearth of performance and quality measures for the increasing HCBS population. As pointed out in a recent report from the NQF project on Home and Community-Based Services Quality: "… there is a lack of systematic measurement of the quality of HCBS across payers and delivery systems (NQF, 2015)."

Rigorously tested quality measures for HCBS is becoming increasingly important as government funding for long-term care has shifted from the provision of care in institutional settings to care at home and in the community. For the first time, in 2013, Medicaid expenditures for HCBS surpassed institutional expenditures, and the trend is expected to continue in the years ahead. The amount of state and federal Medicaid expenditures that are devoted to HCBS has steadily increased since the introduction of Medicaid HCBS programs over 35 years ago. In 2013, Medicaid expenditures for HCBS totaled $74.8 billion (Eiken et al., 2013).

Of all Medicaid funding for individuals receiving long-term services and supports (community-based and institutional care), HCBS accounted for 72% of spending in programs targeting people with developmental disabilities, 40% of spending for programs targeting older people and people with physical disabilities, and 36% of spending for programs serving individuals with serious mental illness or serious emotional disorders (Eiken et al., 2013).

An estimated 3.4 million people used Medicaid HCBS in 2011, 71 percent of all LTSS beneficiaries. This figure includes 1,567,198 people who received services authorized under Section 1915(c) of the Social Security Act, commonly referred to as "HCBS waivers" (Eiken et al., 2015). In a separate report focused on HCBS waivers, CMS-approved State Medicaid reports (from the CMS Reporting Form 372) indicated the following number of people served by population in 2012:

- 792,261 were elders or people with physical disabilities;
- 602,958 were persons with intellectual or developmental disabilities;
- 11,547 were persons with serious mental illness or serious emotional disorder; and
- 10,959 were individuals with brain injury (Eiken, 2012).

It should be noted that these statistics are an undercount of the actual number of individuals receiving 1915(c) waiver services in 2012. Data reported by states on the CMS Form 372 Reports represent 284 of the 305 1915(c) waiver programs in operation that year. Only 372 Reports submitted and approved by CMS are represented in the statistics cited above. In addition to the 1915(c) HCBS waiver programs, in 2012 four states provided services/supports to Medicaid beneficiaries through Medicaid managed long-term services and supports (MLTSS) programs authorized under Section 1115 of the Social Security Act. The numbers served in these MLTSS programs is not available from the 372 Reports.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
National Quality Forum. (2015). Addressing Performance Measure Gaps in Home and Community-Based Services to Support Community Living: Synthesis of Evidence and Environmental Scan, Interim Report. December 18, 2015.

Eiken, S., Sredl, K., Burwell, B., and Saucier, P. (2013). Medicaid Expenditures for Long-Term Services and Supports (LTSS) in FY 2013: Home and Community-Based Services were a Majority of LTSS Spending. Truven Health Analytics, June 30, 2015. https://www.medicaid.gov/medicaid-chip-program-information/by-topics/long-term-services-and-supports/downloads/ltss-expenditures-fy2013.pdf

Eiken, S., Sredl, K., Saucier, P., Burwell, B. (2015). Medicaid Long-Term Services and Supports Beneficiaries in 2011, Truven Health Analytics, September 22, 2015. https://www.medicaid.gov/medicaid-chip-program-information/by-topics/long-term-services-and-supports/downloads/ltss-beneficiaries-report-2011.pdf

Eiken, S. (2012). Medicaid 1915(c) Waiver Data Based on CMS 372 Report, 2011-2012, Truven Health Analytics, September 17, 2015. https://www.medicaid.gov/medicaid-chip-program-information/by-topics/long-term-services-and-supports/downloads/cms-372-report-2012.pdf

1c.5. If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful. *(Describe how and from whom their input was obtained.)*

During the development of the survey from which these measures are, the HCBS beneficiary audience as well as stakeholders in the broader HCBS community have consistently supported the proposed measures as necessary and important.

Truven Health Analytics conducted a literature review for AHRQ that included identifying measures and gaps in the measures. The CAHPS Home- and Community-Based Services Survey developer team conducted a follow-up literature review for the time period of 2007 through 2010.

The research team received input from a focus group and interviews, and CMS posted 60-day and 30-day Federal Register notices on May 18, 2012 and July 24, 2012, respectively, for public comment on the proposed data collection (as required by the OMB Paperwork Requirement Act). No comments were received.

The team also conducted three rounds of cognitive testing in English and one round in Spanish, including comparison of approaches to determine the best method for asking questions of respondents with intellectual impairments. Round 1 of English cognitive testing was conducted in the larger Boston and Raleigh Durham area between January and February 2011. Round 2 of English cognitive testing was conducted in March and April of 2011 in the Raleigh-Durham and greater Boston areas. Round 3 of English cognitive testing was conducted between November 2011 and January 2012. The Spanish cognitive testing was conducted between November 2011 and January 2012 in Florida. Each interview was conducted in person by a trained qualitative interviewer and note taker. Interviews used a structured cognitive testing protocol.

For each round of cognitive testing, the research team assessed comprehension, item order, content, and the respondents' abilities to make judgments about each item and select a response. The research team also tested alternative formats of items and response options. These tests were conducted to develop an alternative version to allow for more participation by individuals with intellectual or cognitive impairments or who otherwise would find the conventional CAHPS format cognitively burdensome. The research team tested the following frequency scales for individual quality ratings:

- Never, sometimes, usually, always (traditional CAHPS 4-point response)
- None, some, most, all of the time (alternate response)
- Mostly yes, mostly no (alternate binary response)

The research team tested the following rating scales to evaluate services overall:

- 0 to 10, with 0 being the worst possible staff and 10 being the best possible staff
- Excellent, very good, good, fair, poor staff rating

The research team also tested the labels of the scale measures in August and September 2015. These were conducted in-person with beneficiaries in Connecticut, Maryland and New Hampshire. The purpose was to provide recommendations for labels that reflected the experiences of HCBS beneficiaries. Prior to testing, two to three labels were developed for each scale taking into account plain language and best practices for public reporting. These labels were then reviewed and slightly revised by state staff participating in the CAHPS Home- and Community-Based Services Survey pilot and field tests.

In addition, there was a Federal Advisory Panel consisting of:

- CMS-Disabled and Elderly Health Programs Group: Anita Yuskauskas (Chair), Mary Sowers, Kathy Poisal, Mary Beth Ribar, Sara Fogler, Carey Appold,
- CMS-Children & Families Health Program Group: Charlie Mackay and John Young
- CMS-Center for Drug and Health Plan Choice: Liz (Elizabeth) Goldstein, Suzanne Rotwein, Lori Teichman, Ted (Edward) Sekscenski, Bill (William) Lehrman, Barb (Barbara) Crawley
- Agency for Healthcare Research and Quality: DEB Potter, Judy Sangl

The research team identified and invited experts and key stakeholders, including representatives of state HCBS programs, self-advocacy groups for people with disabilities, survey development and reporting experts, CAHPS Consortium representatives, and

Federal Government staff, to provide feedback on the development of the survey and the field test process. The organizations represented include:
• Linda Anthony, Disability Rights Network of Pennsylvania and ADAPT, Consumer advocate—adults with physical disabilities
• Julie Brown, RAND Corporation, CAHPS Consortium
• Marcus Canaday, West Virginia Bureau for Medical Services, State HCBS programs for adults with physical disabilities
• Steve Dunaway, Florida Agency for Persons with Disabilities, State HCBS programs for adults with intellectual disabilities
• Chester Finn, Self Advocates Becoming Empowered, Consumer advocate—adults with intellectual disabilities
• Michelle Goody, Massachusetts Medicaid, Medicaid
• Ron Honberg and Sita Diehl, National Alliance on Mental Illness, Consumer advocate—adults with mental illness
• Ari Houser, AARP, Consumer advocate—older adults with disabling/chronic conditions
• Christian Koltonski, Colorado Medicaid, Medicaid
• Jeanne Levelle, Louisiana Medicaid, Medicaid
• Ted Lutterman, National Association of State Mental Health Program Directors, State HCBS programs for adults with mental illness
• Chas Moseley and Nancy Thaler, National Association of State Directors of Developmental Disabilities Services, State HCBS programs for adults with intellectual disabilities
• Sue Palsbo, George Mason University, Survey development for people with physical disabilities
• Teresa Richard, Texas Department of Aging and Disability Services, State HCBS programs—all populations
• Steve Staugaitis, University of Massachusetts Medical School, Performance measures for people with intellectual disabilities
• John Thompson and Kelsey Walter, National Association of States United for Aging and Disabilities, State HCBS programs for older adults with disabilities
• Sally Varney, New Hampshire Medicaid, Medicaid
• Sandeep Wadhwa and Matt Salo, National Association of Medicaid Directors (NAMD) and Colorado Department of Health Care Policy and Financing, State HCBS programs—all populations
• Lorraine Wargo, National Association of State Head Injury Administrators, State HCBS programs for adults with head injuries

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 subcriteria 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):*
Mental Health : Serious Mental Illness, Neurology : Brain Injury, Neurology : Cognitive Impairment/Dementia

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Access, Care Coordination, Functional Status, Health and Functional Status : Development/Wellness, Patient and Family Engagement, Safety

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.)*
The survey and related materials (including information about any NQF endorsed measures derived from the survey) will be available on CMS' Medicaid.gov website and as a link from the CMS CAHPS webpage; a link will also appear on AHRQ's CAHPS website. The survey instruments in English and Spanish are attached for reference.

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:** CAHPS HCBS Supplementary Tables July Submission

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Not applicable.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
The CAHPS Home- and Community-Based Services measures are created using top-box scoring. This refers to the percentage of respondents that give the most positive response. Details regarding the definition of the most positive response are noted below. HCBS service experience is measured in the following areas. Attached Excel Table S.2b includes the specific item wording for each measure and the response options that go into the numerator.

Scale Measures
1. Staff are reliable and helpful – average proportion of respondents that gave the most positive response on 6 survey items
2. Staff listen and communicate well – average proportion of respondents that gave the most positive response on 11 survey items
3. Case manager is helpful - average proportion of respondents that gave the most positive response on 3 survey items
4. Choosing the services that matter to you - average proportion of respondents that gave the most positive response on 2 survey items
5. Transportation to medical appointments - average proportion of respondents that gave the most positive response on 3 survey items
6. Personal safety and respect - average proportion of respondents that gave the most positive response on 3 survey items
7. Planning your time and activities - average proportion of respondents that gave the most positive response on 6 survey items

Global Rating Measures
8. Global rating of personal assistance and behavioral health staff- average proportion of respondents that gave the most positive response of 9 or 10 on a 0-10 scale
9. Global rating of homemaker- average proportion of respondents that gave the most positive response of 9 or 10 on a 0-10 scale
10. Global rating of case manager- average proportion of respondents that gave the most positive response of 9 or 10 on a 0-10 scale

Recommendation Measures
11. Would recommend personal assistance/behavioral health staff to family and friends – average proportion of respondents that gave the most positive response of "Definitely Yes" on a 1-4 scale (Definitely no, Probably no, Probably yes, Definitely yes)
12. Would recommend homemaker to family and friends –– average proportion of respondents that gave the most positive response of "Definitely Yes" on a 1-4 scale (Definitely no, Probably no, Probably yes, Definitely yes)
13. Would recommend case manager to family and friends– average proportion of respondents that gave the most positive response of "Definitely Yes" on a 1-4 scale (Definitely no, Probably no, Probably yes, Definitely yes)

Unmet Needs Measures
14. Unmet need in dressing/bathing due to lack of help–average proportion of respondents that gave the most positive response of "No" score on a 1-2 scale (Yes, No)
15. Unmet need in meal preparation/eating due to lack of help–average proportion of respondents that gave the most positive response of "No" score on a 1-2 scale (Yes, No)
16. Unmet need in medication administration due to lack of help–average proportion of respondents that gave the most positive response of "No" score on a 1-2 scale (Yes, No)
17. Unmet need in toileting due to lack of help–average proportion of respondents that gave the most positive response of "No" score on a 1-2 scale (Yes, No)
18. Unmet need with household tasks due to lack of help–average proportion of respondents that gave the most positive response of "No" score on a 1-2 scale (Yes, No)

Physical Safety Measure
19. Hit or hurt by staff –average proportion of respondents that gave the most positive response of "No" score on a 1-2 scale (Yes, No)

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Originally there was no time reference for measures because cognitive testing showed that this was cognitively burdensome for respondents. This followed the same approach as the CAHPS Nursing Home Long Stay survey, which measures experience of care for a similar population and used a non-specific reference period based on cognitive testing findings (Sangl et al., 2007). The CAHPS Consortium has since requested that "in the last 3 months" to be added to the survey items to maintain consistency with other CAHPS surveys as a condition for trademark designation.

The frequency of data collection/aggregation will be at the discretion of state users, as CMS has determined the survey from which the measures are derived will be conducted on a voluntary by states. It is anticipated that states would field the survey no more frequently than annually per HCBS program. Some states may choose to field it less frequently than annually. Reporting of measures would follow at intervals paralleling data collection time frames.

Sangl, J., Buchanan, J., Cosenza, C., Bernard, S., Keller, S., Mitchell, N., and Larwood D. (2007). The development of a CAHPS instrument for Nursing Home Residents (NHCAHPS). J Aging Soc Policy. 19(2):63-82. PubMed PMID: 17409047.

S.6. **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, 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.
Attached Excel Table S.2b includes the specific item wording for each measure and the response options that go into the numerator[3].

To calculate the program-level scores:
Score each item using the top box method; calculate a mode adjusted score for each respondent; calculate case mix adjusted scores for each program; and calculate means for the scale measures.

Scale Measures:
For each survey item, the top box numerator is the number of respondents who selected the most positive response category.

Staff are reliable and helpful – survey items 13 14 15 19 37 38
13: How often do [personal assistance/behavioral health staff] come to work on time?
14: How often do [personal assistance/behavioral health staff] work as long as they are supposed to?
15: Sometimes staff cannot come to work on a day that they are scheduled. When staff cannot come to work on a day that they are scheduled, does someone let you know if [personal assistance/behavioral health staff] cannot come that day?
19: How often do [personal assistance/behavioral health staff] make sure you have enough personal privacy when you dress, take a shower, or bathe?
37: How often do [homemakers] come to work on time?
38: How often do [homemakers] work as long as they are supposed to?

Staff listen and communicate well – survey items 28 29 30 31 32 33 41 42 43 44 45
28: How often are [personal assistance/behavioral health staff] nice and polite to you?[4]

[3] The CAHPS Home- and Community-Based Services Survey items now include a 3 month time referent to be consistent with other CAHPS surveys as a condition for the CAHPS trademark designation.
[4] "Nice and polite" was changed to "courtesy and respect" to be consistent with other CAHPS surveys as a condition for the CAHPS trademark designation.

29: How often are the explanations [personal assistance/behavioral health staff] gives you hard to understand because of an accent or the way he or she speaks English?*
30: How often do [personal assistance/behavioral health staff] treat you the way you want them to?
31: How often do [personal assistance/behavioral health staff] explain things in a way that is easy to understand?
32: How often do [personal assistance/behavioral health staff] listen carefully to you?
33: Do you feel [personal assistance/behavioral health staff] know what kind of help you need with everyday activities, like getting ready in the morning, getting groceries, or going places in your community?
41: How often are [homemakers] nice and polite to you?[5]
42: How often are the explanations [homemaker] gives you hard to understand because of an accent or the way the provider speaks English?*
43: How often do [homemakers] treat you the way you want them to?
44: How often do [homemakers] listen carefully to you?
45: Do you feel [homemakers] know what kind of help you need?

Case manager is helpful – survey items 49 51 53
49: Can you contact this [case manager] when you need to?
51: Did this [case manager] work with you when you asked for help with getting or fixing equipment?
53: Did this [case manager] work with you when you asked for help with getting other changes to your services?

Choosing the services that matter to you – survey items 56 57
56: Does your [program-specific term for "service plan"] include . . . ?
57: Do you feel [personal assistance/behavioral health staff] know what's on your [program-specific term for "service plan"], including the things that are important to you?

Transportation to medical appointments – survey items 59 61 62
59: Medical appointments include seeing a doctor, a dentist, a therapist, or someone else who takes care of your health. How often do you have a way to get to your medical appointments?
61: Are you able to get in and out of this ride easily?
62: How often does this ride arrive on time to pick you up?

Personal safety and respect – survey items 64 65 68
64: Is there a person you can talk to if someone hurts you or does something to you that you don't like?
65: Do any of the [personal assistance/behavioral health staff, homemakers, or your case managers] that you have now take your money or your things without asking you first?*
68: Do any [staff] that you have now yell, swear, or curse at you?*

Planning your time and activities – survey items 75 77 78 79 80 81
75: When you want to, how often can you get together with these family members who live nearby?
77: When you want to, how often can you get together with these friends who live nearby?
78: When you want to, how often can you do things in the community that you like?
79: Do you need more help than you get now from [personal assistance/behavioral health staff] to do things in your community?*
80: Do you take part in deciding what you do with your time each day?
81: Do you take part in deciding when you do things each day—for example, deciding when you get up, eat, or go to bed?

Global Ratings Measures:
The numerator for each Global measure includes the number of respondents who answered 9 or 10 for the item (on a scale of 0 to 10).

Global rating of personal assistance and behavioral health staff– survey item 35

[5] "Nice and polite" was changed to "courtesy and respect" to be consistent with other CAHPS surveys as a condition for the CAHPS trademark designation.

35: Using any number from 0 to 10, where 0 is the worst help from {personal assistance/behavioral health staff} possible and 10 is the best help from {personal assistance/behavioral health staff} possible, what number would you use to rate the help you get from {personal assistance/behavioral health staff}?

Global rating of homemaker – survey item 46
46: Using any number from 0 to 10, where 0 is the worst help from {homemakers} possible and 10 is the best help from {homemakers} possible, what number would you use to rate the help you get from {homemakers}?

Global rating of case manager– survey item 54
54: Using any number from 0 to 10, where 0 is the worst help from {case manager} possible and 10 is the best help from {case manager}possible, what number would you use to rate the help you get from {case manager}?

Recommendation Measures:
The numerator for each Recommendation measure includes the number of respondents who answered "Definitely yes" for the item (on a scale of "Definitely no", "Probably no", "Probably yes", "Definitely yes"). Item numbers and item text are listed below.

Would recommend personal assistance/behavioral health staff to family and friends – survey item 36
36: Would you recommend the {personal assistance/behavioral health staff} who help you to your family and friends if they needed help with everyday activities? Would you say you recommend the {personal assistance/behavioral health staff} . . .

Would recommend homemaker to family and friends – survey item 47
47: Would you recommend the {homemakers} who help you to your family and friends if they needed {program-specific term for homemaker services}? Would you say you recommend the {homemakers} . . .

Would recommend case manager to family and friends– survey item 55
55: Would you recommend the {case manager} who helps you to your family and friends if they needed {program-specific term for case-management services}? Would you say you recommend the {case manager} . . .

Unmet Needs Measures:
The numerator for each Unmet Needs measure includes the number of respondents who answered "no" for that item (these items are then reverse coded so that higher scores reflect a better experience). Item numbers and item text are listed below.

Unmet need in dressing/bathing due to lack of help - survey item 18
18: Is this because there are no {personal assistance/behavioral health staff} to help you?

Unmet need in meal preparation/eating due to lack of help - survey item 22
22: Is this because there are no {personal assistance/behavioral health staff} to help you?

Unmet need in medication administration due to lack of help - survey item 25
25: Is this because there are no {personal assistance/behavioral health staff} to help you?

Unmet need in toileting due to lack of help - survey item 27
27: Do you get all the help you need with toileting from {personal assistance/behavioral health staff} when you need it? (not reverse coded).

Unmet need with household tasks due to lack of help - survey item 40
40: Is this because there are no {homemakers} to help you? [ASK IF HOMEMAKER IS THE SAME AS PCA STAFF]

Physical Safety Measure:
The numerator for the following Physical Safety measure includes the number of respondents who answered "no" for this item. item (these items are then reverse coded so that higher scores reflect a better experience). The item number and item text is listed below.

Hit or hurt by staff – survey item 71
71: Do any {staff} that you have now hit you or hurt you?

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The denominator for all measures is the number of survey respondents. Individuals eligible for the CAHPS Home- and Community-Based Services survey include Medicaid beneficiaries who are at least 18 years of age in the sample period, and have received HCBS services for 3 months or longer and their proxies.[6] Eligibility is further determined using three cognitive screening items, administered during the interview:

Q1. Does someone come into your home to help you? (Yes, No)
Q2. How do they help you?
Q3. What do you call them?

Individuals who are unable to answer these cognitive screening items are excluded. Some measures also have topic-specific screening items as well. Additional detail is provided in S.9.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
While there are a myriad of home and community-based services and supports (HCBS) that Medicaid programs provide (at their discretion) to beneficiaries with long-term care needs, the proposed provider-related measures in this submission focus on the most common provider types for adults receiving Medicaid HCBS. These include personal assistance providers, behavioral health staff, homemakers and case managers.

While Medicare-certified home health agencies may provide similar services to Medicare beneficiaries, the Medicare benefit is a post-acute care benefit and typically limited to episodes following hospitalization. Medicaid home and community-based services are a long-term care benefit and support persons with long-term care needs over lengthier durations. Personal assistance services, help in the home by behavioral health staff, and homemaker services typically involve assistance with activities of daily living (bathing, dressing, grooming, toileting, eating; mobility) and instrumental activities of daily living (meal preparation, housework, laundry, food shopping). Case management is an integral component of Medicaid HCBS programs; the role of the case manager includes working with the beneficiary to assesses his/her need for services/supports and to develop a person-centered care/service plan, monitoring service delivery, and responding to the individual's changing needs and circumstances.

Not all HCBS beneficiaries receive all services. Q4, Q6, Q8, and Q11 assess which services the beneficiary receives. Beneficiaries are then eligible for different survey questions based on these responses.

These questions are:
Q4. Do you get {program specific term for personal assistance} at home?
Q6. Do you get {program specific term for behavioral health specialist services} at home?
Q8. Do you get {program specific term for homemaker services} at home?

[6] According to guidance produced under the CMS TEFT Technical Assistance contract, individuals who are more likely to be good proxy respondents during the CAHPS Home- and Community-Based Services survey data collection are: (a) those who are willing to respond on behalf of the beneficiary; (b) unpaid caregivers, family members, friends, and neighbors; and (c) those who know the beneficiary well enough that s/he is familiar with the services/supports they are receiving, and has regular, ongoing contact with them. Examples of circumstances that increase the likelihood that someone has knowledge about the beneficiary and their care situation include living with the beneficiary, managing the beneficiary's in-home care for a majority of the day, having regular conversations with the beneficiary about the services they receive, in-person visits with the beneficiary, and being present when services/supports are delivered. Individuals who are less likely to be good proxy respondents are (a) those with paid responsibilities for providing services/supports to the beneficiary, including family members and friends who are paid to help the beneficiary and (b) guardians or conservators whose only responsibility is to oversee the beneficiary's finances.

Q11. Do you get help from {program specific term for case manager services} to help make sure that you have all the services you need?

Scale Measure 1: Staff are reliable and helpful
Q13: the number of surveys completed by all those who responded "yes" to screener Q4 or Q6
Q14: the number of surveys completed by all those who responded "yes" to screener Q4 or Q6
Q15: the number of surveys completed by all those who responded "yes" to screener Q4 or Q6
Q19: the number of surveys completed by all those who responded "yes" to screener Q4 or Q6
Q37: the number of surveys completed by all those who responded "yes" to screener Q8
Q38: the number of surveys completed by all those who responded "yes" to screener Q8

Scale Measure 2: Staff listen and communicate well
Q28: the number of surveys completed by all those who responded "yes" to screener Q4 or Q6
Q29: the number of surveys completed by all those who responded "yes" to screener Q4 or Q6
Q30: the number of surveys completed by all those who responded "yes" to screener Q4 or Q6
Q31: the number of surveys completed by all those who responded "yes" to screener Q4 or Q6
Q32: the number of surveys completed by all those who responded "yes" to screener Q4 or Q6
Q33: the number of surveys completed by all those who responded "yes" to screener Q4 or Q6
Q41: the number of surveys completed by all those who responded "yes" to screener Q8
Q42: the number of surveys completed by all those who responded "yes" to screener Q8
Q43: the number of surveys completed by all those who responded "yes" to screener Q8
Q44: the number of surveys completed by all those who responded "yes" to screener Q8
Q45: the number of surveys completed by all those who responded "yes" to screener Q8

Scale Measure 3: Case manager is helpful
Q49: the number of surveys completed by all those who responded "yes" to screener Q11
Q51: the number of surveys completed by all those who responded "yes" to screener Q11
Q53: the number of surveys completed by all those who responded "yes" to screener Q11

Scale Measure 4: Choosing the services that matter to you
Q56: the number of surveys completed by all those who responded "yes" to screener Q4, Q6, Q8, or Q11
Q57: the number of surveys completed by all those who responded "yes" to screener Q4, Q6, Q8, or Q11

Scale Measure 5: Transportation to medical appointments
Q59: the number of surveys completed by all those who responded "yes" to screener Q4, Q6, Q8, or Q11
Q61: the number of surveys completed by all those who responded "yes" to screener Q4, Q6, Q8, or Q11
Q62: the number of surveys completed by all those who responded "yes" to screener Q4, Q6, Q8, or Q11

Scale Measure 6: Personal safety and respect
Q64: the number of surveys completed by all those who responded "yes" to screener Q4, Q6, Q8, or Q11
Q65: the number of surveys completed by all those who responded "yes" to screener Q4, Q6, Q8, or Q11
Q68: the number of surveys completed by all those who responded "yes" to screener Q4, Q6, Q8, or Q11

Scale Measure 7: Planning your time and activities
Q75: the number of surveys completed by all those who responded "yes" to screener Q4, Q6, Q8, or Q11
Q77: the number of surveys completed by all those who responded "yes" to screener Q4, Q6, Q8, or Q11
Q78: the number of surveys completed by all those who responded "yes" to screener Q4, Q6, Q8, or Q11
Q79: the number of surveys completed by all those who responded "yes" to screener Q4 or Q6
Q80: the number of surveys completed by all those who responded "yes" to screener Q4, Q6, Q8, or Q11
Q81: the number of surveys completed by all those who responded "yes" to screener Q4, Q6, Q8, or Q11

Global Rating Measures:
Global rating of personal assistance and behavioral health staff
Q35: the number of surveys completed by all those who responded "yes" to screener Q4 or Q6
Global rating of homemaker
Q46: the number of surveys completed by all those who responded "yes" to screener Q8

Global rating of case manager
Q54: the number of surveys completed by all those who responded "yes" to screener Q11

Recommendation Measures:
Recommendation of personal assistance and behavioral health staff to family/friends
Q36: the number of surveys completed by all those who responded "yes" to screener Q4 or Q6
Recommendation of homemaker to family/friends
Q47: the number of surveys completed by all those who responded "yes" to screener Q8
Recommendation of case manager to family/friends
Q55: the number of surveys completed by all those who responded "yes" to screener Q11

Unmet Needs Measures:
Unmet need in dressing/bathing due to lack of help -
Q18: the number of surveys completed by all those who responded "yes" to Q17
Unmet need in meal preparation/eating due to lack of help
Q22: the number of surveys completed by all those who responded "yes" to Q21
Unmet need in medication administration due to lack of help
Q25: the number of surveys completed by all those who responded "yes" to Q24
Unmet need in toileting due to lack of help -
Q27: the number of surveys completed by all those who responded "yes" to Q26
Unmet need with household tasks due to lack of help
Q40: the number of surveys completed by all those who responded "yes" to Q39

Personal Safety Measures:
Hit or hurt by staff
Q71: the number of surveys completed by all those who responded "yes" to screener Q4, Q6, Q8, or Q11

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Individuals less than 18 years of age and individuals that have not received HCBS services for at least 3 months should be excluded. During survey administration, additional exclusions include individuals that failed any of the cognitive screening items mentioned in the denominator statement below. There were 227 beneficiaries excluded due to not passing the cognitive screener (53 Aged/Disabled, 59 ID/DD, 25 TBI, and 90 SMI). Allowing proxy respondents in future administrations has the potential to further reduce these numbers.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
Individuals who are unable to answer one or more of the following cognitive screening items should be excluded. If the respondent is not able to answer (e.g., provides an invalid/nonsensical response, does not respond, or indicates "I don't know"), the interviewer should end the interview.

1. Does someone come into your home to help you? (Yes, No)

2. How do they help you? (open ended)
Examples of correct responses include:
- "Helps me get ready every day"
- "Cleans my home"
- "Works with me at my job"
- "Helps me to do things"
- "Drives me around"

3. What do you call them? (open ended)
Examples of sufficient responses include:
- "My worker"
- "My assistant"
- Names of staff ("Jo", "Dawn", etc.)

S.12. Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
The intended primary unit of analysis is the Medicaid HCBS program. However, states may wish to stratify by sub-state agencies such as counties or regional entities with program operational and budgetary authority. In some instances, a state may wish to stratify by case-management agency as well, given they are typically viewed as having substantial responsibility for developing beneficiary service and support plans as well as monitoring whether the service/support plan addresses the person's needs and meet their goals.

States are increasingly moving users of Medicaid long-term services and supports, including HCBS, into managed care arrangements (typically referred to as Managed Long-Term Services and Supports or MLTSS) where the managed care organization (MCO) is the primary accountable entity for ensuring HCBS beneficiary, health, welfare and quality of life. As such, we also anticipate some states may want to stratify based on (MCO).

S.13. Risk Adjustment Type (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Case-mix adjustment is done via regression methodology or a covariance adjustment. We use case-mix adjustment to adjust scores for various patient and survey mode characteristics. The research team suggests general health rating, mental health rating, age, gender, whether respondent lives alone, and response option as case- mix adjusters for the CAHPS Home- and Community-Based Services measures based on our analysis. We also recommend including survey mode as an additional adjustment variable and proxy status if proxy respondents are utilized. Finally, future administrations of the survey should also include education to be consistent with CAHPS survey methodology.

The specific survey items used to develop case mix adjustment are:

General health rating:
In general, how would you rate your overall health? Would you say . . .
Excellent,
Very good,
Good,
Fair, or
Poor?
DON'T KNOW
REFUSED
UNCLEAR RESPONSE

Mental health rating:
In general, how would you rate your overall mental or emotional health? Would you say . . .
Excellent,
Very good,
Good,
Fair, or
Poor?
DON'T KNOW
REFUSED
UNCLEAR RESPONSE

Age:
What is your age?
18 TO 24 YEARS GO TO Q85
25 TO 34 YEARS GO TO Q85

35 TO 44 YEARS GO TO Q85
45 TO 54 YEARS GO TO Q85
55 TO 64 YEARS GO TO Q85
65 TO 74 YEARS GO TO Q85
75 YEARS OR OLDER GO TO Q85
DON'T KNOW
REFUSED? GO TO Q85
UNCLEAR RESPONSE

Gender:
Are you male or female?
MALE
FEMALE
DON'T KNOW
REFUSED
UNCLEAR RESPONSE

Education:
What is the highest grade or level of school that you have completed?
8th grade or less
Some high school, but did not graduate
High school graduate or GED
Some college or 2-year degree
4-year college graduate
More than 4-year college degree
DON'T KNOW
REFUSED
UNCLEAR RESPONSE

Respondent lives alone:
How many adults live at your home, including you?
1 [JUST THE RESPONDENT] ? END SURVEY
2 TO 3
4 OR MORE
DON'T KNOW
REFUSED
UNCLEAR RESPONSE

Proxy response and had help completing survey
Did someone help the respondent complete this survey?
1 YES
2 NO

How did that person help? (Mark all that apply)
1 ANSWERED ALL THE QUESTIONS FOR THE RESPONDENT
2 RESTATED THE QUESTIONS IN A DIFFERENT WAY OR REMINDED/PROMPTED THE RESPONDENT
3 TRANSLATED THE QUESTIONS OR ANSWERS INTO THE RESPONDENT's LANGUAGE
4 HELPED WITH THE USEO F ASSISTIVE OR COMMUNICATION EQUIPMENT SO THAT THE RESPONDENT COULD ANSWER QUESTIONS
5 OTHER, SPECIFY

S.15. Detailed risk model specifications *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Provided in response box S.15a

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
The research team used the CAHPS SAS analysis program to produce the scores which allows users to specify case-mix adjusters. For case-mix adjustment specifications, see pages 54-60 of the Instructions for Analyzing Data from CAHPS® Surveys: Using the CAHPS Analysis Program Version 4.1 available from the downloadable zip file at http://www.ahrq.gov/cahps/surveys-guidance/cg/instructions/index.html

S.16. **Type of score:**
Other (specify):
If other: Case-mix adjusted top box scores

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
Scoring specifications for the measures will follow the same general scoring approach as used by other CAHPS surveys that use the CAHPS analysis program. The measures are based on case-mix adjusted top box scores. The research team suggests general health rating, mental health rating, age, gender, whether respondent lives alone, and response option as case- mix adjusters for these measures. We also recommend including survey mode as an additional adjustment variable and proxy status if proxy responses are permitted. The team is also recommending adjusting for Education in future administrations to be consistent with other CAHPS surveys. More information about case-mix adjustment is available in Instructions for Analyzing Data from CAHPS Surveys (available from the downloadable zip file at http://www.ahrq.gov/cahps/surveys-guidance/cg/instructions/index.html).

To create scores for each scale measure:

1. Calculate the score for each item using the top box method.
2. Calculate a mode adjusted score for each item.
3. Calculate case-mix adjusted scores for each program.
4. Calculate means for the scale measures weighting each item equally.

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 from the downloadable zip file at http://www.ahrq.gov/cahps/surveys-guidance/cg/instructions/index.html.

To create scores for each global rating and individual item measure, follow steps 1-3 above.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Defining the Sample Frame: Eligibility Guidelines
The intended sample for the CAHPS Home- and Community-Based Services survey that the measures are based on is adult Medicaid beneficiaries age 18 or older who have received HCBS services for 3 months or longer from the intended survey administration. Sampling should be stratified by HCBS program within each state, in order to allow comparisons of measure results for each HCBS program to the state mean. The source of the sample frame will be the state Medicaid agency or an entity delegated by the state Medicaid agency (e.g., state agency other than the Medicaid agency that operates the program, a MCO, a case management agency, state county, etc.).

Recommended Number of Completed Surveys
In order to determine the size of the sample, each state should take into account the effective sample size and response rates from the field test. The effective sample size is the number of completed responses needed to obtain a reasonable level of

reliability. The research team conducted a pilot test and a field test of the measures with 26 Medicaid HCBS programs across ten states from October 2013 to March 2015. Results suggest that the effective sample size should be 400 people per stratum (with smaller programs including the census). From field test data, we know that the total response rate was 22.0% and this ranged from 9.8% – 31.1% for HCBS programs and modes of administration. Some states may expect a higher response rate in future administrations because of better outreach, pre-survey communications with potential respondents, as well as use of proxies and can adjust their estimated response rate based on these additional considerations.

Proxy Responses

Proxy responses were permitted for the field test of the measures; however, it should be noted that the proxy data were not collected consistently across states and programs. Proxy here is defined as anyone who provided help to the beneficiary completing the survey. We do expect states to allow proxy responses in future data collection efforts. Most immediately, TEFT grantees who are implementing the survey instrument will have the option of allowing respondents to receive assistance or to have a proxy. They will receive information about considerations and possible approaches to incorporating proxies in data collection.[7] It will be their decision whether and how to incorporate proxies.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*

IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.

Survey Administration Mode

Due to the impairments (i.e., cognitive, hearing) prevalent among individuals served by HCBS programs, stakeholders recommend that the survey be conducted through in-person interviews. However, the CAHPS consortium urged the research team to assess both in-person as well as phone administration modes. Based on field test results, administering the survey by phone was found appropriate if a statistical adjustment for survey mode is made for mixed-mode administrations. For programs using the survey measures to monitor trends, we recommend not switching modes across survey fielding periods. A mail survey is not recommended for the HCBS population due to the prevalence of cognitive disabilities

Survey Response Options

Based on findings from cognitive testing as well as an experiment conducted as part of the field test, a simplified response option of Mostly Yes/ Mostly No was determined more accessible for some respondents than the standard CAHPS response option of Never/ Sometimes/ Usually/ Always. For the field test, within each mode (Computer-assisted telephone interviewing and Computer-assisted personal interviewing), equal numbers of participants were randomly assigned to one of the two response option formats—either the 4-point response option or the 2 point binary response option. Participants assigned to the standard response option were switched to the simplified response option if they had difficulty responding using these cognitively more challenging options. "Difficulty" was determined by how well respondents answered the first three survey questions under Getting Needed Services from Personal Assistant and Behavioral Health Staff . If they were unable to answer the questions or had difficulty answering them, the interviewer switched to the alternative format, similar to the CAHPS Nursing Home Long-Stay Resident method.

The interviewer will need to make the determination as to when to use the alternate response option using the following process. If the respondent is unable to respond using the responses "Never, Sometimes, Usually, And Always" as indicated non-verbally or verbally by stating "I don't understand", "I am not sure of the difference" or a similar response, the interviewer should reread the question providing the "Mostly Yes And Mostly No" response option. For the following question, the interviewer should provide the standard responses "Never, Sometimes, Usually, And Always" again, providing the alternate responses of "mostly yes and mostly no" only if the respondent is unable to respond using the standard response. After three unsuccessful attempts to use the standard response, the interviewer should switch to the alternate response and use it throughout the remaining interview.

Including both response modes will allow more respondents to respond to the survey, including individuals with a developmental disability, intellectual/cognitive impairment, or a traumatic brain injury. In cases where both responses are included, the data from the simplified response should be transformed (mostly yes = always and mostly no= never) and pooled with the standard responses for reporting. It is critical to case mix adjust for survey response if both options are offered.

[7] Raetzman SO, Jackson B, Harris S, Frentzel E. Using Proxy Respondents in TEFT Demonstration Round 2 Experience of Care Data Collection. April 21, 2016. Prepared under Centers for Medicare & Medicaid Services Contract HHSM-500-2010-0025I-T006.

Survey Administration
At least one week prior to survey administration, the states should mail a pre-notification letter on state letterhead to all sampled members, alerting them to expect a phone call about the interview and assuring the sampled members that the survey is endorsed by the state. After the pre-notification letters are mailed, the survey vendors should begin telephone contact of HCBS program participants to introduce the survey, explain the survey's purpose, and schedule the interview date and time. To solicit participation, survey vendors should make at least five call attempts to sampled participants during different call days/times—calling in daytime hours during the week, in the evening, and once on the weekend.

Response Rates
The total response rate was 22.0% from the field test and this ranged from 9.8% – 31.1% for the different HCBS programs. Some states may expect a higher response rate in future administrations because of better outreach, upfront communications, and use of proxies.
The research team calculated the response rate using the American Association for Public Opinion Research (AAPOR) response rate #3 (RR#3):
I/((I+P) + (R+NC+O) + e(UH+UO))
Where:
I = complete interviews (3,226)
P = partial interviews (33)
R = refusals and breakoffs (2,442)
NC = noncontact (3,014)
O = other (3,200)
UH = unknown household (3,868)
UO = unknown other (123)
e = estimated proportion of cases of unknown eligibility that are eligible (0.68)

AAPOR defines several options for calculating response rate. Based on the research team's sampling approach, the formula that is most appropriate for these data was RR#3 (http://www.aapor.org/AAPORKentico/Communications/AAPOR-Journals/Standard-Definitions.aspx). The response rate is the total number of completed surveys divided by the total number of eligible sampled individuals. Households with nonworking or wrong numbers are excluded from the denominator. In some cases, eligibility cannot be determined. For these individuals, RR#3 adjusts the response rate assuming that the rate of response for undetermined households would be the same as the response rate where eligibility could be determined. This is shown in the formula where the number of unknowns (UH + UO) is multiplied by the estimated proportion of cases of unknown eligibility that are eligible (e). The result is a slight upward adjustment of the response rate. Thus, the overall response rate was 21.1 percent (22.3 percent in-person and 20.9 percent for phone).

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Missing data are not imputed for unadjusted scores. Measure scores are calculated at the unit level (e.g. HCBS program) using all available data for individual items. Top box scores for individual survey items are computed individually. These are then averaged across items to calculate the scale measure scores. Therefore, a case with usable data for only some individual survey items can be used in the calculation of scale measure scores for a program. However, only "complete" survey responses (those that answered at least half of key items) are included in all measures calculations.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Patient Reported Data/Survey

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
CAHPS Home- and Community-Based Services Survey
In-person and phone
English and Spanish

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
HCBS Program

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Other
If other: Home and Community-Based Services Program

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not applicable.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: CAHPS Home- and Community-Based Services Measures
Date of Submission: 8/1/2016
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☒ Outcome (*including PRO-PM*)
☐ Cost/resource	☐ Process
☐ 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, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** 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 (2b2-2b6) 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 20 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 sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering 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 **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. 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 **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [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]

2b4. 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 sociodemographic 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.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), 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.

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.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒ other: CAHPS Home- and Community-Based Services Survey Data*	☒ other: CAHPS Home- and Community-Based Services Survey Data

*Metrics presented throughout are derived from analysis of the CAHPS Home- and Community-Based Services Survey funded by the Centers for Medicare and Medicaid Services

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? October 2013 – March 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.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☒ other: Medicaid HCBS programs	☒ other: Medicaid HCBS programs

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*)

The measured entity is a Medicaid HCBS program. Medicaid agencies in the states have administrative authority over these HCBS programs and determine which services and supports to offer beneficiaries who are deemed eligible for a given HCBS program. While Medicaid HCBS programs are administered by state Medicaid agencies under various Medicaid legal authorities,[8] they are frequently operated by other entities including non-Medicaid state agencies (e.g., department of aging, etc.), non-state governmental entities (e.g., county, etc.), or managed care organizations. The operating entities then contract with direct service/support providers and case managers. Therefore, the operating entity is the accountable entity for overseeing quality in a Medicaid HCBS program.

When a Medicaid agency delegates operating authority for a Medicaid HCBS program to another entity, federal regulation requires that the Medicaid agency assert its administrative authority by insuring that the operating entity meets quality requirements. The operating entity is required to demonstrate to the Medicaid authority that they have met quality requirements, or in the case of not having met quality requirements, have remediated

[8] 1915(c) Home and Community-Based Services Waivers: 42 CFR §441.301-308, 310. https://www.medicaid.gov/medicaid-chip-program-information/by-topics/long-term-services-and-supports/home-and-community-based-services/home-and-community-based-services-1915-c.html.

1915(i) Optional State Plan Home and Community-Based Services: 42 CFR §441:700-745 https://www.medicaid.gov/medicaid-chip-program-information/by-topics/long-term-services-and-supports/home-and-community-based-services/home-and-community-based-services-1915-i.html.

1915(k) Community First Choice: 42 CFR §441:500-590. https://www.medicaid.gov/medicaid-chip-program-information/by-topics/long-term-services-and-supports/home-and-community-based-services/community-first-choice-1915-k.html.

Managed Long-Term Services and Supports: Sections 1115, 1915(b) and 1932(a) of the Social Security Act.

problems and are engaged in quality improvement activities to address the problem. It is expected that a HCBS program operating entity will use the measures derived from the CAHPS Home- and Community-Based Services Survey as one tool in assessing program quality and in demonstrating to the state Medicaid authority the level of quality in a program, as well as to track improvement over time. If multiple HCBS programs in a state use these measures, the Medicaid agency will have the ability to compare outcomes and quality across HCBS programs in their state.

The measures derived from the CAHPS Home- and Community-Based Services Survey were developed for, and tested on, Medicaid HCBS programs/beneficiaries. However, there are similar non-Medicaid HCBS programs (e.g., state-funded, Older Americans Act-funded) for which the measures may also be appropriate for use. If these programs offer homemaker, personal assistant, and/or case manager services, then the measures may be applicable for use in these programs as well.

The research team conducted a pilot test and a field test of the survey with 26 Medicaid HCBS programs across ten states. The ten states were geographically dispersed and included AZ, CO, CT, GA, KY, LA, MD, MN, NH, and TN; these states (with the exception of TN) were CMS Testing Experience and Functional Tools (TEFT) Demonstration grantees. These 26 HCBS programs serve a wide array of people including people who are elderly with disabilities, individuals with physical disabilities, persons with intellectual/developmental disability, individuals with brain injury, and those with serious mental illness. Combined, these programs served over 138,000 individuals. A random sample of these (n=21,434) HCBS beneficiaries were invited to complete the survey. The complete analytic dataset consists of surveys from 3,223 total respondents. Of these, 3,003 cases were deemed "complete" (over half of all key items were answered) including proxy respondents and were used in the reliability analysis presented here. The number of returned surveys in each program ranges from 0 to 304. One program was not included in analysis because it did not have any returned surveys.

Exhibit 1. States, Populations, Programs, Authorities, and Total Returned Surveys

State	Population Category	HCBS Program	Funding Authority	Number of Total Returned Surveys
Arizona	Elderly/Physically Disabled	Arizona Long Term Care System (ALTCS), Elderly and Physically Disabled expansion	Medicaid 1115 waiver	127
	ID/DD	Arizona Long Term Care System (ALTCS), Developmental Disability	Medicaid 1115 waiver	58
Colorado	Elderly/Physically Disabled	Elderly, Blind, and Disabled Waiver	Medicaid 1915(c) waiver	151
	ID/DD	Supported Living Services Waiver	Medicaid 1915(c) waiver	92
Connecticut	Elderly	Connecticut Home Care Program for Elders	Medicaid 1915(c) waiver	179
	TBI	Acquired Brain Injury Waiver	Medicaid 1915(c) waiver	115
	SMI	Working for Support and Empowerment (WISE) Waiver	Medicaid 1915(c) waiver	81
Georgia	Physically Disabled, TBI	Independent Care Waiver Program	Medicaid 1915(c) waiver	165
	Elderly/Physically Disabled	Community Care Services Program	Medicaid 1915(c) waiver	98
Kentucky	Elderly/Physically Disabled	Home and Community Based Waiver	Medicaid 1915(c) waiver. ADC delivered through HCBS; not state funded.	150

State	Population Category	HCBS Program	Funding Authority	Number of Total Returned Surveys
	ID/DD	Supports for Community Living Waiver	Medicaid 1915(c) waiver	37
	TBI	Acquired Brain Injury Waiver	Medicaid 1915(c) waiver	26
Louisiana	Elderly/Physically Disabled	Adult Day Health Care Waiver	Medicaid 1915(c) waiver	112
	Elderly/Physically Disabled	Community Choices Waiver	Medicaid 1915(c) waiver	302
	Elderly/Physically Disabled	Long Term Personal Care Services Program	Medicaid State plan option	150
	ID/DD	New Opportunities Waiver	Medicaid 1915(c) waiver	146
Maryland	Elderly/Physically Disabled	Community Options Waiver	Medicaid1915(c) waiver	116
	TBI	Traumatic Brain Injury	Medicaid1915(c) waiver	0*
Minnesota	SMI	Personal Care Assistance Program	Medicaid State plan option	155
	Elderly	Elderly Waiver	Medicaid 1915(c) waiver	155
	TBI	Brain Injury Waiver	Medicaid 1915(c) waiver	72
New Hampshire	Elderly/Physically Disabled	Choices for Independence Home and Community Based Care Waiver	Medicaid 1915(c) waiver	147
	ID/DD	Developmental Disabilities Waiver	Medicaid 1915(c) waiver	91
	TBI	Acquired Brain Disorder Waiver	Medicaid 1915(c) waiver	20
	SMI	Bureau of Behavioral Health, Community Mental Health Services	Medicaid State plan, NH general funds, private insurance	174
Tennessee	Elderly/Physically Disabled	TennCare CHOICES in Long-Term Care	Medicaid 1115 waiver	304

*There are 0 completes because of a combined effect of a low number of individuals in the TBI program and the data collection ended before the vendor was able to begin data collection.

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*)

There were 3,003 completed CAHPS Home- and Community-Based Services surveys from 26 Medicaid HCBS programs included in the analysis of the survey data. The breakdown of individuals who completed the survey included:

- 70.2 % in programs serving elderly (age 65+) Medicaid beneficiaries with disabilities, or programs serving working age (age 18-64) Medicaid beneficiaries with physical disabilities (68.4percent without proxies);
- 8.3% served by programs for Medicaid beneficiaries with intellectual or developmental disabilities (12.8 percent without proxies);
- 8.7% enrolled in programs targeting Medicaid beneficiaries with a traumatic brain injury (8.5% without proxies); and
- 13.0% enrolled in Medicaid and receiving services due to a serious mental illness (10.3% without proxies).

Demographics for those completing the survey included:

- Race: White 63.6% (64.2%, without proxies), Black 28.7% (29.5% without proxies), Other Race 7.7% (7.85% without proxies),
- Language: English 90.8% (89.6% without proxies), Spanish 3.6 % (3.8% without proxies), other 5.5% (6.5% without proxies);
- Gender: Male 36.7%, (37.7% without proxies), Female 63.6% (62.3% without proxies);
- Age: 18-24 3.2% (3.2% without proxies), 25-34 7.4% (7.4% without proxies), 35-44 8.8% (8.8% without proxies), 45-54 17.1% (17.1% without proxies), 55-64 22.7% (22.7%), 65-74 19.2% (19.2% without proxies), 75+ 21.6 (21.6% without proxies);,
- Living Arrangement: Lives alone 48.2% (48.2% without proxies), Lives with others 51.8% (51.8% without proxies);
- Metropolitan Statistical Area: Yes 70.8% (70.8% without proxies), No 29.2% (29.2% without proxies).

Other characteristics for those completing the survey included:

- Self-reported general health: Good, Very Good or Excellent 47.6% (50.1% without proxies), Fair or Poor 52.4% (50.0% without proxies),
- Self-reported mental health: Good, Very Good or Excellent 68.3% (66.4% without proxies), Fair or Poor 31.7% (33.6% without proxies).

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. The same data were used for each aspect of testing below.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

The respondent characteristics that were available and evaluated as potential case-mix adjusters included self-reported general health rating, self-reported mental health rating, age, gender, and whether respondent lives alone. We also evaluated the differences in scores by HCBS population. Age, education, and health status are the most common CAHPS variables used in case- mix adjustment. The Medicaid HCBS population, by definition, has low income; therefore, income was not used as a case mix adjuster. The fielded version of the survey did not include an item to assess education; however, this has been added to the final survey and is recommended as a case-mix adjuster moving forward as a condition for the CAHPS trademark designation and to be consistent with other CAHPS surveys. The survey was translated into Spanish, but the number of respondents responding in Spanish (46 respondents) were too few to conduct a language comparison.

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*)

The research team ran a confirmatory factor analysis to test the hypothesized factor structure for the scale measures. The model was an acceptable fit (RMSEA=0.044, CFI=0.954, TLI=0.950).We estimated Cronbach's Alpha values to assess internal consistency reliability, of survey items used in the scale measures. Cronbach's Alpha is a common measure for surveys with scale-type questions. A scale should have an alpha of 0.70 or greater to be considered reliable.[9]

We also looked at HCBS program-level reliability, or inter-unit reliability (IUR). Unit-level reliability indicates the extent to which the experiences of respondents within a unit (e.g., HCBS program) correlate with one another compared to the amount that reported experiences differ among units. As such, it reflects the signal-to-noise ratio; that is, the fraction of total variation due to signal (true variation in scores across units). One of the primary purposes of these measures is to be able to detect difference among HCBS programs, and thus, this ratio is a good indicator of the extent to which the scale measures and other survey items accomplish this goal. It also indicates how reliable a measure is across different respondents. This statistic represents a transformation of the F-statistic for testing differences among programs on a measure (IUR = (F-1)/F). IUR can be interpreted as the fraction of the variation among HCBS program scores that is due to real differences, rather than due to chance. If the IUR is higher, the ability of the item or scale measure to discriminate across programs is greater. Scales with reliability coefficients above 0.70 provide adequate precision for use in statistical analysis of unit-level comparisons.[10] As the IUR gets smaller, a larger sample is needed in order to reliably discriminate across programs. We also calculated the ICC as the between-unit variance minus the within-unit variance over the total variance adjusted for the average number of respondents per reporting unit. The IUR provides the reliability based on the sample size associated with the data while the ICC indicates the reliability of a measure for a single respondent. The IUR 200 values are the projected IUR for a sample of 200. This uses the Spearman-Brown prophecy formula to calculate the projected IUR with a sample of 200 per unit. The Effective Sample Size (ESS) is the average number of usable responses to a particular measure to obtain a target IUR of 0.70.

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*)

Below are Cronbach's Alpha values for scale measures and inter-unit reliability (IUR) statistics for all measures (Exhibit 2). Please reference tab 1.b.2a in the supplementary tables file for item-level IUR statistics for survey items used in the scale measures in Exhibit 2.

Exhibit 2. CAHPS Home- and Community-Based Services Reliability Statistics

Measure	Cronbach's Alpha	IUR	IUR 200	ICC	ESS 70	Measure Response Rate
Scale Measures						
Staff are reliable and helpful	0.74	0.67	0.78	0.0176	142	91.70%
Staff are reliable and helpful	0.79	0.75	0.84	0.0259	95	92.10%
Case manager is helpful	0.55	0.44	0.59	0.0072	374	86.30%
Choosing the services that matter to you	0.27	0.87	0.92	0.057	45	86.40%

[9] Nunnally JC, Bernstein IH (1994). Psychometric Theory. New York: McGraw Hill.

[10] Nunnally, J. C. (1978). *Psychometric theory* (2nd ed). New York: McGraw-Hill .

Transportation to medical appointments	0.45	0.72	0.80	0.0199	116	99.10%
Planning your time and activities	0.50	0.64	0.74	0.0142	162	99.70%
Personal safety and respect	0.15	0.39	0.50	0.0051	460	99.80%
Global Ratings						
Global rating of personal assistance/behavioral health staff	NA	0.67	0.79	0.0188	141	86.40%
Global rating of homemaker	NA	0.65	0.86	0.029	219	35.70%
Global rating of case manager	NA	0.72	0.83	0.0234	115	84.90%
Recommendations Measures						
Recommendation of personal assistance/behavioral health staff	NA	0.74	0.84	0.0263	103	83.90%
Recommendation of homemaker	NA	0.80	0.93	0.059	105	35.50%
Recommendation of case manager	NA	0.71	0.82	0.0225	120	84.50%
Unmet Needs Measures						
Unmet need in dressing/ bathing	NA	0.19	0.87	0.0332	1983	3.40%
Unmet need in meal preparation/ eating		0.50	0.98	0.1858	439	2.30%
Unmet need in medication administration		0.35	0.94	0.079	736	3.70%
Unmet need in toileting		0.22	0.54	0.0058	1149	34.70%
Unmet need with household tasks		0.42	0.96	0.1093	529	3.60%
Physical Safety Measure						
Not hit or hurt by staff		0.22	0.31	0.0022	1042	99.70%

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 Cronbach's Alpha scores range from 0.79 to 0.15, with multiple measures falling below the recommended 0.70 threshold. All of the measures with values below the recommended threshold were deemed critical by the technical expert panel for assessing the quality of a HCBS program.

The IUR values range from 0.87 to 0.19, with multiple measures falling below the 0.70 threshold. This indicates that these measures will need a larger sample size to effectively discriminate among programs which should be attainable in future administrations. However, there are other important goals for using these measures, such as quality improvement for the states, where these measures will still be important.

The IUR 200 values show the projected IUR values for 200 complete surveys for the respective measure. These values show that most measures cross this 0.70 threshold with a sample of 200. The ESS shows the number of usable responses that will be needed in future administrations in order to obtain an IUR of 0.70. Based on these values, the next round of data collection will aim for 400 completed surveys per program and will thus reach the required reliability threshold for the majority of measures.

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. 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)*

Criterion validity refers to the extent to which the HCBS scale measures agree with some criterion of the "true" value of the measure, and can be predictive or concurrent. To evaluate the latter, we estimated correlation coefficients between each global rating measure and each scale measure. If the scale measures have good concurrent validity, then they should have a moderate to strong correlation ($r > 0.30$) with a conceptually related global rating measure. For example, we expect a strong correlation between the *Overall Rating of Case Manager* with the *Case Manager is Helpful* scale measure.

We also examined correlations among the scale measures to determine if they measure different constructs. As these are all measures of beneficiary experience with HCB services, we expect these factors to be related; however, all inter-scale measure correlations should be below 0.80 to indicate that these 7 factors, while related, do not overlap to the point of being redundant.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Exhibit 3. Correlation of Scale Measures and Related Global Rating Measures

Measure	Correlation with Global Rating of Personal Assistance Staff	Correlation with Recommendation of Personal Assistance Staff
Staff are reliable and helpful	0.21	0.35
Staff listen and communicate well	0.24	0.35
Personal safety and respect	0.13	0.17
Measure	**Correlation with Global Rating of Homemaker**	**Correlation with Recommendation of Homemaker**
Staff are reliable and helpful	0.16	0.25
Staff listen and communicate well	0.25	0.29
Personal safety and respect	0.13	0.13
Measure	**Correlation with Global Rating of Case Manager**	**Correlation with Recommendation of Case Manager**
Case manager is helpful	0.24	0.29
Choosing the services that matter to you	0.26	0.25

Exhibit 4. Inter-Scale Correlations

Scale Measures	Staff are reliable and helpful	Staff are reliable and helpful	Case manager is helpful	Choosing the services that matter to you	Transportation to medical appointments	Personal safety and respect	Planning your time and activities
Staff are reliable and helpful	1	-	-	-	-	-	-
Staff listen and communicate well	0.48	1	-	-	-	-	-
Case manager is helpful	0.10	0.11	1	-	-	-	-
Choosing the services that matter to you	0.25	0.26	0.17	1	-	-	-
Transportation to medical appointments	0.21	0.26	0.11	0.14	1	-	-
Personal safety and respect	0.16	0.19	0.13	0.16	0.10	1	-
Planning your time and activities	0.25	0.26	0.12	0.17	0.30	0.12	1

2b2.4. 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?*)

For most measures, the correlations between the scale measures and the related global rating measures were moderate, suggesting that the scale measures are valid measures of beneficiary experience with these providers. The correlation for Personal Safety and Respect was low; however, it should be noted that there was not much variance in the items for this measure.

The scale measures were somewhat correlated with each other as they are all measures of beneficiary experience. However, no values were above 0.80, suggesting that these scales are measuring unique concepts.

2b3. EXCLUSIONS ANALYSIS

NA ☒ no exclusions — *skip to section 2b4*

2b3.1. 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*)

2b3.2. 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*)

2b3.3. 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*)

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with user-selected risk factors***

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

*The CAHPS analysis program was employed as the statistical risk model, and this program allows researchers to select adjustment factors.

2b4.2. If an outcome or resource use measure is <u>not risk adjusted or stratified</u>, provide <u>rationale and analyses</u> to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

Not applicable

2b4.3. Describe the conceptual/clinical <u>and</u> statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

The goals of case-mix adjustment are to help remove the effects of individual respondent characteristics that may affect ratings, remove effects that might be considered spurious (i.e., that reflect something other than quality of care), and remove incentives for providers to avoid "hard-to-treat" individuals. The most common CAHPS case-mix adjusters are age, education, and health status (both general health and emotional/mental health).

Three conditions were required in the selection of variables for case-mix adjustment:

- Within reporting units (HCBS programs), the case-mix variables must be related to the outcome measures (ratings). That is, the variables must have sufficient predictive power in relation to the outcomes (e.g., older respondents give higher ratings of their care). These variables are referred to as "predictors" of the outcome being examined.
- There must be variation between reporting units (HCBS programs) on these predictor variables. That is, the predictors must be unevenly distributed across reporting units (e.g., one program might have a population that tends to be much younger than the population of another program). This condition is the heterogeneity factor of the predictor.
- The case-mix variables must be appropriate for adjustment because they are not themselves determined by the provider's actions. That is, they must be characteristics that are brought to the program by the beneficiary (e.g., age or education), not characteristics that might be consequences of the beneficiary's satisfaction with, or assessment of, the program (e.g., number of visits with a provider). Predictors that are consequences of the beneficiary's satisfaction with the program are endogenous.

We tested the beneficiary characteristics of age, health status (both general health and emotional/mental health), gender, and whether the respondent lived alone as case-mix adjusters. These characteristics typically have the strongest and most consistent associations with patient-reported problems in other CAHPS surveys.[11] We also tested several survey design characteristics – survey mode (in-person vs. phone), response option (standard vs. alternate[12]), proxy status (whether someone completed the survey for the respondent), and assistance with

[11] 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.
[12] The research team opted to have two different response options for many of the survey items: the standard 4-point CAHPS frequency response (never, sometimes, usually, and always) and an alternate binary response (mostly yes and mostly no). This allows

survey (whether someone helped the respondent complete the survey) -- as potential case mix adjusters.[13] The document "Instructions for Analyzing Data from CAHPS Surveys" dated April 2012 (available from the downloadable zip file at http://www.ahrq.gov/cahps/surveys-guidance/cg/instructions/index.html) contains instructions for coding these variables and for including them in analyses using the CAHPS Analysis Program in SAS.

Our analysis for case-mix selection followed four steps:

1. Selection of potential case mix adjusters;
2. Estimation of heterogeneity;
3. Estimation of predictive power of the selected adjusters; and
4. Estimation of the impact of each adjuster.

The research team used stepwise regression to select a subset of the potential case-mix adjusters for further analysis. Stepwise regression analyses evaluated the strength of the relationship of each potential adjuster to ten global rating and scale measures in separate models in which each measure was regressed on all of the potential adjusters. In the stepwise regression models, the potential adjuster variables are added one by one to the model. For a variable to remain in the model, its F-statistic had to be significant at $p < 0.05$. Upon addition of a new variable to the model, each variable already in the model was reassessed, and variables that no longer retained an F-statistic significant at the retention p-level ($p < 0.05$) were excluded from the model. Only after this check was made and the necessary deletions accomplished was another variable added to the model. The stepwise process was complete for a given model when none of the variables outside the model had an F statistic significant at $p < 0.05$ and every variable in the model was statistically significant at $p < 0.05$. Adjuster variables selected in any of the models formed a core set of potential case mix adjusters eligible for final selection.

The research team then estimated the **heterogeneity factor**, **predictive power**, **explanatory power**, and **impact factor** for each potential case-mix variable selected in the regression models. **Heterogeneity of the predictor variables** across programs was measured as the ratio of between-program to within-program variance of the residuals when the variable was regressed on all other potential case-mix adjusters in a random effects model, where the program was included in the model as a random effect. **Heterogeneity of outcome variables** across programs was measured as the ratio of between-program to within-program variance of the residuals when the variable was regressed on program in a random effects model. The research team measured **predictive power** as the incremental amount of variance explained by the predictor (represented as the partial r2 x 1,000) in the stepwise regression analyses, controlling for the other potential case-mix adjusters. To measure **explanatory power**, which considers both the predictive power of each potential adjuster and the heterogeneity of the adjusters across programs, the predictive power was multiplied by the adjuster heterogeneity factor. Finally, the research team calculated the **impact factor**, which standardizes explanatory power with respect to the overall variance in the outcome being assessed as explanatory power/outcome heterogeneity. Variables that had an impact factor >1.0 were considered as candidates for case mix adjusters.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

Results are shown in Exhibits 5-7 below.

---

respondents who can use the 4-point frequency response to do so; for those that cannot, they are still able to participate in the survey using a modified response version. Similarly, based on input from the CAHPS Consortium and Julie Brown, the research team included the two different response scales for the global rating measures.

[13] 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.

Exhibit 5. Parameter Estimates and Selection Status for Variable Selection Models - PCA, Homemaker and Case Manager Global Rating Measures

		Personal Assistance/Behavioral Health Staff Rating		Homemaker Rating		Case Manager Rating	
		Outcome Heterogeneity= 0.006		Outcome Heterogeneity= 0.014		Outcome Heterogeneity= 0.019	
Case-mix Adjustment Variables	*Adjuster Heterogeneity*	*Predictive Power**	*Impact Factor** >1.0*	*Predictive Power**	*Impact Factor* >1.0*	*Predictive Power**	*Impact Factor** >1.0*
General Health	0.044	-	-	-	-	-	-
Mental health	0.049	-	-	-	-	3.40	Yes
Age (18-34)	0.175	2.7	Yes	-	-	-	-
Age (25-34)	0.336	-	-	-	-	-	-
Age (35-44)	0.161	-	-	-	-	-	-
Age (45-54)	0.081	-	-	-	-	-	-
Age (65-74)	0.216	-	-	-	-	2.00	Yes
Age (75+)	0.332	-	-	-	-	2.40	Yes
Gender	0.031	-	-	-	-	-	-
Proxy	0.161	-	-	-	-	-	-
Assistance with survey	0.118	-	-	-	-	-	-
Respondent lives alone	0.042	-	-	-	-	-	-
Survey mode	0.026	-	-	-	-	4.20	Yes
Response option mode	0.039	37.6	Yes	23	Yes	57.70	-

*Predictive power = partial R^2 * 1000)
** Impact factor = (Adjuster Heterogeneity x (R^2 x 1,000)) / (Outcome heterogeneity)
Dashes indicate that the variable was not selected into the stepwise model

Exhibit 6. Parameter Estimates and Selection Status for Variable Selection Models – Getting Needed Care, Communication, and Case Management Scale Measures

		Getting Needed Care		Communication		Case Management	
		Outcome Heterogeneity= 0.029		Outcome Heterogeneity= 0.018		Outcome Heterogeneity= 0.007	
Case-mix Adjustment Variables	*Adjuster Heterogeneity*	*Predictive Power**	*Impact Factor** >1.0*	*Predictive Power**	*Impact Factor** >1.0*	*Predictive Power**	*Impact Factor** >1.0*
General Health	0.044	-	-	-	-	4	Yes
Mental health	0.049	-	-	-	-	-	-
Age (18-34)	0.175	-	-	-	-	-	-
Age (25-34)	0.336	-	-	-	-	-	-
Age (35-44)	0.161	-	-	-	-	-	-
Age (45-54)	0.081	-	-	-	-	-	-
Age (65-74)	0.216	3.2	Yes	-	-	4.3	Yes
Age (75+)	0.332	-	-	-	-	-	-
Gender	0.031	-	-	-	-	-	-
Proxy	0.161	-	-	-	-	-	-
Assistance with survey	0.118	3.3	Yes	-	-	2	Yes
Respondent lives alone	0.042	-	-	-	-	-	-
Survey mode	0.026	-	-	2.3	Yes	6.3	Yes
Response option mode	0.039	17.1	Yes	12.6	Yes	4	Yes

*Predictive power = partial R^2 * 1000)

** Impact factor = (Adjuster Heterogeneity x (R^2 x 1,000)) / (Outcome heterogeneity)
Dashes indicate that the variable was not selected into the stepwise model

Exhibit 7. Parameter Estimates and Selection Status for Variable Selection Models – Choosing Your Services, Transportation, Personal Safety, and Community Inclusion Scale Measure Score Scale Measures

		Choosing Your Services		Transportation		Personal Safety		Community Inclusion	
		Outcome Heterogeneity= 0.035		Outcome Heterogeneity = 0.022		Outcome Heterogeneity= 0.004		Outcome Heterogeneity= 0.010	
Case-mix Adjustment Variables	*Adjuster Heterogeneity*	*Predictive Power**	*Impact Factor** >1.0*	*Predictive Power**	*Impact Factor** >1.0*	*Predictive Power**	*Impact Factor** >1.0*	*Predictive Power**	*Impact Factor** >1.0*
General Health	0.044	-	-	7.7	Yes	5	Yes	6.3	Yes
Mental health	0.049	8.6	Yes	-	-	-		35.4	Yes
Age (18-34)	0.175	-	-	-	-	3.6	Yes	-	-
Age (25-34)	0.336	-	-	-	-	-	-	-	-
Age (35-44)	0.161	-	-	5.3	Yes	-	-	-	-
Age (45-54)	0.081	-	-	-	-	-	-	-	-
Age (65-74)	0.216	-	-	-	-	-	-	-	-
Age (75+)	0.332	-	-	-	-	-	-	-	-
Gender	0.031	-	-	-	-	-	-	-	-
Proxy	0.161	-	-	-	-	-	-	7.4	Yes
Assistance with survey	0.118	-	-	6.2	Yes	1.8	Yes	-	-
Respondent lives alone	0.042	-	-	3.2	Yes	4	Yes	-	-
Survey mode	0.026	-	-	-	-	-	-	-	-
Response option mode	0.039	-	-	15.7	Yes	5	Yes	31	Yes

*Predictive power = partial R^2 * 1000)
** Impact factor = (Adjuster Heterogeneity x (R^2 x 1,000)) / (Outcome heterogeneity)
Dashes indicate that the variable was not selected into the stepwise model

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

See sections 1.8, 2b4.3. and 2b4.4a.

2b4.5. 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*)

See sections 1.8, 2b4.3. and 2b4.4a.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**
Not applicable.

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):
Not applicable.

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
Not applicable.

2b4.9. Results of Risk Stratification Analysis:
Not applicable.

2b4.10. 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*)

Variables that had an impact factor >1.0, and were therefore eligible to be considered as case- mix adjusters, included general health rating, mental health rating, age, whether respondent lives alone, survey administration mode, response option, and proxy status. Gender did qualify for inclusion as a case-mix adjuster when the proxy data were not included but then did not with the additional proxy data included. Therefore, we recommend testing this as an adjuster with future data. If proxies are permitted in future administrations, we recommend including adjustments for both proxy assisted and proxy completed, consistent with other CAHPS survey methodology. We also recommend including education in future administrations to be consistent with other CAHPS surveys.

2b4.11. 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*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

We used t-tests to compare the case-mix adjusted mean scores of each item, scale score, and global rating for each HCBS program within a state to the mean score of all programs combined within the state. A p-value of <0.05 was used to determine whether the scores were statistically significantly different from each other.

2b5.2. 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*)

Exhibit 8 shows counts of programs that were statistically significantly different above or below their state mean for each measure. The exhibit also reports the percentage of programs that were statistically significant in either direction from their state mean.

Exhibit 8. Number and Percentage of Programs with Scores Differing from State Mean

Item #	Measure	Number of Programs Above Respective State Mean	Number of Programs Below Respective State Mean	% of Programs Differing from State Mean
Global Ratings Measures				
35	Global Rating of Personal Assistance/Behavioral Health Staff	6	1	28.0%
46	Global Rating of Homemaker	5	0	27.8%
54	Global Rating of Case Manager	6	3	36.0%
Recommendation Measures				
36	Recommendation of Personal Assistance/Behavioral Health Staff	4	2	24.0%
47	Recommendation of Homemaker	4	2	33.3%
55	Recommendation of Case Manager	5	2	28.0%
Scale Measures				
Staff are reliable and helpful		**5**	**2**	**28.0%**
13	Staff come to work on time	5	1	24.0%
14	Staff work as long as they are supposed to	5	1	24.0%
15	Someone tells you if staff cannot come	7	1	32.0%
19	Staff make sure you have enough privacy for dressing, showering, bathing	4	1	20.0%
37	Homemakers come to work on time	1	2	16.7%
38	Homemakers work as long as they are supposed to	1	1	11.1%
Staff listen and communicate well		**6**	**1**	**28.0%**
28	Staff are nice and polite	6	1	28.0%
29	Staff explanations are easy to understand	7	5	48.0%
30	Staff treat you the way you want them to	7	3	40.0%
31	Staff explain things in a way that is easy to understand	4	0	16.0%
32	Staff listen carefully to you	5	1	24.0%
33	Staff know what kind of help you need with everyday activities	5	2	28.0%
41	Homemakers are nice and polite	2	0	11.1%
42	Homemaker explanations are easy to understand	3	3	33.3%
43	Homemakers treat you the way you want them to	6	1	38.9%
44	Homemakers listen carefully	3	0	16.7%
45	Homemakers know what kind of help you need	3	0	16.7%
Case manager is helpful		**5**	**0**	**20.0%**
49	Able to contact this case manager when needed	4	1	20.0%
51	Case manager helped when asked for help with getting or fixing equipment	5	0	20.0%
53	Case manager helped when asked for help with getting other changes to services	6	0	24.0%
Choosing the services that matter to you		**4**	**3**	**28.0%**
56	Person-centered service plan included all of the things that are important	4	4	32.0%
57	Case manager knows what's on the service plan, including the things that are important	4	1	20.0%
Transportation to medical appointments		**5**	**4**	**36.0%**

59	Always have a way to get to your medical appointments	5	4	36.0%
61	Able to get in and out of this ride easily	4	1	20.0%
62	Ride arrives on time to pick you up	6	3	36.0%
Personal safety and respect		**2**	**0**	**8.0%**
64	Have someone to talk to if someone hurts you or does something to you that you don't like	4	0	16.0%
65	None of the staff take money or things without asking*	7	0	28.0%
68	None of the staff yell, swear, or curse*	6	0	24.0%
Planning your time and activities		**3**	**3**	**24.0%**
75	Can get together with nearby family	4	2	24.0%
77	Can get together with nearby friends	3	4	28.0%
78	Can do things in community	3	3	24.0%
79	Needs more help to do things in community	4	2	24.0%
80	Takes part in deciding what to do with their time	5	0	20.0%
81	Takes part in deciding when they do things each day	9	1	40.0%
Unmet Needs Measures				
18	There are no staff to help dress, shower, or bathe	1	2	18.8%
22	Sufficient staff to help you with meals	2	2	23.5%
25	Sufficient staff to help you with medications	4	2	31.6%
27	Sufficient staff to help you with toileting	8	0	34.8%
40	Sufficient homemakers to help you with household tasks	2	3	26.3%
Physical Safety Measure				
71	Do any staff hit or hurt you	13	0	52.0%

*Programs marked as above or below state means were statistically significantly different at p<.05

2b5.3. 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 findings demonstrate that the measures produce results that adequately discriminate between service recipients' experience of care in their program compared to all programs within a state.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

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

Not applicable.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS 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 SDS 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.

2b6.1. 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*)

2b6.2. 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*)

2b6.3. 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*)

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*)

We conducted a nonresponse bias analysis to evaluate whether respondents and nonrespondents differed significantly. Response bias could be present if there is evidence that the responding population differed in important ways from the population of interest. Our response bias analysis involved comparing respondents to nonrespondents by mode of survey administration, HCBS population, and demographic characteristics using bivariate cross tabulations with chi-square tests (differences were considered statistically significant at $p < 0.05$).

The research team evaluated whether respondents and nonrespondents differed significantly across various characteristics using available data from the sample frame. Complete sample frame data were available only for a subset of the states; therefore, the total number of respondents for the nonresponse bias analysis is fewer than in the psychometric analyses.

2b7.2. 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; <u>if no empirical sensitivity analysis</u>, identify the approaches for handling missing data that were considered and pros and cons of each*)

Respondents and nonrespondents differed significantly by HCBS population, metropolitan statistical area (MSA) residence, state of residence, and guardian status. For example, more respondents were in the disabled (< age 65) group than non-respondents (42 percent vs. 36 percent, respectively); more respondents lived in an MSA than nonrespondents (77 percent vs. 74 percent, respectively); and more nonrespondents reported having a guardian than respondents (10 percent vs. 4 percent, respectively). There were no differences in response by assigned survey administration mode, survey response option, gender, or primary language.

Exhibit 9. Sample Frame Demographic Characteristics Comparison

Characteristics	Nonrespondents n=13,940	Respondents n=1,624	Total (Nonrespondents and Respondents Combined) N=15,564
HCBS Population*			
Aged (65+)	34.0	31.0	33.7
Disabled (<65)	36.4	41.8	36.9
ID/DD	19.0	11.3	18.2
TBI	4.2	6.3	4.4
SMI	6.4	9.6	6.8
Primary Language			
English	97.1	97.7	97.2
Spanish	2.0	1.9	2.0
Other	0.9	0.4	0.8
Metropolitan Statistical Area*			
Yes	74.3	76.5	74.5
No	25.7	23.5	25.5
Gender			
Male	41.9	43.0	42.0
Female	58.2	57.0	58.0
Assigned Survey Response			
Alternate	50.1	49.0	49.9
Standard CAHPS	50.0	51.1	50.1
Assigned Survey Mode			
In-person	80.6	79.2	80.4
Phone	19.4	20.8	19.6
State†*			
AZ	9.4	11.4	9.6
CO	17.7	15.0	17.4
GA	14.1	16.2	14.3
MD	19.2	7.1	18.0
MN	14.5	23.7	15.4

Characteristics	Nonrespondents n=13,940	Respondents n=1,624	Total (Nonrespondents and Respondents Combined) N=15,564
NH	25.2	26.6	25.3
Guardian*			
Yes	10.3	4.0	9.7
No	89.7	96.0	90.4

*Nonrespondents and respondents significantly differ by this characteristics at p <0.05

2b7.3. 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*)

Respondents and nonrespondents did differ by HCBS population, which will be a challenge with future data collection efforts. The team had difficulty reaching desired response rates from beneficiaries with intellectual or developmental disabilities. Future survey administrations may consider allowing proxy assistance with the survey, which will likely increase response rates.

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 beneficiaries

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)*
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.**
It is recommended that the CAHPS Home- and Community-Based Services Survey be administered in-person or by phone. CATI or CAPI data collection is recommended which allow for the creation of electronic databases post data collection.

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.**
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. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

Data Collection:

• Despite a substantial amount of training and an extensive guide provided to survey vendors, all did not follow the data collection instructions exactly. These aspects can be reinforced when reviewing and modifying the materials for future administrations.
• In addition, implementers will need to be thoroughly educated about skip patterns in the EoC survey instrument, applicability of questions to their programs, and how to explain this to data collectors and survey programmers (who will need to take these patterns into effect when analyzing the data). Some of these skip patterns may be adapted to specific states, in which case additional work will be required with survey vendors (e.g., to explain why the skip patterns were adapted and conduct additional review of the field disks to ensure the surveys were appropriately adapted).
• It will be important for states to provide clear specifications about the nature of the work and realistic information about the context in which vendors will need to work. This is especially critical if they decide to use a survey vendor that is not familiar with the data collection instrument or HCBS populations.

Sampling:

• We recommend screening the sample for deceased individuals to the greatest extent possible.

Response Rate:

• Many beneficiaries of Medicaid HCBS programs have guardians from whom consent for the beneficiary's participation in a survey must be secured. For many states, this information is not centrally or readily available, or not updated. Accessing this information prior to contact will help increase participation.
• The AAPOR response rates considers individuals who are deceased or who are physically or mentally unable to respond as eligible respondents resulting in lower response rates. An alternative is to calculate a response rate that does not include such individuals as eligible respondents.
• To avoid alarming potential survey participants and to enhance the recruitment process, any pre-notification letters to the beneficiary should clearly identify the primary survey vendor.
• Programs should employ additional strategies for recruiting challenging populations, including using proxies. Additional outreach can involve case/care managers, or states might enlist advocacy groups to communicate to beneficiaries the importance of participating in the survey.

While field test response rates were less than optimal – 22.0% on average, they ranged from a high of 27.7% for respondents from programs serving the aged/disabled to a low of 9.8% for respondents in programs serving those with intellectual/developmental disabilities (ID/DD). The research team is confident that when states assume responsibility for survey administration the response rate will increase. Our confidence is based on: 1) the use of proxy respondents; 2) enhanced program-specific recruitment efforts; and 3) historical experience of response rates associated with a long-standing survey targeting people with intellectual and developmental disabilities. Taken together, these provide evidence of the feasibility of achieving sample sizes sufficient to discern variation and performance across programs. Each of these reasons for expecting increased response rates in the future is addressed in more detail below.

1. *Use Of Proxy Respondents*: A larger proportion of beneficiaries enrolled in programs serving people with ID/D tend to have guardians. It was our experience in the field test that guardians tended to act as gate keepers, refusing access to the beneficiaries. Many of these guardians said that the beneficiary was not able to complete the survey but that as their guardian they wished to do so. In the beginning of the field test, proxy respondents were not allowed; at the time the CAHPS Consortium did not allow proxy respondents and since CMS was seeking a CAHPS trademark for the survey, the survey team did not allow proxy respondents. However, as field data started coming in, survey vendors began

reporting that interviewers were allowing proxy respondents. Subsequently, the research team decided to allow and document proxy respondents so there would have opportunity to assess their contributions. However, because proxy respondents were allowed only beginning in September 2014 and data collection occurred over the period of July 2014-February 2015, it is not possible to make definitive statements about the effect of proxy respondents on the response rate. That said, there is suggestive evidence that the response rate for the ID/DD population may be increased if proxy respondents are allowed. *Our analyses show that if proxy responses for persons with ID/DD are counted, the number of respondents increases by approximately one hundred percent.*

The table below provides some additional information on proxy use during the field test for all populations, including information pertinent to proxy respondents for surveys targeting people with ID/DD. Proxy respondents were most prevalent in programs serving people with ID/DD. Nearly half of respondents for ID/DD programs were proxy respondents whereas the proxy respondent rate was substantially less in programs serving people with other disabilities.

Proxy Respondents in the EoC Survey Field Test			
Population	**Proxy Complete (N)**	**Percent Proxy Complete Surveys**	**Range of State % Proxy Complete**
Intellectual/developmental disability	192	49.9%	36.1% - 85.7%
Aged and/or disabled	414	20.2%	5.2% -36.6%
Traumatic brain injury	53	20.7%	5.7% - 39.3%
Serious mental illness	8	2.6%	0.0% - 5.1%
Overall	667	22.2%	0.0% - 85.7%

If proxy respondents are allowed in future administrations of the survey one would expect an increase in response rates for the ID/DD population. As noted above, we found that guardians, family members and caregivers acted as gatekeepers, not allowing access to the beneficiary. For example, 14% of all eligible ID/DD sample members had a guardian who did not allow access to the beneficiary because they were either "physically or mentally incompetent" (an AAPOR category for non-response). Converting at least some gate-keeper guardians to proxy respondents in the future should increase response rates substantially.

2. *Enhanced Program-Specific Recruitment Efforts*: Since the conclusion of the field test data collection, the TEFT Demonstration state grantees have identified improvements that they intend to implement for the next round of data collection which they are conducting themselves. It is expected that some of these enhancements will result in improved response rates. They include:
   - Insuring that pre-notification letters originate from the state agency operating the HCBS program being surveyed so that those receiving the letter have familiarity with the letterhead.
   - Ensuring that beneficiary contact information is accurate by requiring that case managers verify beneficiary and guardian contact information for persons sampled.
   - Ensuring that survey vendors have experience and specialized qualifications with the populations being surveyed so they are sensitized to particular considerations in interacting with people with certain types of disability. This is likely to increase rapport and result in improved recruitment.
   - Targeting survey mode to persons/groups more likely to respond to a certain mode (rather than randomization to mode as happened in the field test).
   - Conducting outreach to relevant stakeholders about the survey. This includes case managers and providers so they can encourage beneficiaries to participate when they receive inquiries from sampled members about the legitimacy of the survey. It may also include family and caregiver support groups. Stakeholders are more likely to encourage survey participation if they understand who is sponsoring the survey, its purpose and benefits.
   - Not fielding the survey during the winter holiday season.
   - Not fielding the survey during winter months in colder climates due to the risk of inclement weather prohibiting travel.

3. *Response Rates from An Other Survey of People with ID/DD*: Some of the TEFT Demonstration state grantees sponsor another survey – the National Core Indicators (NCI) survey -- that elicits feedback from people with ID/DD. Some state

ID/DD agencies have conducted the NCI repeatedly over many years. Consequently beneficiaries, family members and guardians are very familiar with the survey. Also, the NCI allows proxy respondents.

Four TEFT grantee states shared the response rates that they have attained in recent years for the adult NCI survey:

- Arizona: 87% response rate
- Kentucky: 94.5% response rate
- Colorado: 39% response rate
- Connecticut: In order to complete 400 surveys, they pull a sample of upwards of 1,000.

In addition to the information provided by the TEFT states, the National Association for Directors of Developmental Disability Services (NASDDDS), one of the sponsors of the NCI survey for people with ID/DD, reports that for their face-to-face surveys: *"Most states interview about 500 people to get the 400 sample size number (and most have to pull about 800 names to get the sample size)."* (http://www.nasddds.org/uploads/files/NCI_Description_and_Costs.pdf). While the NCI project does not report average response rates, this information from the NASDDDS website indicates the feasibility of achieving much higher response rates for the ID/DD subgroup than realized in the Experience of Care survey field test.

Timing of Data Collection:
• States that experience snow/ice during the winter should be encouraged to schedule data collection in other seasons.

Administration time:
Prospectively, for the PRA package submitted to OMB, the research team estimated 30 minutes per administration. This estimate was based on the length of the survey and CMS's experience with previous CAHPS surveys of comparable length that were fielded with a similar, although not identical, population – the long-stay nursing home resident CAHPS. The nursing home CAHPS reported an administration time of 20 minutes on average. Since there are more items in the CAHPS Home and Community-Based Services Survey than the Nursing Home CAHPS (96 items versus 45 items), the research team estimated 30 minutes for the PRA package.

Retrospective analysis post field test showed that on average, respondents answered 51 (out of 96 items). Skip patterns account for the majority of the discrepancy between total number of items and number answered. General guidance from a survey expert is that it takes a person can answer approximately 4 items per minute. This estimate would put the time to administer the HCBS CAHPS in the vicinity of 13 minutes. Although field test data are not available on length of time to administer, we can safely assume that survey administration, on average, did not exceed 30 minutes based on these other sources.

Routine use of satisfaction/experience surveys in HCBS programs:
State Medicaid programs have been administering surveys to their HCBS beneficiaries for many years. Many are home-grown and state-specific, others were developed with an intended wider audience; only some of these surveys have undergone limited testing. Beneficiaries and their caregivers are very accustomed to these surveys. We know that at least 46 states currently administer a satisfaction/experience survey to their beneficiaries in at least one HCBS program. However, these surveys are all disability-specific, unlike the CAHPS HCBS Survey which was designed and tested as a cross-disability survey.

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 final CAHPS Home- and Community-Based Services survey will be available to state Medicaid Agencies for use free of charge. In addition to the survey instrument, users will have access to comprehensive materials supporting fielding, analysis, and reporting as well as CAHPS Analysis Program that performs analysis and significance testing.

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.

Planned	Current Use (for current use provide URL)
Public Reporting Quality Improvement (Internal to the specific organization)	

4a.1. For each CURRENT use, checked above, provide:

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

4a.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?*)

Testing of the survey and measures was recently completed. Plans for voluntary use by HCBS programs are underway.

4a.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.*)

The measures in this submission derive from the CAHPS Home and Community-Based Services Survey. The survey was developed with CAHPS principles and was recently approved by AHRQ and the CAHPS Consortium for a CAHPS trademark. The survey will be released publicly. Because the survey was developed for voluntary use in Medicaid HCBS programs, it is expected that many state Medicaid programs will begin using the survey within the next few years. Thus, it is expected that the measures derived from the survey will likely be used by HCBS programs for their internal assessment of program quality and related quality improvement projects, as well as for public reporting. It is also possible that some measures may be considered as metrics in value based purchasing initiatives most typically associated with state Medicaid managed long-term services and supports. However, the survey and related measure use in state HCBS programs will be voluntary; at this time CMS has no plans to use the measures for national public reporting.

4b. 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.

4b.1. Progress on Improvement. (Not required for initial endorsement unless available.)

Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

Not applicable.

4b.2. If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

See 4a.3.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**

There were no unintended consequences identified.

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

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 completely harmonized?

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

Not applicable.

5b. Competing Measures

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

OR

Multiple measures are justified.

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

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

Not applicable.

Appendix

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

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare and Medicaid Services
Co.2 **Point of Contact:** Kerry, Lida, Kerry.Lida@cms.hhs.gov, 410-786-4826-
Co.3 **Measure Developer if different from Measure Steward:** Truven Health Analytics
Co.4 **Point of Contact:** Beth, Jackson, Beth.Jackson@truvenhealth.com, 508-520-1507-

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.
See section 1c.5. for list of technical expert panel members.
The research team involved in the development of the measures includes:
Centers for Medicare & Medicaid Services
Kerry Lida, Ph.D. kerry.lida@cms.hhs.gov
Michael R. Smith, MPA michael.smith1@cms.hhs.gov
Other Investigators
Beth Jackson, Ph.D., Truven Health Analytics
Susan Raetzman, M.S.P.H., Truven Health Analytics
Elizabeth Frentzel, M.P.H., American Institutes for Research
Coretta Mallery, Ph.D., American Institutes for Research
Chris Pugliese, M.P.P., American Institutes for Research
Lee Hargraves, Ph.D., American Institutes for Research
Tandrea Hilliard, Ph.D., American Institutes for Research
Chris Evensen, M.A., American Institutes for Research
Steven Garfinkel, Ph.D., American Institutes for Research

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

Ad.6 **Copyright statement:**
Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:** CMS is in the process of renewing their Measure Steward Agreement and received approval from NQF to submit these measures while this is in process.

CAHPS® Home- and Community-Based Services Survey

Version: 1.0

Population: Adult

Language: English

File name: CAHPS HCBS Survey English (8Jul2016) for CoP
Last updated: July 8, 2016

Instructions for Vendor

- The interview is intended as an interviewer-administered survey; thus all text that appears in initial uppercase and lowercase letters should be read aloud. Text that appears in **bold, lowercase letters** should be emphasized.

- Text in {*italics and in braces*} will be provided by the HCBS program's administrative data. However, if the interviewee provides another term, that term should be used in place of the program-specific term wherever indicated. For example, some interviewees may refer to their case manager by another title, which should be used instead throughout the survey.

- For response options of "never," "sometimes," "usually," and "always," if the respondent cannot use that scale, the alternate version of the survey with response options of "mostly yes" and "mostly no" should be used. These alternate response options are reserved for respondents who find the "never," "sometimes," "usually," "always" response scale cognitively challenging.

- For response options of 0 to 10, if the respondent cannot use that scale, the alternate version of the survey with response options of "excellent," "very good," "good," "fair," or "poor" should be used. These alternate response options are reserved for respondents who find the numeric scale cognitively challenging.

- All questions include a "REFUSED" response option. In this case, "refused" means the respondent did not provide any answer to the question.

- All questions include a "DON'T KNOW" response option. This is used when the respondent indicates that he or she does not know the answer and cannot provide a response to the question.

- All questions include an "UNCLEAR" response option. This should be used when a respondent answers, but the interviewer cannot clarify the meaning of the response even after minor probing **or** the response is completely unrelated to the question, (e.g., the response to "In the last 3 months, how often did your homemakers listen carefully to what you say?" is "I like to sit by Mary").

- Some responses have skip patterns, which are expressed as "→ GO TO Q#." The interviewer should be advanced to the next appropriate item to ask the respondent.

- Not all respondents receive all home and community-based services asked about in this instrument. Items Q4 through Q12 help to confirm which services a respondent receives. The table after it summarizes the logic of which items should be used.

- Survey users may add questions to this survey before the "About You" section. A separate supplemental employment module can be added.

- Use singular/plural as needed. In most cases, questions are written assuming there is more than one staff person supporting a respondent or it is written without an indication of whether there is more than one staff person. Based on information collected from Q4 through Q12, it is possible to modify questions to be singular or plural as they relate to staff.

- Use program-specific terms. Where appropriate, add in the program-specific terms for staff (e.g., [*program-specific term for these types of staff*]) but allow the interviewer to modify the term based on the respondent's choice of the word. It will be necessary to obtain information for program-specific terms. State administrative data should include the following information:
  - Agency name(s)
  - Titles of staff who provide care
  - Names of staff who provide care
  - Activities that each staff member provides (this will help with identifying appropriate skip logic)
  - Hours of staff who come to the home

COGNITIVE SCREENING QUESTIONS

People might be paid to help you get ready in the morning, with housework, go places, or get mental health services. This survey is about the people who are paid to help you in your home and community with everyday activities. It also asks about the services you get.

1. Does someone come into your home to help you?

[1] ☐ YES
[2] ☐ NO → END SURVEY
[-1] ☐ DON'T KNOW → END SURVEY
[-2] ☐ REFUSED → END SURVEY
[-3] ☐ UNCLEAR RESPONSE → END SURVEY

2. How do they help you?

_____

[EXAMPLES OF CORRECT RESPONSES INCLUDE]

- HELPS ME GET READY EVERY DAY
- CLEANS MY HOME
- WORKS WITH ME AT MY JOB
- HELPS ME DO THINGS
- DRIVES ME AROUND

[-1] ☐ DON'T KNOW → GO TO THE ABOUT YOU SECTION
[-2] ☐ REFUSED → GO TO THE ABOUT YOU SECTION
[-3] ☐ UNCLEAR RESPONSE → GO TO THE ABOUT YOU SECTION

3. What do you call them?

_____

[EXAMPLES OF SUFFICIENT RESPONSES INCLUDE]

- MY WORKER
- MY ASSISTANT
- NAMES OF STAFF (JO, DAWN, ETC.)

[-1] ☐ DON'T KNOW → GO TO THE ABOUT YOU SECTION
[-2] ☐ REFUSED → GO TO THE ABOUT YOU SECTION
[-3] ☐ UNCLEAR RESPONSE → GO TO THE ABOUT YOU SECTION

CSQPASS.

[IF ALL 3 QUESTIONS WERE ANSWERED CORRECTLY, ENTER 1 TO CONTINUE.]
1 PASS - ALL 3 QUESTIONS WERE ANSWERED CORRECTLY → GO TO Q4
2 FAIL - AT LEAST 1 QUESTION WAS NOT ANSWERED CORRECTLY → GO TO SURVEND

SURVEND.
Thank you for your time. Those are all the questions we have.
Have a nice day/evening. [ENTER 1 TO EXIT SURVEY]

IDENTIFICATION QUESTIONS

Now I would like to ask you some more questions about the types of people who come to your home.

4. In the last 3 months, did you get {*program specific term for personal assistance*} at home?

1☐ YES
2☐ NO → GO TO Q6
-1☐ DON'T KNOW → GO TO Q6
-2☐ REFUSED → GO TO Q6
-3☐ UNCLEAR RESPONSE → GO TO Q6

5. What do you call the person or people who gave you {*program-specific term for personal assistance*}? For example, do you call them {*program-specific term for personal assistance*}, staff, personal care attendants, PCAs, workers, or something else?

_____

[ADD RESPONSE WHEREVER IT SAYS "*personal assistance/behavioral health staff"*]

6. In the last 3 months, did you get {*program specific term for behavioral health specialist services*} at home?

1☐ YES
2☐ NO → GO TO Q8
-1☐ DON'T KNOW → GO TO Q8
-2☐ REFUSED → GO TO Q8
-3☐ UNCLEAR RESPONSE → GO TO Q8

7. What do you call the person or people who gave you {*program specific term for behavioral health specialist services*}? For example, do you call them {*program-specific term for behavioral health specialists*}, counselors, peer supports, recovery assistants, or something else?

_____

[ADD RESPONSE WHEREVER IT SAYS "*personal assistance/behavioral health staff."* IF Q4 ALSO = YES, LIST BOTH TITLES]

8. In the last 3 months, did you get {*program specific term for homemaker services*} at home?

[1] ☐ YES
[2] ☐ NO → GO TO Q11
[-1] ☐ DON'T KNOW → GO TO Q11
[-2] ☐ REFUSED → GO TO Q11
[-3] ☐ UNCLEAR RESPONSE → GO TO Q11

9. What do you call the person or people who gave you {*program specific term for homemaker services*}? For example, do you call them {*program-specific term for homemaker*}, aides, homemakers, chore workers, or something else?

_____

[ADD RESPONSE WHEREVER IT SAYS "*homemaker"*]

10. [IF (Q4 *OR* Q6) *AND* Q8 = YES, ASK] In the last 3 months, did the same people who help you with everyday activities also help you clean your home?

[1] ☐ YES
[2] ☐ NO
[-1] ☐ DON'T KNOW
[-2] ☐ REFUSED
[-3] ☐ UNCLEAR RESPONSE

11. In the last 3 months, did you get help from {*program specific term for case manager services}* to help make sure that you had all the services you needed?

[1] ☐ YES
[2] ☐ NO
[-1] ☐ DON'T KNOW
[-2] ☐ REFUSED
[-3] ☐ UNCLEAR RESPONSE

12. What do you call the person who gave you {*program specific term for case manager services*}? For example, do you call the person a {*program-specific term for case manager*}, case manager, care manager, service coordinator, supports coordinator, social worker, or something else?

_____

[ADD RESPONSE WHEREVER IT SAYS "*case manager"*]

BELOW ARE INSTRUCTIONS FOR WHICH QUESTIONS TO ASK FOR EACH RESPONSE ABOVE.

ITEM AND RESPONSE—FOLLOW ALL ROWS THAT APPLY	**ACTION**
IF Q4 OR Q6 = YES (PERSONAL ASSISTANCE OR BEHAVIORAL HEALTH SPECIALIST SERVICES), AND Q8 = NO, DON'T KNOW, REFUSE, UNCLEAR (HOMEMAKER SERVICES)	ASK Q13–Q36, AND Q48 ONWARD
IF Q4 OR Q6 = YES (PERSONAL ASSISTANCE OR BEHAVIORAL HEALTH SPECIALIST SERVICES), AND Q8 = YES (HOMEMAKER SERVICES)	ASK Q13 ONWARD
IF Q4 AND Q6 = NO (PERSONAL ASSISTANCE OR BEHAVIORAL HEALTH SPECIALIST SERVICES)	SKIP Q13–36, Q57 AND Q79
IF Q8 = YES (HOMEMAKER SERVICES)	ASK Q37 ONWARD
IF Q10 = YES (HOMEMAKER AND PERSONAL ASSISTANCE STAFF SAME)	ASK Q13–Q36, Q39, Q40, AND Q48 ONWARD
IF Q11 = ANY RESPONSE (CASE MANAGER)	ASK Q48 ONWARD

GETTING NEEDED SERVICES FROM PERSONAL ASSISTANT AND BEHAVIORAL HEALTH STAFF

13. First I would like to talk about the {*personal assistance/behavioral health staff*} who are paid to help you with everyday activities—for example, getting dressed, using the bathroom, taking a bath or shower, or going places. In the last 3 months, how often did {personal assistance/behavioral health staff} come to work on time? Would you say . . .

[1] ☐ Never,
[2] ☐ Sometimes,
[3] ☐ Usually, or

4☐ Always?
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

ALTERNATE VERSION: First I would like to talk about the {*personal assistance/behavioral health staff*} who are paid to help you with everyday activities—for example, getting dressed, using the bathroom, taking a bath or shower, or going places. In the last 3 months, did {*personal assistance/behavioral health staff*} come to work on time? Would you say. . .

1☐ Mostly yes or
2☐ Mostly no?
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

14. In the last 3 months, how often did {personal assistance/behavioral health staff} work as long as they were supposed to? Would you say. . .

1☐ Never,
2☐ Sometimes,
3☐ Usually, or
4☐ Always?
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

ALTERNATE VERSION: In the last 3 months, did {*personal assistance/behavioral health staff*} work as long as they were supposed to? Would you say . . .

1☐ Mostly yes or
2☐ Mostly no?
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

15. Sometimes staff cannot come to work on a day that they are scheduled. In the last 3 months, when staff could not come to work on a day that they were scheduled, did someone let you know that {*personal assistance/behavioral health staff*} could not come that day?

1☐ YES
2☐ NO
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

16. In the last 3 months, did you need help from {*personal assistance/behavioral health staff*} to get dressed, take a shower, or bathe?

[1] ☐ YES
[2] ☐ NO → GO TO Q20
[-1] ☐ DON'T KNOW → GO TO Q20
[-2] ☐ REFUSED → GO TO Q20
[-3] ☐ UNCLEAR RESPONSE → GO TO Q20

17. In the last 3 months, did you **always** get dressed, take a shower, or bathe when you needed to?

[1] ☐ YES → GO TO Q19
[2] ☐ NO
[-1] ☐ DON'T KNOW → GO TO Q19
[-2] ☐ REFUSED → GO TO Q19
[-3] ☐ UNCLEAR RESPONSE → GO TO Q19

18. In the last 3 months, was this because there were no {*personal assistance/behavioral health staff*} to help you?

[1] ☐ YES
[2] ☐ NO
[-1] ☐ DON'T KNOW
[-2] ☐ REFUSED
[-3] ☐ UNCLEAR RESPONSE

19. In the last 3 months, how often did {*personal assistance/behavioral health staff*} make sure you had enough personal privacy when you dressed, took a shower, or bathed? Would you say. . .

[1] ☐ Never,
[2] ☐ Sometimes,
[-3] ☐ Usually, or
[-3] ☐ Always?
[-1] ☐ DON'T KNOW
[-2] ☐ REFUSED
[-3] ☐ UNCLEAR RESPONSE

ALTERNATE VERSION: In the last 3 months, did {*personal assistance/behavioral health staff*} make sure you had enough personal privacy when you dressed, took a shower, or bathed? Would you say. . .

[1] ☐ Mostly yes or
[2] ☐ Mostly no?
[-1] ☐ DON'T KNOW
[-2] ☐ REFUSED
[-3] ☐ UNCLEAR RESPONSE

20. In the last 3 months, did you need help from {*personal assistance/behavioral health staff*} with your meals, such as help making or cooking meals or help eating?

[1] ☐ YES

2☐ NO → GO TO Q23
-1☐ DON'T KNOW → GO TO Q23
-2☐ REFUSED → GO TO Q23
-3☐ UNCLEAR RESPONSE → GO TO Q23

21. In the last 3 months, were you **always** able to get something to eat when you were hungry?

1☐ YES → GO TO Q23
2☐ NO
-1☐ DON'T KNOW → GO TO Q23
-2☐ REFUSED → GO TO Q23
-3☐ UNCLEAR RESPONSE → GO TO Q23

22. In the last 3 months, was this because there were no {*personal assistance/behavioral health staff*} to help you?

1☐ YES
2☐ NO
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

23. Sometimes people need help taking their medicines, such as reminders to take a medicine, help pouring them, or setting up their pills. In the last 3 months, did you need help from {*personal assistance/behavioral health staff*} to take your medicines?

1☐ YES
2☐ NO → GO TO Q26
-1☐ DON'T KNOW → GO TO Q26
-2☐ REFUSED → GO TO Q26
-3☐ UNCLEAR RESPONSE → GO TO Q26

24. In the last 3 months, did you **always** take your medicine when you were supposed to?

1☐ YES → GO TO Q26
2☐ NO
-1☐ DON'T KNOW → GO TO Q26
-2☐ REFUSED → GO TO Q26
-3☐ UNCLEAR RESPONSE → GO TO Q26

25. In the last 3 months, was this because there were no {*personal assistance/behavioral health staff*} to help you?

1☐ YES
2☐ NO
-1☐ DON'T KNOW
-2☐ REFUSED

-3 ☐ UNCLEAR RESPONSE

26. Help with toileting includes helping someone get on and off the toilet or help changing disposable briefs or pads. In the last 3 months, did you need help from {*personal assistance/behavioral health staff*} with toileting?

1 ☐ YES
2 ☐ NO → GO TO Q28
-1 ☐ DON'T KNOW → GO TO Q28
-2 ☐ REFUSED → GO TO Q28
-3 ☐ UNCLEAR RESPONSE → GO TO Q28

27. In the last 3 months, did you get all the help you needed with toileting from {personal assistance/behavioral health staff} when you needed it?

1 ☐ YES
2 ☐ NO
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

HOW WELL PERSONAL ASSISTANT AND BEHAVIORAL HEALTH STAFF COMMUNICATE WITH AND TREAT YOU

The next several questions ask about how {personal assistance/behavioral health staff} treat you.

28. In the last 3 months, how often did {personal assistance/behavioral health staff} treat you with courtesy and respect? Would you say . . .

1 ☐ Never,
2 ☐ Sometimes,
3 ☐ Usually, or
4 ☐ Always?
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

ALTERNATE VERSION: In the last 3 months, did {*personal assistance/behavioral health staff*} treat you with courtesy and respect? Would you say . . .

1 ☐ Mostly yes or
2 ☐ Mostly no?
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

29. In the last 3 months, how often were the explanations {*personal assistance/behavioral health staff*} gave you hard to understand because of an accent or the way {*personal assistance/behavioral health staff}* spoke English? Would you say …

[1] ☐ Never,
[2] ☐ Sometimes,
[3] ☐ Usually, or
[4] ☐ Always?
[-1] ☐ DON'T KNOW
[-2] ☐ REFUSED
[-3] ☐ UNCLEAR RESPONSE

ALTERNATE VERSION: In the last 3 months, were the explanations {*personal assistance/behavioral health staff*} gave you hard to understand because of an accent or the way {*personal assistance/behavioral health staff*} spoke English? Would you say. . .

[1] ☐ Mostly yes or
[2] ☐ Mostly no?
[-1] ☐ DON'T KNOW
[-2] ☐ REFUSED
[-3] ☐ UNCLEAR RESPONSE

30. In the last 3 months, how often did {*personal assistance/behavioral health staff*} treat you the way you wanted them to? Would you say . . .

[1] ☐ Never,
[2] ☐ Sometimes,
[3] ☐ Usually, or
[4] ☐ Always?
[-1] ☐ DON'T KNOW
[-2] ☐ REFUSED
[-3] ☐ UNCLEAR RESPONSE

ALTERNATE VERSION: In the last 3 months, did {*personal assistance/behavioral health staff*} treat you the way you wanted them to? Would you say . . .

[1] ☐ Mostly yes or
[2] ☐ Mostly no?
[-1] ☐ DON'T KNOW
[-2] ☐ REFUSED
[-3] ☐ UNCLEAR RESPONSE

31. In the last 3 months, how often did {*personal assistance/behavioral health staff*} explain things in a way that was easy to understand? Would you say . . .

[1] ☐ Never,
[2] ☐ Sometimes,
[3] ☐ Usually, or

4☐ Always?
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

ALTERNATE VERSION: In the last 3 months, did {*personal assistance/behavioral health staff*} explain things in a way that was easy to understand? Would you say . . .

1☐ Mostly yes or
2☐ Mostly no?
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

32. In the last 3 months, how often did {*personal assistance/behavioral health staff*} listen carefully to you? Would you say . . .

1☐ Never,
2☐ Sometimes,
3☐ Usually, or
4☐ Always?
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

ALTERNATE VERSION: In the last 3 months, did {*personal assistance/behavioral health staff*} listen carefully to you?
Would you say . . .

1☐ Mostly yes or
2☐ Mostly no?
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

33. In the last 3 months, did you feel {*personal assistance/behavioral health staff*} knew what kind of help **you** needed with everyday activities, like getting ready in the morning, getting groceries, or going places in your community?

1☐ YES
2☐ NO
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

34. In the last 3 months, did *{personal assistance/behavioral health staff}* encourage you to do things for yourself if you could?

1☐ YES

2 ☐ NO
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

35. Using any number from 0 to 10, where 0 is the worst help from {*personal assistance/behavioral health staff*} possible and 10 is the best help from {*personal assistance/behavioral health staff*} possible, what number would you use to rate the help you get from {*personal assistance/behavioral health staff*}?

__0 TO 10
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

ALTERNATE VERSION: How would you rate the help you get from {*personal assistance/behavioral health staff*}? Would you say . . .

1 ☐ Excellent,
2 ☐ Very good,
3 ☐ Good,
4 ☐ Fair, or
5 ☐ Poor?
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

36. Would you recommend the {*personal assistance/behavioral health staff}* who help you to your family and friends if they needed help with everyday activities? Would you say you would recommend the {*personal assistance/behavioral health staff}* . . .

1 ☐ Definitely no,
2 ☐ Probably no,
3 ☐ Probably yes, or
4 ☐ Definitely yes?
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

GETTING NEEDED SERVICES FROM HOMEMAKERS

The next several questions are about the {*homemakers*}, the staff who are paid to help you do tasks around the home—such as cleaning, grocery shopping, or doing laundry.

37. In the last 3 months, how often did {*homemakers*} come to work on time? Would you say . . .

1 ☐ Never,
2 ☐ Sometimes,

3☐ Usually, or
4☐ Always?
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

ALTERNATE VERSION: In the last 3 months, did {*homemakers*} come to work on time? Would you say . . .

1☐ Mostly yes or
2☐ Mostly no?
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

38. In the last 3 months, how often did {*homemakers*} work as long as they were supposed to? Would you say . . .

1☐ Never,
2☐ Sometimes,
3☐ Usually, or
4☐ Always?
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

ALTERNATE VERSION: In the last 3 months, did {*homemakers*} work as long as they were supposed to? Would you say . . .

1☐ Mostly yes or
2☐ Mostly no?
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

39. In the last 3 months, did your household tasks, like cleaning and laundry, **always** get done when you needed them to? [ASK IF HOMEMAKER IS THE SAME AS PCA STAFF]

1☐ YES → GO TO Q41
2☐ NO
-1☐ DON'T KNOW → GO TO Q41
-2☐ REFUSED → GO TO Q41
-3☐ UNCLEAR RESPONSE → GO TO Q41

40. In the last 3 months, was this because there were no {*homemakers*} to help you? [ASK IF HOMEMAKER IS THE SAME AS PCA STAFF]

1☐ YES
2☐ NO
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

HOW WELL HOMEMAKERS COMMUNICATE WITH AND TREAT YOU

The next several questions ask about how {homemakers} treat you.

41. In the last 3 months, how often did {*homemakers*} treat you with courtesy and respect? Would you say . . .

1☐ Never,
2☐ Sometimes,
3☐ Usually, or
4☐ Always?
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

ALTERNATE VERSION: In the last 3 months, did {*homemakers*} treat you with courtesy and respect? Would you say . . .

1☐ Mostly yes or
2☐ Mostly no?
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

42. In the last 3 months, how often were the explanations {*homemakers*} gave you hard to understand because of an accent or the way the {*homemakers}* spoke English? Would you say . . .

1☐ Never,
2☐ Sometimes,
3☐ Usually, or
4☐ Always?
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

ALTERNATE VERSION: In the last 3 months, were the explanations {*homemakers*} gave you hard to understand because of an accent or the way {*homemakers*} spoke English? Would you say. . .

1☐ Mostly yes or

[2] ☐ Mostly no?
[-1] ☐ DON'T KNOW
[-2] ☐ REFUSED
[-3] ☐ UNCLEAR RESPONSE

43. In the last 3 months, how often did {*homemakers*} treat you the way you wanted them to? Would you say . . .

[1] ☐ Never,
[2] ☐ Sometimes,
[3] ☐ Usually, or
[4] ☐ Always?
[-1] ☐ DON'T KNOW
[-2] ☐ REFUSED
[-3] ☐ UNCLEAR RESPONSE

ALTERNATE VERSION: In the last 3 months, did {*homemakers*} treat you the way you wanted them to? Would you say . . .

[1] ☐ Mostly yes or
[2] ☐ Mostly no?
[-1] ☐ DON'T KNOW
[-2] ☐ REFUSED
[-3] ☐ UNCLEAR RESPONSE

44. In the last 3 months, how often did {*homemakers*} listen carefully to you? Would you say . . .

[1] ☐ Never,
[2] ☐ Sometimes,
[3] ☐ Usually, or
[4] ☐ Always?
[-1] ☐ DON'T KNOW
[-2] ☐ REFUSED
[-3] ☐ UNCLEAR RESPONSE

ALTERNATE VERSION: In the last 3 months, did {*homemakers*} listen carefully to you? Would you say . . .

[1] ☐ Mostly yes or
[2] ☐ Mostly no?
[-1] ☐ DON'T KNOW
[-2] ☐ REFUSED
[-3] ☐ UNCLEAR RESPONSE

45. In the last 3 months, did you feel {*homemakers*} knew what kind of help you needed?

[1] ☐ YES

2☐ NO
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

46. Using any number from 0 to 10, where 0 is the worst help from {*homemakers*} possible and 10 is the best help from {*homemakers*} possible, what number would you use to rate the help you get from {*homemakers*}?

__0 TO 10
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

ALTERNATE VERSION: How would you rate the help you get from {*homemakers*}? Would you say . . .

1☐ Excellent,
2☐ Very good,
3☐ Good,
4☐ Fair, or
5☐ Poor?
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

47. Would you recommend the {*homemakers}* who help you to your family and friends if they needed {*program-specific term for homemaker services*}? Would you say you would recommend the {*homemakers*} . . .

1☐ Definitely no,
2☐ Probably no,
3☐ Probably yes, or
4☐ Definitely yes?
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

YOUR CASE MANAGER

Now I would like to talk to you about your {*case manager*}, the person who helps make sure you have the services you need.

48. Do you know who your {*case manager*} is?

1☐ YES
2☐ NO → GO TO Q56
-1☐ DON'T KNOW → GO TO Q56

-2 ☐ REFUSED → GO TO Q56
-3 ☐ UNCLEAR RESPONSE → GO TO Q56

49. In the last 3 months, could you contact this {*case manager*} when you needed to?

1 ☐ YES
2 ☐ NO
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

50. Some people need to get equipment to help them, like wheelchairs or walkers, and other people need their equipment replaced or fixed. In the last 3 months, did you ask this {*case manager*} for help with getting or fixing equipment?

1 ☐ YES
2 ☐ NO → GO TO Q52
3 ☐ DON'T NEED → GO TO Q52
-1 ☐ DON'T KNOW → GO TO Q52
-2 ☐ REFUSED → GO TO Q52
-3 ☐ UNCLEAR RESPONSE → GO TO Q52

51. In the last 3 months, did this {*case manager*} work with you when you asked for help with getting or fixing equipment?

1 ☐ YES
2 ☐ NO
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

52. In the last 3 months, did you ask this {*case manager*} for help in getting any changes to your services, such as more help from {*personal assistance/behavioral health staff and/or homemakers if applicable*}, or for help with getting places or finding a job?

1 ☐ YES
2 ☐ NO → GO TO 54
3 ☐ DON'T NEED → GO TO Q54
-1 ☐ DON'T KNOW → GO TO Q54
-2 ☐ REFUSED → GO TO Q54
-3 ☐ UNCLEAR RESPONSE → GO TO Q54

53. In the last 3 months, did this {*case manager*} work with you when you asked for help with getting other changes to your services?

1 ☐ YES
2 ☐ NO

-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

54. Using any number from 0 to 10, where 0 is the worst help from {*case manager*} possible and 10 is the best help from {*case manager*} possible, what number would you use to rate the help you get from {*case manager*}?

__0 TO 10

-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

ALTERNATE VERSION: How would you rate the help you get from the {*case manager*}? Would you say . . .

1 ☐ Excellent,
2 ☐ Very good,
3 ☐ Good,
4 ☐ Fair, or
5 ☐ Poor?
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

55. Would you recommend the {*case manager*} who helps you to your family and friends if they needed {*program-specific term for case-management services*}? Would you say you would recommend the {*case manager*} . . .

1 ☐ Definitely no,
2 ☐ Probably no,
3 ☐ Probably yes, or
4 ☐ Definitely yes?
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

CHOOSING YOUR SERVICES

56. In the last 3 months, did your [*program-specific term for "*service *plan"*] include . . .

1 ☐ **None** of the things that are important to you,
2 ☐ **Some** of the things that are important to you,
3 ☐ **Most** of the things that are important to you, or
4 ☐ **All** of the things that are important to you?
-1 ☐ DON'T KNOW → GO TO Q58
-2 ☐ REFUSED → GO TO Q58

-3 ☐ UNCLEAR RESPONSE → GO TO Q58

57. In the last 3 months, did you feel {*personal assistance/behavioral health staff*} knew what's on your [*program-specific term for "service plan"*], including the things that are important to you?

1 ☐ YES
2 ☐ NO
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

58. In the last 3 months, who would you have talked to if you wanted to change your [*program-specific term for "service plan"*]? Anyone else? [INTERVIEWER MARKS ALL THAT APPLY]

1 ☐ CASE MANAGER
2 ☐ OTHER STAFF
3 ☐ FAMILY/FRIENDS
4 ☐ SOMEONE ELSE, PLEASE SPECIFY _____
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

TRANSPORTATION

The next questions ask about how you get to places in your community.

59. Medical appointments include seeing a doctor, a dentist, a therapist, or someone else who takes care of your health. In the last 3 months, how often did you have a way to get to your medical appointments? Would you say . . .

1 ☐ Never,
2 ☐ Sometimes,
3 ☐ Usually, or
4 ☐ Always?
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

ALTERNATE VERSION: Medical appointments include seeing a doctor, a dentist, a therapist, or someone else who takes care of your health. In the last 3 months, did you have a way to get to your medical appointments? Would you say . . .

1 ☐ Mostly yes or
2 ☐ Mostly no?
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

60. In the last 3 months, did you use a van or some other transportation service? Do not include a van you own.

1☐ YES
2☐ NO → GO TO Q63
-1☐ DON'T KNOW → GO TO Q63
-2☐ REFUSED → GO TO Q63
-3☐ UNCLEAR RESPONSE → GO TO Q63

61. In the last 3 months, were you able to get in and out of this ride easily?

1☐ YES
2☐ NO
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

62. In the last 3 months, how often did this ride arrive on time to pick you up? Would you say . . .

1☐ Never,
2☐ Sometimes,
3☐ Usually, or
4☐ Always?
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

ALTERNATE VERSION: In the last 3 months, did this ride arrive on time to pick you up? Would you say . . .

1☐ Mostly yes or
2☐ Mostly no?
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

PERSONAL SAFETY

The next few questions ask about your personal safety.

63. Who would you contact in case of an emergency? [INTERVIEWER MARKS ALL THAT APPLY]

1☐ FAMILY MEMBER OR FRIEND
2☐ CASE MANAGER
3☐ AGENCY THAT PROVIDES HOME- AND COMMUNITY-BASED SERVICES
4☐ PAID EMERGENCY RESPONSE SERVICE (E.G., LIFELINE)
5☐ 9–1–1 (FIRST RESPONDERS, POLICE, LAW ENFORCEMENT)

6 ☐ SOMEONE ELSE, PLEASE SPECIFY _____
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

64. In the last 3 months, was there a person you could talk to if someone hurt you or did something to you that you didn't like?

1 ☐ YES
2 ☐ NO
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

The next few questions ask if anyone paid to help you treated you badly in the last 3 months. This includes {*personal assistance/behavioral health staff*, *homemakers*, *or your case manager*}. We are asking everyone the next questions—not just you. [ADD STATE-SPECIFIC LANGUAGE HERE REGARDING MANDATED REPORTING, IF APPROPRIATE—"I want to remind you that, although your answers are confidential, I have a legal responsibility to tell {*STATE*} if I hear something that makes me think you are being hurt or are in danger."]

65. In the last 3 months, did **any** {*personal assistance/behavioral health staff, homemakers, or your case managers*} take your money or your things without asking you first?

1 ☐ YES
2 ☐ NO → GO TO Q68
-1 ☐ DON'T KNOW → GO TO Q68
-2 ☐ REFUSED → GO TO Q68
-3 ☐ UNCLEAR RESPONSE → GO TO Q68

66. In the last 3 months, did someone work with you to fix this problem?

1 ☐ YES
2 ☐ NO → GO TO Q68
-1 ☐ DON'T KNOW → GO TO Q68
-2 ☐ REFUSED → GO TO Q68
-3 ☐ UNCLEAR RESPONSE → GO TO Q68

67. In the last 3 months, who has been working with you to fix this problem? Anyone else? [INTERVIEWER MARKS ALL THAT APPLY]

1 ☐ FAMILY MEMBER OR FRIEND
2 ☐ CASE MANAGER
3 ☐ AGENCY
4 ☐ SOMEONE ELSE, PLEASE SPECIFY _____
-1 ☐ DON'T KNOW
-2 ☐ REFUSED

-3 ☐ UNCLEAR RESPONSE

68. In the last 3 months, did any {*staff*} yell, swear, or curse at you?

1 ☐ YES
2 ☐ NO → GO TO Q71
-1 ☐ DON'T KNOW → GO TO Q71
-2 ☐ REFUSED → GO TO Q71
-3 ☐ UNCLEAR RESPONSE → GO TO Q71

69. In the last 3 months, did someone work with you to fix this problem?

1 ☐ YES
2 ☐ NO → GO TO Q71
-1 ☐ DON'T KNOW → GO TO Q71
-2 ☐ REFUSED → GO TO Q71
-3 ☐ UNCLEAR RESPONSE → GO TO Q71

70. In the last 3 months, who has been working with you to fix this problem? Anyone else? [INTERVIEWER MARKS ALL THAT APPLY]

1 ☐ FAMILY MEMBER OR FRIEND
2 ☐ CASE MANAGER
3 ☐ AGENCY
4 ☐ SOMEONE ELSE, PLEASE SPECIFY _____
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

71. In the last 3 months, did any {*staff*} hit you or hurt you?

1 ☐ YES
2 ☐ NO → GO TO Q74
-1 ☐ DON'T KNOW → GO TO Q74
-2 ☐ REFUSED → GO TO Q74
-3 ☐ UNCLEAR RESPONSE → GO TO Q74

72. In the last 3 months, did someone work with you to fix this problem?

1 ☐ YES
2 ☐ NO → GO TO Q74
-1 ☐ DON'T KNOW → GO TO Q74
-2 ☐ REFUSED → GO TO Q74
-3 ☐ UNCLEAR RESPONSE → GO TO Q74

73. In the last 3 months, who has been working with you to fix this problem? Anyone else? [INTERVIEWER MARKS ALL THAT APPLY]

1☐ FAMILY MEMBER OR FRIEND
2☐ CASE MANAGER
3☐ AGENCY
4☐ SOMEONE ELSE, PLEASE SPECIFY _____
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

COMMUNITY INCLUSION AND EMPOWERMENT

Now I'd like to ask you about the things you do in your community.

74. Do you have any **family** members who live nearby? Do not include family members you live with.

1☐ YES
2☐ NO → GO TO Q76
-1☐ DON'T KNOW → GO TO Q76
-2☐ REFUSED → GO TO Q76
-3☐ UNCLEAR RESPONSE → GO TO Q76

75. In the last 3 months, when you wanted to, how often could you get together with these family members who live nearby? Would you say . . .

1☐ Never,
2☐ Sometimes,
3☐ Usually, or
4☐ Always?
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

ALTERNATE VERSION: In the last 3 months, when you wanted to, could you get together with these family members who live nearby? Would you say . . .
1☐ Mostly yes or
2☐ Mostly no?
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

76. Do you have any **friends** who live nearby?

1☐ YES
2☐ NO → GO TO Q78
-1☐ DON'T KNOW → GO TO Q78
-2☐ REFUSED → GO TO Q78
-3☐ UNCLEAR RESPONSE → GO TO Q78

77. In the last 3 months, when you wanted to, how often could you get together with these friends who live nearby? Would you say . . .

1☐ Never,
2☐ Sometimes,
3☐ Usually, or
4☐ Always?
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

ALTERNATE VERSION: In the last 3 months, when you wanted to, could you get together with these friends who live nearby? Would you say . . .
1☐ Mostly yes or
2☐ Mostly no?
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

78. In the last 3 months, when you wanted to, how often could you do things in the community that you like? Would you say . . .

1☐ Never,
2☐ Sometimes,
3☐ Usually, or
4☐ Always?
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

ALTERNATE VERSION: In the last 3 months, when you wanted to, could you do things in the community that you like? Would you say . . .
1☐ Mostly yes or
2☐ Mostly no?
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

79. In the last 3 months, did you need more help than you get from {*personal assistance/behavioral health staff*} to do things in your community?

1☐ YES
2☐ NO
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

80. In the last 3 months, did you take part in deciding **what** you do with your time each day?

1☐ YES
2☐ NO
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

81. In the last 3 months, did you take part in deciding **when** you do things each day—for example, deciding when you get up, eat, or go to bed?

1☐ YES
2☐ NO
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

ABOUT YOU

Now I just have a few more questions about you.

82. In general, how would you rate your overall health? Would you say . . .

1☐ Excellent,
2☐ Very good,
3☐ Good,
4☐ Fair, or
5☐ Poor?
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

83. In general, how would you rate your overall mental or emotional health? Would you say . . .

1☐ Excellent,
2☐ Very good,
3☐ Good,
4☐ Fair, or
5☐ Poor?
-1☐ DON'T KNOW
-2☐ REFUSED
-3☐ UNCLEAR RESPONSE

84. What is your age?

1☐ 18 TO 24 YEARS

2 ☐ 25 TO 34 YEARS
3 ☐ 35 TO 44 YEARS
4 ☐ 45 TO 54 YEARS
5 ☐ 55 TO 64 YEARS
6 ☐ 65 TO 74 YEARS
7 ☐ 75 YEARS OR OLDER
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

ALTERNATE VERSION: In what year were you born?
_____ (YEAR)
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

85. [IF NECESSARY, ASK, AND VERIFY IF OVER THE PHONE] Are you male or female?

1 ☐ MALE
2 ☐ FEMALE
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

86. What is the highest grade or level of school that you have completed?

1 ☐ 8th grade or less
2 ☐ Some high school, but did not graduate
3 ☐ High school graduate or GED
4 ☐ Some college or 2-year degree
5 ☐ 4-year college graduate
6 ☐ More than 4-year college degree
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

87. Are you of Hispanic, Latino, or Spanish origin?

1 ☐ YES, HISPANIC, LATINO, OR SPANISH
2 ☐ NO, NOT HISPANIC, LATINO, OR SPANISH → GO TO Q89
-1 ☐ DON'T KNOW → GO TO Q89
-2 ☐ REFUSED → GO TO Q89
-3 ☐ UNCLEAR RESPONSE → GO TO Q89

88. Which group best describes you? [READ ALL ANSWER CHOICES. CODE ALL THAT APPLY.]

1 ☐ Mexican, Mexican American, Chicano, Chicana

2 ☐ Puerto Rican
3 ☐ Cuban
4 ☐ Another Hispanic, Latino, or Spanish origin
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

89. What is your race? You may choose one or more of the following. Would you say you are. . .

1 ☐ White → GO TO Q92
2 ☐ Black or African-American → GO TO Q92
3 ☐ Asian → GO TO Q90
4 ☐ Native Hawaiian or other Pacific Islander → GO TO Q91
5 ☐ American Indian or Alaska Native → GO TO Q92
6 ☐ OTHER → GO TO Q92
-1 ☐ DON'T KNOW → GO TO Q92
-2 ☐ REFUSED → GO TO Q92
-3 ☐ UNCLEAR RESPONSE → GO TO Q92

90. Which group best describes you? [READ ALL ANSWER CHOICES. CODE ALL THAT APPLY.]

1 ☐ Asian Indian → GO TO Q92
2 ☐ Chinese → GO TO Q92
3 ☐ Filipino → GO TO Q92
4 ☐ Japanese → GO TO Q92
5 ☐ Korean → GO TO Q92
6 ☐ Vietnamese → GO TO Q92
7 ☐ Other Asian → GO TO Q92
-1 ☐ DON'T KNOW → GO TO Q92
-2 ☐ REFUSED → GO TO Q92
-3 ☐ UNCLEAR RESPONSE → GO TO Q92

91. Which group best describes you? [READ ALL ANSWER CHOICES. CODE ALL THAT APPLY.]

1 ☐ Native Hawaiian
2 ☐ Guamanian or Chamorro
3 ☐ Samoan
4 ☐ Other Pacific Islander
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

92. Do you speak a language other than English at home? [READ CHOICES ONLY IF NEEDED…]

1 ☐ Yes

2 ☐ No → GO TO Q94
-1 ☐ DON'T KNOW → GO TO Q94
-2 ☐ REFUSED → GO TO Q94
-3 ☐ UNCLEAR RESPONSE → GO TO Q94

93. What is the language you speak at home?

1 ☐ Spanish,
2 ☐ Some other language → Which one? _____
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

94. [IF NECESSARY, ASK] How many adults live at your home, including you?

1 ☐ 1 [JUST THE RESPONDENT] → END SURVEY
2 ☐ 2 TO 3
3 ☐ 4 OR MORE
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

95. [IF NECESSARY, ASK] Do you live with any family members?

1 ☐ YES
2 ☐ NO
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

96. [IF NECESSARY, ASK] Do you live with people who are not family or are not related to you?

1 ☐ YES
2 ☐ NO
-1 ☐ DON'T KNOW
-2 ☐ REFUSED
-3 ☐ UNCLEAR RESPONSE

INTERVIEWER QUESTIONS

THE FOLLOWING QUESTIONS SHOULD BE ANSWERED AFTER THE INTERVIEW IS CONDUCTED.

97. WAS THE RESPONDENT ABLE TO GIVE VALID RESPONSES?

1 ☐ YES
2 ☐ NO

98. WAS ANY ONE ELSE PRESENT DURING THE INTERVIEW?

1☐ YES
2☐ NO → END SURVEY

99. WHO WAS PRESENT DURING THE INTERVIEW? (MARK ALL THAT APPLY.)

1☐ SOMEONE **NOT** PAID TO PROVIDE SUPPORT TO THE RESPONDENT
2☐ STAFF OR SOMEONE PAID TO PROVIDE SUPPORT TO THE RESPONDENT

100. DID SOMEONE HELP THE RESPONDENT COMPLETE THIS SURVEY?

1☐ YES
2☐ NO → END SURVEY

101. HOW DID THAT PERSON HELP? [MARK ALL THAT APPLY.]

1☐ ANSWERED ALL THE QUESTIONS FOR RESPONDENT
2☐ RESTATED THE QUESTIONS IN A DIFFERENT WAY OR REMINDED/PROMPTED THE RESPONDENT
3☐ TRANSLATED THE QUESTIONS OR ANSWERS INTO THE RESPONDENT'S LANGUAGE
4☐ HELPED WITH THE USE OF ASSISTIVE OR COMMUNICATION EQUIPMENT SO THAT THE RESPONDENT COULD ANSWER THE QUESTIONS
5☐ OTHER, SPECIFY_____

102. WHO HELPED THE RESPONDENT? (MARK ALL THAT APPLY.)

1☐ SOMEONE **NOT** PAID TO PROVIDE SUPPORT TO THE RESPONDENT
2☐ STAFF OR SOMEONE PAID TO PROVIDE SUPPORT TO THE RESPONDENT

CAHPS® Home- and Community-Based Services Survey

Version: 1.0

Population: Adult

Language: Spanish

Instructions for Vendor

- The interview is intended as an interviewer-administered survey; thus all text that appears in initial uppercase and lowercase letters should be read aloud. Text that appears in **bold, lowercase letters** should be emphasized.

- Text in {*italics and in braces*} will be provided by the HCBS program's administrative data. However, if the interviewee provides another term, that term should be used in place of the program-specific term wherever indicated. For example, some interviewees may refer to their case manager by another title, which should be used instead throughout the survey.

- For response options of "never," "sometimes," "usually," and "always," if the respondent cannot use that scale, the alternate version of the survey with response options of "mostly yes" and "mostly no" should be used. These alternate response options are reserved for respondents who find the "never," "sometimes," "usually," "always" response scale cognitively challenging.

- For response options of 0 to 10, if the respondent cannot use that scale, the alternate version of the survey with response options of "excellent," "very good," "good," "fair," or "poor" should be used. These alternate response options are reserved for respondents who find the numeric scale cognitively challenging.

- All questions include a "REFUSED" response option. In this case, "refused" means the respondent did not provide any answer to the question.

- All questions include a "DON'T KNOW" response option. This is used when the respondent indicates that he or she does not know the answer and cannot provide a response to the question.

- All questions include an "UNCLEAR" response option. This should be used when a respondent answers, but the interviewer cannot clarify the meaning of the response even after minor probing **or** the response is completely unrelated to the question, (e.g., the response to "In the last 3 months, how often did your homemakers listen carefully to what you say?" is "I like to sit by Mary").

- Some responses have skip patterns, which are expressed as "→ GO TO Q#." The interviewer should be advanced to the next appropriate item to ask the respondent.

- Not all respondents receive all home and community-based services asked about in this instrument. Items Q4 through Q12 help to confirm which services a respondent receives. The table after it summarizes the logic of which items should be used.

- Survey users may add questions to this survey before the "About You" section. A separate supplemental employment module can be added.

- Use singular/plural as needed. In most cases, questions are written assuming there is more than one staff person supporting a respondent or it is written without an indication of whether there is more

than one staff person. Based on information collected from Q4 through Q12, it is possible to modify questions to be singular or plural as they relate to staff.

- Use program-specific terms. Where appropriate, add in the program-specific terms for staff (e.g., [*program-specific term for these types of staff*]) but allow the interviewer to modify the term based on the respondent's choice of the word. It will be necessary to obtain information for program-specific terms. State administrative data should include the following information:
  - Agency name(s)
  - Titles of staff who provide care
  - Names of staff who provide care
  - Activities that each staff member provides (this will help with identifying appropriate skip logic)
  - Hours of staff who come to the home

COGNITIVE SCREENING QUESTIONS

Es posible que algunas personas se les pague para que le ayuden a alistarse por la mañana, a hacer los oficios de la casa, a ir a algún sitio o a recibir servicios de salud mental. Esta encuesta es sobre las personas a las que se les paga para que le ayuden con las actividades que hace normalmente o comúnmente en la casa y en la comunidad. También contiene preguntas sobre los servicios que recibe.

1. ¿Viene alguien a su casa para ayudarle?

[1] ☐ SÍ
[2] ☐ NO → END SURVEY
[-1] ☐ NO SABE → END SURVEY
[-2] ☐ SE NEGÓ A CONTESTAR → END SURVEY
[-3] ☐ RESPUESTA POCO CLARA → END SURVEY

2. ¿Cómo le ayudan?

_____

[EXAMPLES OF CORRECT RESPONSES INCLUDE]

- ME AYUDA A ALISTARME TODOS LOS DIAS (HELPS ME GET READY EVERY DAY)
- LIMPIA MI CASA (CLEANS MY HOME)
- TRABAJA CONMIGO EN MI EMPLEO (WORKS WITH ME AT MY JOB)
- ME AYUDA HACER COSAS (HELPS ME DO THINGS)
- ME AYUDA CON TRANSPORTE (DRIVES ME AROUND)

[-1] ☐ NO SABE → GO TO THE ABOUT YOU SECTION
[-2] ☐ SE NEGÓ A CONTESTAR → GO TO THE ABOUT YOU SECTION
[-3] ☐ RESPUESTA POCO CLARA → GO TO THE ABOUT YOU SECTION

3. ¿Cómo llama usted a esa(s) persona(s)?

_____

[EXAMPLES OF SUFFICIENT RESPONSES INCLUDE]

- MI TRABAJADOR(A) (MY WORKER)
- MI ASISTENTE (MY ASSISTANT)
- POR SU(S) NOMBRE(S), MARIA, ANA, ETCC. (NAMES OF STAFF (JO, DAWN, ETC.))

[-1] ☐ NO SABE → GO TO THE ABOUT YOU SECTION
[-2] ☐ SE NEGÓ A CONTESTAR → GO TO THE ABOUT YOU SECTION
[-3] ☐ RESPUESTA POCO CLARA → GO TO THE ABOUT YOU SECTION

CSQPASS.

[IF ALL 3 QUESTIONS WERE ANSWERED CORRECTLY, ENTER 1 TO CONTINUE]

1 PASS - ALL 3 QUESTIONS WERE ANSWERED CORRECTLY → GO TO Q4

2 FAIL - AT LEAST 1 QUESTION WAS NOT ANSWERED CORRECTLY → GO TO SURVEND

SURVEND.
GRACIAS POR SU TIEMPO. ESAS SON TODAS LAS PREGUNTAS QUE TENEMOS.
TENGA UN BUEN DIA/TARDE [ENTER 1 TO EXIT SURVEY]

PREGUNTAS DE IDENTIFICACIÓN

Ahora me gustaría hacerle más preguntas sobre el tipo de personas que vienen a su casa para ayudarle.

4. En los últimos 3 meses, ¿recibió usted {*program specific term for personal assistance*} en casa?

   1 ☐ SÍ
   2 ☐ NO → Go to Q6
   -1 ☐ NO SABE → Go to Q6
   -2 ☐ SE NEGÓ A CONTESTAR → Go to Q6
   -3 ☐ RESPUESTA POCO CLARA → Go to Q6

5. ¿Cómo llama usted a la(s) persona(s) que le dio(dieron) {*program specific term for personal assistance*}? Por ejemplo, ¿les llama {*program specific term for personal assistance}*, personal, auxiliares de cuidados personales (*PCAs* por su sigla en inglés), trabajadores o alguna otra cosa?

   _____

   [ADD RESPONSE WHEREVER IT SAYS "*personal assistance/behavioral health staff", "el personal de salud mental / los auxiliares de cuidados personales*"]

6. En los últimos 3 meses, ¿recibió usted {*program specific term for behavioral health specialist services*} en casa?

   1 ☐ SÍ
   2 ☐ NO → Go to Q8
   -1 ☐ NO SABE → Go to Q8
   -2 ☐ SE NEGÓ A CONTESTAR → Go to Q8
   -3 ☐ RESPUESTA POCO CLARA → Go to Q8

7. ¿Cómo llama usted a la(s) persona(s) que le dio(dieron) {*program specific term for behavioral health specialist services*}? Por ejemplo, ¿les llama {*program specific term for behavioral health specialists*}, consejeros, apoyo de personas en la misma situación (*peer support* en inglés), asistentes de recuperación o alguna otra cosa?

   _____

   [ADD RESPONSE WHEREVER IT SAYS "*personal assistance/behavioral health staff."* IF Q4 ALSO= YES, LIST BOTH TITLES]

8. En los últimos 3 meses, ¿recibió usted {*program specific term for homemaker services*} en casa?

   1 ☐ SÍ
   2 ☐ NO → GO to Q11
   -1 ☐ NO SABE → Q11
   -2 ☐ SE NEGÓ A CONTESTAR → GO to Q11
   -3 ☐ RESPUESTA POCO CLARA → GO to Q11

9. ¿Cómo llama usted a la(s) persona(s) que le dio(dieron) {*program specific term for homemaker services*}? Por ejemplo, ¿les llama {*program specific term for homemaker*}, ayudantes de oficios domésticos, ayudantes para tareas de la casa o alguna otra cosa?

_____

[ADD RESPONSE WHEREVER IT SAYS "*homemaker"*]

10. [IF (Q4 *OR* Q6) *AND* Q8= YES, ASK] En los últimos 3 meses, las personas que le ayudan con las actividades que hace normalmente o comúnmente ¿también le ayudan a limpiar la casa?

1 ☐ SÍ
2 ☐ NO
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

11. En los últimos 3 meses, ¿recibió usted ayuda de {*program specific term for case manager services*} para asegurarse de que usted recibió todos los servicios que necesitó?

1 ☐ SÍ
2 ☐ NO
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

12. ¿Cómo llama usted a la persona que le dio {*program specific term for case manager services*}? Por ejemplo, ¿llama a esa persona {*program specific term for case manager*}, encargado de caso, encargado de cuidados, coordinador de servicios, coordinador de servicios de apoyo, trabajador social o alguna otra cosa?

_____

[ADD RESPONSE WHEREVER IT SAYS "*case manager"*]

BELOW ARE INSTRUCTIONS FOR WHICH QUESTIONS TO ASK FOR EACH RESPONSE ABOVE

ITEM AND RESPONSE—FOLLOW ALL ROWS THAT APPLY	**ACTION**
IF Q4 OR Q6 = YES (PERSONAL ASSISTANCE OR BEHAVIORAL HEALTH SPECIALIST SERVICES), AND Q8 = NO, DON'T KNOW, REFUSE, UNCLEAR (HOMEMAKER SERVICES)	ASK Q13–Q36, AND Q48 ONWARD

IF Q4 OR Q6 = YES (PERSONAL ASSISTANCE OR BEHAVIORAL HEALTH SPECIALIST SERVICES), AND Q8 = YES (HOMEMAKER SERVICES)	ASK Q13 ONWARD
IF Q4 AND Q6 = NO (PERSONAL ASSISTANCE OR BEHAVIORAL HEALTH SPECIALIST SERVICES)	SKIP Q13–36, Q57 AND Q79
IF Q8 = YES (HOMEMAKER SERVICES)	ASK Q37–47 AND ONWARD
IF Q10 = YES (HOMEMAKER AND PERSONAL ASSISTANCE STAFF SAME)	ASK Q13–Q36, Q39, Q40, AND Q48 ONWARD
IF Q11 = ANY RESPONSE (CASE MANAGER)	ASK Q48–Q55 AND Q56 ONWARD

Obtención de los servicios necesarios de parte de los auxiliares de cuidados personales y del personal de salud mental

13. Primero me gustaría hablar sobre {*el personal de salud mental / los auxiliares de cuidados personales*} la(s) persona(s) a la(s) que se le(s) paga para que le ayude(n) en sus actividades diarias, como vestirse, ir al baño, bañarse o ducharse, o ir a algún sitio. En los últimos 3 meses, ¿con qué frecuencia el/los {*el personal de salud mental / los auxiliares de cuidados personales*} llegó/llegaron a trabajar a tiempo? ¿Diría que…?

- 1 ☐ Nunca,
- 2 ☐ A veces,
- 3 ☐ Casi siempre, o
- 4 ☐ Siempre?
- -1 ☐ NO SABE
- -2 ☐ SE NEGÓ A CONTESTAR
- -3 ☐ RESPUESTA POCO CLARA

Versión Alternativa: Primero me gustaría hablar sobre el/los{*el personal de salud mental / los auxiliares de cuidados personales*}, a quien(es) se le(s) paga para que le ayude(n) en sus actividades diarias, como vestirse, ir al baño, bañarse o ducharse, o ir a algún sitio. ¿ En los últimos 3 meses, llegó/llegaron el/los {*el personal de salud mental / los auxiliares de cuidados personales*} a trabajar a tiempo? ¿Diría que…?

- 1 ☐ En general, sí, o
- 2 ☐ En general, no?
- -1 ☐ NO SABE
- -2 ☐ SE NEGÓ A CONTESTAR
- -3 ☐ RESPUESTA POCO CLARA

14. En los últimos 3 meses, ¿con qué frecuencia {*el personal de salud mental / los auxiliares de cuidados personales*} trabajaron todo el tiempo que se suponía que deberían trabajar? ¿Diría que…?

- [1] ☐ Nunca,
- [2] ☐ A veces,
- [3] ☐ Casi siempre, o
- [4] ☐ Siempre?
- [-1] ☐ NO SABE
- [-2] ☐ SE NEGÓ A CONTESTAR
- [-3] ☐ RESPUESTA POCO CLARA

Versión Alternativa: En los últimos 3 meses, ¿trabajaron el/los {*el personal de salud mental / los auxiliares de cuidados personales*} todo el tiempo que se suponía que deberían trabajar? ¿Diría que…?

- [1] ☐ En general, sí, o
- [2] ☐ En general, no?
- [-1] ☐ NO SABE
- [-2] ☐ SE NEGÓ A CONTESTAR
- [-3] ☐ RESPUESTA POCO CLARA

15. A veces el personal no puede llegar al trabajo en un día en que tenga programado hacerlo. En los últimos 3 meses, cuando el personal no pudo llegar al trabajo en un día en que tenía programado hacerlo, ¿le avisó alguien que {*el personal de salud mental / los auxiliares de cuidados personales*} no podía llegar ese día?

- [1] ☐ SÍ
- [2] ☐ NO
- [-1] ☐ NO SABE
- [-2] ☐ SE NEGÓ A CONTESTAR
- [-3] ☐ RESPUESTA POCO CLARA

16. En los últimos 3 meses, ¿necesitó ayuda de {*el personal de salud mental / los auxiliares de cuidados personales*} para vestirse, ducharse o bañarse?

- [1] ☐ SÍ
- [2] ☐ NO → GO TO Q20
- [-1] ☐ NO SABE → GO TO Q20
- [-2] ☐ SE NEGÓ A CONTESTAR → GO TO Q20
- [-3] ☐ RESPUESTA POCO CLARA → GO TO Q20

17. En los últimos 3 meses, ¿**Siempre** se vistió, se duchó o se bañó cuando lo necesitaba?

- [1] ☐ SÍ → GO TO Q19
- [2] ☐ NO
- [-1] ☐ NO SABE → GO TO Q19
- [-2] ☐ SE NEGÓ A CONTESTAR → GO TO Q19
- [-3] ☐ RESPUESTA POCO CLARA → GO TO Q19

18. En los últimos 3 meses, ¿esto pasó porque no había {*auxiliares de cuidados personales / personal de salud mental*} que le ayude(n)?

- [1] ☐ SÍ
- [2] ☐ NO

-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

19. En los últimos 3 meses, ¿con qué frecuencia {*el personal de salud mental / los auxiliares de cuidados personales*} se aseguró de que usted tuviera suficiente privacidad cuando se vestía, se duchará o se bañará? ¿Diría que…?

1 ☐ Nunca,
2 ☐ A veces,
3 ☐ Casi siempre, o
4 ☐ Siempre?
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

Versión Alternativa: En los últimos 3 meses, ¿se aseguró el/los {*el personal de salud mental / los auxiliares de cuidados personales*} de que usted tuviera suficiente privacidad cuando se vestía, se duchará o se bañará ? ¿Diría que…?

1 ☐ En general, sí, o
2 ☐ En general, no?
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

20. En los últimos 3 meses, ¿necesitó que {*el personal de salud mental / los auxiliares de cuidados personales*} le ayude(n) con las comidas, por ejemplo, para preparar o cocinar las comidas o para ayudarle a comer?

1 ☐ SÍ
2 ☐ NO → GO TO Q23
-1 ☐ NO SABE → GO TO Q23
-2 ☐ SE NEGÓ A CONTESTAR → GO TO Q23
-3 ☐ RESPUESTA POCO CLARA → GO TO Q23

21. En los últimos 3 meses, ¿**siempre** pudo conseguir algo para comer cuando tenía hambre?

1 ☐ SÍ → GO TO Q23
2 ☐ NO
-1 ☐ NO SABE → GO TO Q23
-2 ☐ SE NEGÓ A CONTESTAR → GO TO Q23
-3 ☐ RESPUESTA POCO CLARA → GO TO Q23

22. En los últimos 3 meses, ¿esto pasó porque no había {*auxiliares de cuidados personales / personal de salud mental*} que le ayude(n)?

1 ☐ SÍ
2 ☐ NO
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

23. A veces las personas necesitan ayuda para tomarse sus medicinas, por ejemplo, necesitan ayuda para acordarse de tomárselas, para servirlas o para alistar las pastillas. En los últimos 3 meses, ¿necesitó que {*el personal de salud mental / los auxiliares de cuidados personales*} le ayude(n) a tomarse sus medicinas?

1 ☐ SÍ
2 ☐ NO → GO TO Q26
-1 ☐ NO SABE → GO TO Q26
-2 ☐ SE NEGÓ A CONTESTAR → GO TO Q26
-3 ☐ RESPUESTA POCO CLARA → GO TO Q26

24. En los últimos 3 meses, ¿**siempre** se tomó su medicina cuando debía tomársela?

1 ☐ SÍ → GO TO Q26
2 ☐ NO
-1 ☐ NO SABE → GO TO Q26
-2 ☐ SE NEGÓ A CONTESTAR → GO TO Q26
-3 ☐ RESPUESTA POCO CLARA → GO TO Q26

25. En los últimos 3 meses, ¿esto pasó porque no había {*auxiliares de cuidados personales / personal de salud mental*} que le ayude(n)?

1 ☐ SÍ
2 ☐ NO
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

26. La ayuda para ir al baño incluye ayudarle a alguien a sentarse y levantarse del inodoro o ayudarle a cambiarse de ropa interior o de toallas desechables. En los últimos 3 meses, ¿necesitó que {*el personal de salud mental / los auxiliares de cuidados personales*} le ayude(n) a ir al baño?

1 ☐ SÍ
2 ☐ NO → GO TO Q28
-1 ☐ NO SABE → GO TO Q28
-2 ☐ SE NEGÓ A CONTESTAR → GO TO Q28
-3 ☐ RESPUESTA POCO CLARA → GO TO Q28

27. En los últimos 3 meses, ¿recibió usted toda la ayuda que necesitó de {*el personal de salud mental / los auxiliares de cuidados personales*} para ir al baño cuando lo necesitó?

1 ☐ SÍ
2 ☐ NO
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

Qué tan bien se comunica(n) con usted los auxiliares de cuidados personales o el personal de salud mental y qué tan bien lo(a) tratan

Las siguientes preguntas se refieren a cómo lo(a) trata(n) {*el personal de salud mental / los auxiliares de cuidados personales*}.

28. En los últimos 3 meses, ¿con qué frecuencia {*el personal de salud mental / los auxiliares de cuidados personales*} lo(a) trató con cortesía y respeto? ¿Diría que…?

[1] ☐ Nunca,
[2] ☐ A veces,
[3] ☐ Casi siempre, o
[4] ☐ Siempre?
[-1] ☐ NO SABE
[-2] ☐ SE NEGÓ A CONTESTAR
[-3] ☐ RESPUESTA POCO CLARA

Versión Alternativa: En los últimos 3 meses, ¿{*el personal de salud mental / los auxiliares de cuidados personales*} lo(a) trató con cortesía y respeto? ¿Diría que…?

[1] ☐ En general, sí, o
[2] ☐ En general, no?
[-1] ☐ NO SABE
[-2] ☐ SE NEGÓ A CONTESTAR
[-3] ☐ RESPUESTA POCO CLARA

29. En los últimos 3 meses, ¿con qué frecuencia fue difícil entender las explicaciones que le dio/dieron {*el personal de salud mental / los auxiliares de cuidados personales*} porque tiene(n) acento o por la forma en que ellos o ellas hablaron español? ¿Diría que…?

[1] ☐ Nunca,
[2] ☐ A veces,
[3] ☐ Casi siempre, o
[4] ☐ Siempre?
[-1] ☐ NO SABE
[-2] ☐ SE NEGÓ A CONTESTAR
[-3] ☐ RESPUESTA POCO CLARA

Versión Alternativa: En los últimos 3 meses, ¿Fue difícil entender las explicaciones que le dio/dieron{*el personal de salud mental / los auxiliares de cuidados personales*} porque estos tiene(n) acento o por la forma en que hablaron español? ¿Diría que…?

[1] ☐ En general, sí, o
[2] ☐ En general, no?
[-1] ☐ NO SABE
[-2] ☐ SE NEGÓ A CONTESTAR
[-3] ☐ RESPUESTA POCO CLARA

30. En los últimos 3 meses, ¿con qué frecuencia {l*os auxiliares de cuidados personales / el personal de salud mental*} lo(a) trató como usted quiso? ¿Diría que…?

[1] ☐ Nunca,
[2] ☐ A veces,
[3] ☐ Casi siempre, o
[4] ☐ Siempre?
[-1] ☐ NO SABE
[-2] ☐ SE NEGÓ A CONTESTAR
[-3] ☐ RESPUESTA POCO CLARA

Versión Alternativa: En los últimos 3 meses,¿{*el personal de salud mental / los auxiliares de cuidados personales*} lo(a) trató como usted quiso? ¿Diría que…?

[1] ☐ En general, sí, o
[2] ☐ En general, no?
[-1] ☐ NO SABE
[-2] ☐ SE NEGÓ A CONTESTAR
[-3] ☐ RESPUESTA POCO CLARA

31. En los últimos 3 meses, ¿con qué frecuencia {*el personal de salud mental / los auxiliares de cuidados personales*} le explicó las cosas de una manera fácil de entender? ¿Diría que…?

[1] ☐ Nunca,
[2] ☐ A veces,
[3] ☐ Casi siempre, o
[4] ☐ Siempre?
[-1] ☐ NO SABE
[-2] ☐ SE NEGÓ A CONTESTAR
[-3] ☐ RESPUESTA POCO CLARA

Versión Alternativa: En los últimos 3 meses, ¿{*el personal de salud mental / los auxiliares de cuidados personales*} le explicó las cosas de una manera fácil de entender? ¿Diría que…?

[1] ☐ En general, sí, o
[2] ☐ En general, no?
[-1] ☐ NO SABE
[-2] ☐ SE NEGÓ A CONTESTAR
[-3] ☐ RESPUESTA POCO CLARA

32. En los últimos 3 meses, ¿con qué frecuencia {*el personal de salud mental / los auxiliares de cuidados personales*} lo(a) escuchó con atención? ¿Diría que…?

[1] ☐ Nunca,
[2] ☐ A veces,
[3] ☐ Casi siempre, o
[4] ☐ Siempre?
[-1] ☐ NO SABE
[-2] ☐ SE NEGÓ A CONTESTAR
[-3] ☐ RESPUESTA POCO CLARA

Versión Alternativa: En los últimos 3 meses, ¿{*el personal de salud mental / los auxiliares de cuidados personales*} lo(a) escuchó con atención? ¿Diría que…?

[1] ☐ En general, sí, o
[2] ☐ En general, no?
[-1] ☐ NO SABE
[-2] ☐ SE NEGÓ A CONTESTAR
[-3] ☐ RESPUESTA POCO CLARA

33. En los últimos 3 meses, ¿cree usted que {*el personal de salud mental / los auxiliares de cuidados personales*} supieron el tipo de ayuda que **usted** necesita con las actividades diarias, como alistarse por la mañana, hacer mercado o ir a alguna parte de su comunidad?

- [1] ☐ SÍ
- [2] ☐ NO
- [-1] ☐ NO SABE
- [-2] ☐ SE NEGÓ A CONTESTAR
- [-3] ☐ RESPUESTA POCO CLARA

34. En los últimos 3 meses, ¿{*el personal de salud mental / los auxiliares de cuidados personales*} lo(a) animó a hacer cosas sin ayuda si usted podía hacerlas?

- [1] ☐ SÍ
- [2] ☐ NO
- [-1] ☐ NO SABE
- [-2] ☐ SE NEGÓ A CONTESTAR
- [-3] ☐ RESPUESTA POCO CLARA

35. Usando un número del 0 al 10, el 0 siendo la peor ayuda que recibe de {*el personal de salud mental / los auxiliares de cuidados personales*} posible y el 10 es la mejor ayuda que recibe de {*el personal de salud mental / los auxiliares de cuidados personales*} posible, ¿qué número usaría para calificar la ayuda que recibe de {*el personal de salud mental / los auxiliares de cuidados personales*}?

____ 0 a 10

- [-1] ☐ NO SABE
- [-2] ☐ SE NEGÓ A CONTESTAR
- [-3] ☐ RESPUESTA POCO CLARA

Versión Alternativa: ¿Cómo calificaría la ayuda que recibe de {*el personal de salud mental / los auxiliares de cuidados personales*}? ¿Diría que es…?

- [1] ☐ Excelente,
- [2] ☐ Muy buena,
- [3] ☐ Buena,
- [4] ☐ Regular, o
- [5] ☐ Mala?
- [-1] ☐ NO SABE
- [-2] ☐ SE NEGÓ A CONTESTAR
- [-3] ☐ RESPUESTA POCO CLARA

36. ¿Les recomendaría a sus familiares y amigos {*el personal de salud mental / los auxiliares de cuidados personales*} que le ayuda(n) si ellos necesitaran ayuda para realizar las actividades diarias? ¿Diría que recomendaría {*el personal de salud mental / los auxiliares de cuidados personales*}?

- [1] ☐ Definitivamente no,
- [2] ☐ Probablemente no,
- [3] ☐ Probablemente sí, o
- [4] ☐ Definitivamente sí?
- [-1] ☐ NO SABE
- [-2] ☐ SE NEGÓ A CONTESTAR

-3 ☐ RESPUESTA POCO CLARA

Obtención de los servicios necesarios de los ayudantes de oficios domésticos

Las siguientes preguntas son acerca de los {*ayudantes de oficios domésticos*}, el personal a quien se le paga para que haga tareas de la casa, como limpiar, hacer mercado o lavar la ropa.

37. En los últimos 3 meses, ¿con qué frecuencia {*los ayudantes de oficios domésticos*} llegaron a tiempo al trabajo? ¿Diría que…?

- [1] ☐ Nunca,
- [2] ☐ A veces,
- [3] ☐ Casi siempre, o
- [4] ☐ Siempre?
- [-1] ☐ NO SABE
- [-2] ☐ SE NEGÓ A CONTESTAR
- [-3] ☐ RESPUESTA POCO CLARA

Versión Alternativa: En los últimos 3 meses, ¿llegaron los {*los ayudantes de oficios domésticos*} a tiempo al trabajo? ¿Diría que…?

- [1] ☐ En general, sí, o
- [2] ☐ En general, no?
- [-1] ☐ NO SABE
- [-2] ☐ SE NEGÓ A CONTESTAR
- [-3] ☐ RESPUESTA POCO CLARA

38. En los últimos 3 meses, ¿con qué frecuencia {l*os ayudantes de oficios domésticos*} trabajaron todo el tiempo que se suponía que debían trabajar? ¿Diría que…?

- [1] ☐ Nunca,
- [2] ☐ A veces,
- [3] ☐ Casi siempre, o
- [4] ☐ Siempre?
- [-1] ☐ NO SABE
- [-2] ☐ SE NEGÓ A CONTESTAR
- [-3] ☐ RESPUESTA POCO CLARA

Versión Alternativa: En los últimos 3 meses, ¿trabajaron {l*os ayudantes de oficios domésticos*} todo el tiempo que se suponía que debían trabajar? ¿Diría que…?

- [1] ☐ En general, sí, o
- [2] ☐ En general, no?
- [-1] ☐ NO SABE
- [-2] ☐ SE NEGÓ A CONTESTAR
- [-3] ☐ RESPUESTA POCO CLARA

39. En los últimos 3 meses, las tareas de la casa, como limpiar y lavar la ropa ¿se hicieron **siempre** cuando usted necesitaba que se hicieran? [ASK IF HOMEMAKER IS THE SAME AS PCA STAFF]

- [1] ☐ SÍ → GO TO Q41
- [2] ☐ NO
- [-1] ☐ NO SABE → GO TO Q41
- [-2] ☐ SE NEGÓ A CONTESTAR → GO TO Q41

-3 ☐ RESPUESTA POCO CLARA → GO TO Q41

40. ¿Fue porque no había {*ayudantes de oficios domésticos*} que le ayudaran? [ASK IF HOMEMAKER IS THE SAME AS PCA STAFF]

1 ☐ SÍ
2 ☐ NO
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

Qué tan bien se comunican con usted los ayudantes de oficios domésticos y qué tan bien lo(a) tratan

Las siguientes preguntas se refieren a la forma en que lo(a) tratan *{los ayudantes de oficios domésticos}.*

41. En los últimos 3 meses, ¿con qué frecuencia {*los ayudantes de oficios domésticos*} lo(a) trataron con cortesía y respeto? ¿Diría que…?

1 ☐ Nunca,
2 ☐ A veces,
3 ☐ Casi siempre, o
4 ☐ Siempre?
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

Versión Alternativa: En los últimos 3 meses, ¿{*los ayudantes de oficios domésticos*} lo(a) trataron con cortesía y respeto? ¿Diría que…?

1 ☐ En general, sí, o
2 ☐ En general, no?
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

42. En los últimos 3 meses, ¿con qué frecuencia fue difícil entender las explicaciones que le dieron {*los ayudantes de oficios domésticos*} porque tienen acento o por la forma en que ellos o ellas hablaron español? ¿Diría que…?

1 ☐ Nunca,
2 ☐ A veces,
3 ☐ Casi siempre, o
4 ☐ Siempre?
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

Versión Alternativa: En los últimos 3 meses, ¿Fue difícil entender las explicaciones que le dieron {*los ayudantes de oficios domésticos*} porque estos tienen acento o por la forma en que {*los ayudantes de oficios domésticos*} hablaron español? ¿Diría que…?

1 ☐ En general, sí, o

2 ☐ En general, no?
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

43. En los últimos 3 meses, ¿con qué frecuencia {*los ayudantes de oficios domésticos*} lo(a) trataron como usted quiere? ¿Diría que…?

1 ☐ Nunca,
2 ☐ A veces,
3 ☐ Casi siempre, o
4 ☐ Siempre?
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

Versión Alternativa: En los últimos 3 meses, ¿{*los ayudantes de oficios domésticos*} lo(a) trataron como usted quiere? ¿Diría que…?

1 ☐ En general, sí, o
2 ☐ En general, no?
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

44. En los últimos 3 meses, ¿con qué frecuencia {*los ayudantes de oficios domésticos*} le escucharon con atención? ¿Diría que…?

[1] ☐ Nunca,
[2] ☐ A veces,
[3] ☐ Casi siempre, o
[4] ☐ Siempre?
[-1] ☐ NO SABE
[-2] ☐ SE NEGÓ A CONTESTAR
[-3] ☐ RESPUESTA POCO CLARA

Versión Alternativa: En los últimos 3 meses, ¿{*los ayudantes de oficios domésticos*} le escucharon con atención? ¿Diría que…?

[1] ☐ En general, sí, o
[2] ☐ En general, no?
[-1] ☐ NO SABE
[-2] ☐ SE NEGÓ A CONTESTAR
[-3] ☐ RESPUESTA POCO CLARA

45. En los últimos 3 meses, ¿cree usted que {*los ayudantes de oficios domésticos*} supieron el tipo de ayuda que usted necesitaba?

[1] ☐ SÍ
[2] ☐ NO
[-1] ☐ NO SABE
[-2] ☐ SE NEGÓ A CONTESTAR
[-3] ☐ RESPUESTA POCO CLARA

46. ¿Usando un número del 0 al 10, el 0 siendo la peor ayuda que recibe de {*los ayudantes de oficios domésticos*} posible y el 10 es la mejor ayuda que recibe de {*los ayudantes de oficios domésticos*} posible, ¿qué número usaría para calificar la ayuda que recibe de {*los ayudantes de oficios domésticos*}?

____ 0 a 10

[-1] ☐ NO SABE
[-2] ☐ SE NEGÓ A CONTESTAR
[-3] ☐ RESPUESTA POCO CLARA

Versión Alternativa: ¿Cómo calificaría la ayuda que recibe de {*los ayudantes de oficios domésticos*}? ¿Diría que es…?

[1] ☐ Excelente,
[2] ☐ Muy buena,
[3] ☐ Buena,
[4] ☐ Regular, o
[5] ☐ Mala?
[-1] ☐ NO SABE
[-2] ☐ SE NEGÓ A CONTESTAR
[-3] ☐ RESPUESTA POCO CLARA

47. ¿Les recomendaría a sus familiares y amigos {*los ayudantes de oficios domésticos*} que le ayudan si ellos necesitaran {*término específico del encuestado para "servicios de ayuda con los oficios domésticos"*}? ¿Diría que recomendaría {los ayudantes de oficios domésticos}?

1 ☐ Definitivamente no,
2 ☐ Probablemente no,
3 ☐ Probablemente sí, o
4 ☐ Definitivamente sí?
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

Su encargado de caso

Ahora me gustaría hablarle de su *{encargado de caso}*, la persona que se asegura de que usted reciba los servicios que necesita.

48. ¿Sabe quién es su {*encargado de caso*}?

1 ☐ SÍ
2 ☐ NO → GO TO Q56
-1 ☐ NO SABE → GO TO Q56
-2 ☐ SE NEGÓ A CONTESTAR → GO TO Q56
-3 ☐ RESPUESTA POCO CLARA → GO TO Q56

49. En los últimos 3 meses, ¿Pudo comunicarse con este {*encargado de caso*} cuando necesitó hacerlo?

1 ☐ SÍ
2 ☐ NO
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

50. Algunas personas necesitan conseguir equipo, como sillas de ruedas o andadores, que les sirvan de ayuda y otras personas necesitan que el equipo que tienen sea remplazado o reparado. En los últimos 3 meses, ¿le pidió ayuda a este {*encargado de caso*} para conseguir o reparar un equipo?

1 ☐ SÍ
2 ☐ NO → GO TO Q52
3 ☐ NO NECESITA→ GO TO Q52
-1 ☐ NO SABE → GO TO Q52
-2 ☐ SE NEGÓ A CONTESTAR → GO TO Q52
-3 ☐ RESPUESTA POCO CLARA → GO TO Q52

51. En los últimos 3 meses, ¿Este {*encargado de caso*} colaboró con usted cuando le pidió ayuda para conseguir o reparar un equipo?

1 ☐ SÍ
2 ☐ NO
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

52. En los últimos 3 meses, ¿le pidió ayuda a este {*encargado de caso*} para hacer cambios en los servicios que recibe, como más ayuda de {*el personal de salud mental/los auxiliares de cuidados personales y/o los ayudantes de oficios domésticos*}, o para ir a lugares o buscar trabajo?

[1] ☐ SÍ
[2] ☐ NO → GO TO 54
[3] ☐ NO NECESITA→ GO TO Q54
[-1] ☐ NO SABE → GO TO 54
[-2] ☐ SE NEGÓ A CONTESTAR → GO TO 54
[-3] ☐ RESPUESTA POCO CLARA → GO TO 54

53. En los últimos 3 meses, ¿este {*encargado de caso*} colaboró con usted cuando le pidió ayuda para hacer otros cambios en los servicios que recibe?

[1] ☐ SÍ
[2] ☐ NO
[-1] ☐ NO SABE
[-2] ☐ SE NEGÓ A CONTESTAR
[-3] ☐ RESPUESTA POCO CLARA

54. ¿Usando un número del 0 al 10, el 0 siendo la peor ayuda que recibe del {*encargado de caso*} posible y el 10 es la mejor ayuda que recibe del {*encargado de caso*} posible, ¿qué número usaría para calificar la ayuda que recibe del {*encargado de caso*}?

____ 0 a 10

[-1] ☐ NO SABE
[-2] ☐ SE NEGÓ A CONTESTAR
[-3] ☐ RESPUESTA POCO CLARA

Versión Alternativa: ¿Cómo calificaría la ayuda que recibe del {*encargado de caso*}? ¿Diría que es…?

[1] ☐ Excelente,
[2] ☐ Muy buena,
[3] ☐ Buena,
[4] ☐ Regular, o
[5] ☐ Mala?
[-1] ☐ NO SABE
[-2] ☐ SE NEGÓ A CONTESTAR
[-3] ☐ RESPUESTA POCO CLARA

55. ¿Les recomendaría a sus familiares y amigos el {*encargado de caso*} que le ayuda a usted si ellos necesitaran {*término específico del encuestado para "servicios que presta un encargado de caso"*}? ¿Diría que les recomendaría el {*encargado de caso*}?

[1] ☐ Definitivamente no,
[2] ☐ Probablemente no,
[3] ☐ Probablemente sí, o
[4] ☐ Definitivamente sí?
[-1] ☐ NO SABE
[-2] ☐ SE NEGÓ A CONTESTAR
[-3] ☐ RESPUESTA POCO CLARA

La elección de sus servicios

56. En los últimos 3 meses, ¿se incluyó en su [*término específico de cada programa que se refiere a un "plan de servicios"*]?

- 1 ☐ **Ninguna** de las cosas que son importantes para usted
- 2 ☐ **Algunas** de las cosas que son importantes para usted
- 3 ☐ **La mayoría** de las cosas que son importantes para usted
- 4 ☐ **Todas** las cosas que son importantes para usted
- -1 ☐ NO SABE → GO TO Q58
- -2 ☐ SE NEGÓ A CONTESTAR → GO TO Q58
- -3 ☐ RESPUESTA POCO CLARA → GO TO Q58

57. En los últimos 3 meses, ¿Cree que {*los auxiliares de cuidados personales / el personal de salud mental*} supo(supieron) qué se incluyó en su [*término específico de cada programa que se refiere a un "plan de servicios"*], incluso las cosas que son importantes para usted?

- 1 ☐ SÍ
- 2 ☐ NO
- -1 ☐ NO SABE
- -2 ☐ SE NEGÓ A CONTESTAR
- -3 ☐ RESPUESTA POCO CLARA

58. En los últimos 3 meses, ¿con quién hubiera hablado si quisiera cambiar su [*término específico de cada programa que se refiere a un "plan de servicios*"]? ¿Hablaría con alguien más? [INTERVIEWER MARKS ALL THAT APPLY]

- 1 ☐ ENCARGADO DE CASO
- 2 ☐ OTROS MIEMBROS DEL PERSONAL
- 3 ☐ FAMILIARES/ AMIGOS
- 4 ☐ ALGUIEN MÁS, ESPECIFIQUE _____
- -1 ☐ NO SABE
- -2 ☐ SE NEGÓ A CONTESTAR
- -3 ☐ RESPUESTA POCO CLARA

Transporte

El tema de las siguientes preguntas es cómo va usted a sitios de su comunidad.

59. Entre las citas médicas se incluye ir a ver al doctor, al dentista, al terapeuta o a otra persona que se encargue del cuidado de su salud. En los últimos 3 meses, ¿con qué frecuencia tuvo forma de llegar a sus citas médicas? ¿Diría que...?

- 1 ☐ Nunca,
- 2 ☐ A veces,
- 3 ☐ Casi siempre, o
- 4 ☐ Siempre?
- -1 ☐ NO SABE
- -2 ☐ SE NEGÓ A CONTESTAR
- -3 ☐ RESPUESTA POCO CLARA

Versión Alternativa: Entre las citas médicas se incluye ir a ver al doctor, al dentista, al terapeuta o a otra persona que se encargue del cuidado de su salud. En los últimos 3 meses, ¿tuvo forma de llegar a sus citas médicas? ¿Diría que...?

1 ☐ En general, sí, o
2 ☐ En general, no?
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

60. En los últimos 3 meses, ¿usó una camioneta van o algún otro servicio de transporte? No incluya una camioneta van que le pertenezca a usted.

1 ☐ SÍ
2 ☐ NO → GO TO Q63
-1 ☐ NO SABE → GO TO Q63
-2 ☐ SE NEGÓ A CONTESTAR → GO TO Q63
-3 ☐ RESPUESTA POCO CLARA → GO TO Q63

61. En los últimos 3 meses, ¿pudo subirse y bajarse de este vehículo fácilmente?

1 ☐ SÍ
2 ☐ NO
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

62. En los últimos 3 meses, ¿con qué frecuencia llegó este vehículo a tiempo a recogerlo(a)? ¿Diría que…?

1 ☐ Nunca,
2 ☐ A veces,
3 ☐ Casi siempre, o
4 ☐ Siempre?
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

Versión Alternativa: En los últimos 3 meses, ¿llegó este vehículo a tiempo a recogerlo(a)? ¿Diría que…?

1 ☐ En general, sí, o
2 ☐ En general, no?
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

Seguridad personal

Las siguientes preguntas se refieren a su seguridad personal.

63. En los últimos 3 meses, ¿con quién se comunicaría en caso de emergencia? [INTERVIEWER MARKS ALL THAT APPLY]

1 ☐ PARIENTE O AMIGO

2 ☐ ENCARGADO DE CASO
3 ☐ AGENCIA
4 ☐ SERVICIO DE EMERGENCIA PAGADOS (EJEMPLO LIFELINE)
5 ☐ 911/ PERSONAL DE PRIMEROS AUXILIOS (POLICIA, ETC)
6 ☐ ALGUIEN MÁS, ESPECIFIQUE _____
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

64. En los últimos 3 meses, ¿Hubo una persona con quien podía hablar si alguien lo(a) lastimó o le hizo algo que a usted no le gustó?

1 ☐ SÍ
2 ☐ NO
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

Las siguientes preguntas se refieren a si alguna persona a quien se le paga para ayudarle lo(a) tratado mal en los últimos 3 meses. Esto incluye a *{personal assistance/behavioral health staff, homemakers, or your case manager}.* Les estamos haciendo a todos las siguientes preguntas, no solo a usted. [ADD STATE-SPECIFIC LANGUAGE HERE REGARDING MANDATED REPORTING, IF APPROPRIATE: Quiero recordarle que, aunque sus respuestas son confidenciales, tengo la responsabilidad legal de informarle al estado de *{STATE}* si oigo algo que me haga pensar que alguien lo(a) está lastimando o que usted está en peligro.]

65. En los últimos 3 meses, ¿**alguno** de {*los auxiliar(es) de cuidados personales, el personal de salud mental, los ayudantes de oficios domésticos o los encargados de caso*} tomó su dinero o sus cosas sin preguntarle primero?

- [1] ☐ SÍ
- [2] ☐ NO → GO TO Q68
- [-1] ☐ NO SABE → GO TO Q68
- [-2] ☐ SE NEGÓ A CONTESTAR → GO TO 68
- [-3] ☐ RESPUESTA POCO CLARA → GO TO Q68

66. En los últimos 3 meses, ¿alguien estuvo colaborando con usted para solucionar este problema?

- [1] ☐ SÍ
- [2] ☐ NO → GO TO Q68
- [-1] ☐ NO SABE → GO TO Q68
- [-2] ☐ SE NEGÓ A CONTESTAR → GO TO Q68
- [-3] ☐ RESPUESTA POCO CLARA → GO TO Q68

67. En los últimos 3 meses, ¿quién colaboró con usted para solucionar este problema? ¿Alguna otra persona? [INTERVIEWER MARKS ALL THAT APPLY]

- [1] ☐ PARIENTE O AMIGO
- [2] ☐ ENCARGADO DE CASO
- [3] ☐ AGENCIA
- [4] ☐ ALGUIEN MÁS, ESPECIFIQUE _____
- [-1] ☐ NO SABE
- [-2] ☐ SE NEGÓ A CONTESTAR
- [-3] ☐ RESPUESTA POCO CLARA

68. En los últimos 3 meses, ¿algún {*empleado*} a le gritó, lo(a) insultó o le dijo malas palabras?

- [1] ☐ SÍ
- [2] ☐ NO → GO TO Q71
- [-1] ☐ NO SABE → GO TO Q71
- [-2] ☐ SE NEGÓ A CONTESTAR → GO TO Q71
- [-3] ☐ RESPUESTA POCO CLARA → GO TO Q71

69. En los últimos 3 meses, ¿alguien colaboró con usted para solucionar este problema?

- [1] ☐ SÍ
- [2] ☐ NO → GO TO Q71
- [-1] ☐ NO SABE → GO TO Q71
- [-2] ☐ SE NEGÓ A CONTESTAR → GO TO Q71
- [-3] ☐ RESPUESTA POCO CLARA → GO TO Q71

70. En los últimos 3 meses, ¿quién estuvo colaborando con usted para solucionar este problema? ¿Alguna otra persona? [INTERVIEWER MARKS ALL THAT APPLY]

- [1] ☐ PARIENTE O AMIGO
- [2] ☐ ENCARGADO DE CASO
- [3] ☐ AGENCIA
- [4] ☐ ALGUIEN MÁS, ESPECIFIQUE _____
- [-1] ☐ NO SABE
- [-2] ☐ SE NEGÓ A CONTESTAR
- [-3] ☐ RESPUESTA POCO CLARA

71. En los últimos 3 meses, ¿algún {*empleado*} le pegó o lo(a) lastimó?

- [1] ☐ SÍ
- [2] ☐ NO → GO TO Q74
- [-1] ☐ NO SABE → GO TO Q74
- [-2] ☐ SE NEGÓ A CONTESTAR → GO TO Q74
- [-3] ☐ RESPUESTA POCO CLARA → GO TO Q74

72. En los últimos 3 meses, ¿alguien colaboró con usted para solucionar este problema?

- [1] ☐ SÍ
- [2] ☐ NO → GO TO Q74
- [-1] ☐ NO SABE → GO TO Q74
- [-2] ☐ SE NEGÓ A CONTESTAR → GO TO Q74
- [-3] ☐ RESPUESTA POCO CLARA → GO TO Q74

73. En los últimos 3 meses, ¿quién estuvo colaborando con usted para solucionar este problema? ¿Alguna otra persona? [INTERVIEWER MARKS ALL THAT APPLY]

- [1] ☐ PARIENTE O AMIGO
- [2] ☐ ENCARGADO DE CASO
- [3] ☐ AGENCIA
- [4] ☐ ALGUIEN MÁS, ESPECIFIQUE _____
- [-1] ☐ NO SABE
- [-2] ☐ SE NEGÓ A CONTESTAR
- [-3] ☐ RESPUESTA POCO CLARA

Comunidad y empoderamiento

Ahora me gustaría preguntarle sobre las cosas que hace en su comunidad.

74. ¿Tiene **familiares** que vivan cerca? No incluya a los miembros de la familia con los que vive.

- [1] ☐ SÍ
- [2] ☐ NO → GO TO Q76
- [-1] ☐ NO SABE → GO TO Q76
- [-2] ☐ SE NEGÓ A CONTESTAR → GO TO Q76
- [-3] ☐ RESPUESTA POCO CLARA → GO TO Q76

75. En los últimos 3 meses, cuando usted lo deseó, ¿con qué frecuencia pudo reunirse con estos familiares que viven cerca? ¿Diría que…?

- [1] ☐ Nunca,
- [2] ☐ A veces,
- [3] ☐ Casi siempre, o
- [4] ☐ Siempre?
- [-1] ☐ NO SABE
- [-2] ☐ SE NEGÓ A CONTESTAR
- [-3] ☐ RESPUESTA POCO CLARA

Versión Alternativa: En los últimos 3 meses, cuando usted lo deseó, ¿pudo reunirse con estos familiares que viven cerca? ¿Diría que…?

- [1] ☐ En general, sí, o
- [2] ☐ En general, no?
- [-1] ☐ NO SABE
- [-2] ☐ SE NEGÓ A CONTESTAR
- [-3] ☐ RESPUESTA POCO CLARA

76. ¿Tiene **amigos** que vivan cerca?

- [1] ☐ SÍ
- [2] ☐ NO → GO TO Q78
- [-1] ☐ NO SABE → GO TO Q78
- [-2] ☐ SE NEGÓ A CONTESTAR → GO TO Q78
- [-3] ☐ RESPUESTA POCO CLARA → GO TO Q78

77. En los últimos 3 meses, cuando usted lo deseó, ¿con qué frecuencia pudo reunirse con estos amigos que viven cerca? ¿Diría que…?

- [1] ☐ Nunca,
- [2] ☐ A veces,
- [3] ☐ Casi siempre, o
- [4] ☐ Siempre?
- [-1] ☐ NO SABE
- [-2] ☐ SE NEGÓ A CONTESTAR
- [-3] ☐ RESPUESTA POCO CLARA

Versión Alternativa: En los últimos 3 meses, cuando usted lo deseó, ¿pudo reunirse con estos amigos que viven cerca? ¿Diría que…?

- [1] ☐ En general, sí, o
- [2] ☐ En general, no?
- [-1] ☐ NO SABE
- [-2] ☐ SE NEGÓ A CONTESTAR
- [-3] ☐ RESPUESTA POCO CLARA

78. En los últimos 3 meses, cuando usted lo deseó, ¿con qué frecuencia pudo hacer lo que le gusta en la comunidad? ¿Diría que…?

- [1] ☐ Nunca,
- [2] ☐ A veces,
- [3] ☐ Casi siempre, o

4 ☐ Siempre?
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

Versión Alternativa: En los últimos 3 meses, cuando usted lo deseó, ¿pudo hacer lo que le gusta en la comunidad? ¿Diría que…?

1 ☐ En general, sí, o
2 ☐ En general, no?
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

79. En los últimos 3 meses, ¿necesitó más ayuda de la que recibe de {*el personal de salud mental / los auxiliares de cuidados personales}* para hacer cosas en su comunidad?

1 ☐ SÍ
2 ☐ NO
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

80. En los últimos 3 meses, ¿participó en decidir **qué** hace cada día?

1 ☐ SÍ
2 ☐ NO
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

81. En los últimos 3 meses, ¿participó en decidir el **horario** de sus actividades de cada día? Por ejemplo, cuándo se levanta, cuándo come o cuándo se acuesta.

1 ☐ SÍ
2 ☐ NO
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

Sobre usted

Ahora tengo unas cuantas preguntas sobre usted.

82. En general, ¿cómo calificaría toda su salud? ¿Diría que…?

1 ☐ Excelente,
2 ☐ Muy buena,
3 ☐ Buena,
4 ☐ Regular, o
5 ☐ Mala?
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR

-3 ☐ RESPUESTA POCO CLARA

83. En general, ¿cómo calificaría toda su salud mental o emocional? ¿Diría que…?

1 ☐ Excelente,
2 ☐ Muy buena,
3 ☐ Buena,
4 ☐ Regular, o
5 ☐ Mala?
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

84. ¿Qué edad tiene?

1 ☐ ENTRE 18 Y 24 AÑOS
2 ☐ ENTRE 25 Y 34 AÑOS
3 ☐ ENTRE 35 Y 44 AÑOS
4 ☐ ENTRE 45 Y 54 AÑOS
5 ☐ ENTRE 55 Y 64 AÑOS
6 ☐ ENTRE 65 Y 74 AÑOS
7 ☐ 75 AÑOS O MÁS
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

Versión Alternativa: ¿En qué año nació?

_____ (AÑO)

-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

85. [IF NECESSARY, ASK, AND VERIFY IF OVER THE PHONE] ¿Es usted hombre o mujer?

1 ☐ Hombre
2 ☐ Mujer
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

86. ¿Cuál es el grado o nivel escolar más alto que usted ha completado?

1 ☐ 8 años de escuela o menos
2 ☐ 9 a 12 años de escuela, pero sin graduarse,
3 ☐ Graduado de la escuela secundaria (high school), Diploma de escuela secundaria preparatoria, o su equivalente (o GED)
4 ☐ Algunos cursos universitarios o un título universitario de un programa de 2 años
5 ☐ Título universitario de 4 años
6 ☐ Título universitario de más de 4 años
-1 ☐ NO SABE
-2 ☐ SE NEGÓ A CONTESTAR
-3 ☐ RESPUESTA POCO CLARA

87. ¿Es de origen o ascendencia hispana o latina o española?

- [1] ☐ Sí, hispano(A), latino(A) o ESPAÑOL(A)
- [2] ☐ No, ni hispano(A) ni latino(A) ni ESPAÑOL(A) → GO TO Q89
- [-1] ☐ NO SABE → GO TO Q89
- [-2] ☐ SE NEGÓ A CONTESTAR → GO TO Q89
- [-3] ☐ RESPUESTA POCO CLARA → GO TO Q89

88. ¿Qué grupo lo describe mejor? [READ ALL ANSWER CHOICES. CODE ALL THAT APPLY.]

- [1] ☐ Mexicano, mexicano americano, chicano
- [2] ☐ Puertorriqueño
- [3] ☐ Cubano
- [4] ☐ De otro origen hispano, latino o español
- [-1] ☐ NO SABE
- [-2] ☐ SE NEGÓ A CONTESTAR
- [-3] ☐ RESPUESTA POCO CLARA

89. ¿A qué raza pertenece? Puede escoger una o más de las siguientes. ¿Diría que es…?

- [1] ☐ Blanco(a) → GO TO Q92
- [2] ☐ Negro(a) o afroamericano(a) → GO TO Q92
- [3] ☐ Asiático(a) → GO TO Q90
- [4] ☐ Nativo(a) de Hawái o de otras islas del Pacifico → GO TO Q91
- [5] ☐ Indígena americano(a) o nativo(a) de Alaska ➔ GO TO Q92
- [6] ☐ OTRO → GO TO Q92
- [-1] ☐ NO SABE → GO TO Q92
- [-2] ☐ SE NEGÓ A CONTESTAR → GO TO Q92
- [-3] ☐ RESPUESTA POCO CLARA → GO TO Q92

90. ¿Qué grupo lo describe mejor? [READ ALL ANSWER CHOICES. CODE ALL THAT APPLY.]

- [1] ☐ Indio asiático → GO TO Q92
- [2] ☐ Chino → GO TO Q92
- [3] ☐ Filipino → GO TO Q92
- [4] ☐ Japonés → GO TO Q92
- [5] ☐ Coreano → GO TO Q92
- [6] ☐ Vietnamita → GO TO Q92
- [7] ☐ Otra asiático → GO TO Q92
- [-1] ☐ NO SABE → GO TO Q92
- [-2] ☐ SE NEGÓ A CONTESTAR → GO TO Q92
- [-3] ☐ RESPUESTA POCO CLARA → GO TO Q92

91. ¿Qué grupo lo describe mejor? READ ALL ANSWER CHOICES. CODE ALL THAT APPLY.

- [1] ☐ Nativo de Hawái
- [2] ☐ Guameño o chamorro
- [3] ☐ Samoano
- [4] ☐ Nativa de otras islas del Pacífico
- [-1] ☐ NO SABE
- [-2] ☐ SE NEGÓ A CONTESTAR
- [-3] ☐ RESPUESTA POCO CLARA

92. ¿Habla algún otro idioma aparte de español en casa? READ CHOICES ONLY IF NEEDED…

- [1] ☐ SÍ
- [2] ☐ No → GO TO Q94
- [-1] ☐ NO SABE → GO TO Q94
- [-2] ☐ SE NEGÓ A CONTESTAR → GO TO Q94
- [-3] ☐ RESPUESTA POCO CLARA → GO TO Q94

93. ¿Qué idioma habla usted en casa?

- [1] ☐ Inglés
- [2] ☐ Español
- [3] ☐ Ambos: inglés y español
- [4] ☐ Español y otro idioma
- [5] ☐ Otro idioma → ¿Cuál? _____
- [-1] ☐ NO SABE
- [-2] ☐ SE NEGÓ A CONTESTAR
- [-3] ☐ RESPUESTA POCO CLARA

94. **[**IF NECESSARY, ASK**]** ¿Incluyendo a usted, cuantos adultos viven en su casa?

- [1] ☐ 1 [SOLO EL RESPONDENTE] → END SURVEY
- [2] ☐ ENTRE 2 A 3
- [3] ☐ 4 O MÁS
- [-1] ☐ NO SABE
- [-2] ☐ SE NEGÓ A CONTESTAR
- [-3] ☐ RESPUESTA POCO CLARA

95. **[**IF NECESSARY, ASK**]** ¿Vive con familiares?

- [1] ☐ SÍ
- [2] ☐ NO
- [-1] ☐ NO SABE
- [-2] ☐ SE NEGÓ A CONTESTAR
- [-3] ☐ RESPUESTA POCO CLARA

96. [IF NECESSARY, ASK] ¿Vive con personas que no son de su familia ni tienen ningún parentesco con usted?

- [1] ☐ SÍ
- [2] ☐ NO
- [-1] ☐ NO SABE
- [-2] ☐ SE NEGÓ A CONTESTAR
- [-3] ☐ RESPUESTA POCO CLARA

Interviewer Questions

LAS SIGUIENTES PREGUNTAS DBERAN CONTESTARSE DESPUES DE LA ENTREVISTA.

97. ¿EL ENTREVISTADO PUDO DAR RESPUESTAS VÁLIDAS?

1 ☐ SÍ
2 ☐ NO

98. ¿ESTUVO ALGUNA OTRA PERSONA PRESENTE DURANTE LA ENTREVISTA?

1 ☐ SÍ
2 ☐ NO → END SURVEY

99. ¿QUIÉN ESTUVO PRESENTE DURANTE LA ENTREVISTA? (MARQUE TODAS LAS OPCIONES QUE CORRESPONDAN)

1 ☐ ALGUIEN A QUIEN **NO** SE LE PAGA PARA QUE LE PROPORCIONE APOYO AL ENTREVISTADO
2 ☐ UN MIEMBRO DEL PERSONAL O ALGUIEN A QUIEN SE LE PAGA PARA QUE LE PROPORCIONE APOYO AL ENTREVISTADO

100. ¿ALGUIEN LE AYUDÓ AL ENTREVISTADO A RESPONDER ESTA ENCUESTA?

1 ☐ SÍ
2 ☐ NO → END SURVEY

101. ¿CÓMO LE AYUDÓ ESA PERSONA? MARQUE TODAS LAS OPCIONES QUE CORRESPONDAN.

1 ☐ RESPONDIÓ A TODAS LAS PREGUNTAS POR EL ENTREVISTADO
2 ☐ FORMULÓ LAS PREGUNTAS DE DIFERENTE MANERA O LE RECORDÓ / LE DIO PISTAS AL ENTREVISTADO
3 ☐ TRADUJO LAS PREGUNTAS O RESPUESTAS AL IDIOMA DEL ENTREVISTADO
4 ☐ AYUDÓ MEDIANTE EL USO DE UN EQUIPO DE ASISTENCIA O DE COMUNICACIONES PARA QUE EL ENTREVISTADO PUDIERA RESPONDER A LAS PREGUNTAS
5 ☐ OTRA (ESPECIFIQUE) _____

102. ¿QUIÉN LE AYUDÓ AL ENTREVISTADO? (MARQUE TODAS LAS OPCIONES QUE CORRESPONDAN)

1 ☐ ALGUIEN A QUIEN **NO** SE LE PAGA PARA QUE LE PROPORCIONE APOYO AL ENTREVISTADO
2 ☐ UN MIEMBRO DEL PERSONAL O ALGUIEN A QUIEN SE LE PAGA PARA QUE LE PROPORCIONE APOYO AL ENTREVISTADO

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 subcriterion 1b).

Brief Measure Information

NQF #: 2612
De.2. **Measure Title:** CARE: Improvement in Mobility
Co.1.1. **Measure Steward:** American Health Care Association
De.3. **Brief Description of Measure:** The measure calculates a skilled nursing facility's (SNFs) average change in mobility for patients admitted from a hospital who are receiving therapy. The measure calculates the average change in mobility score between admission and discharge for all residents admitted to a SNF from a hospital or another post-acute care setting for therapy (i.e., PT or OT) regardless of payor status. This is a risk adjusted outcome measure, based on the mobility subscale of the Continuity Assessment and Record Evaluation (CARE) Tool and information from the admission MDS 3.0 assessment. The measure is calculated on a rolling 12 month, average updated quarterly.
1b.1. **Developer Rationale:** Therapies in skilled nursing facilities (SNFs) serve a critical role in helping individuals achieve and maintain maximum physical and functional well-being following hospitalization. One of the principle reasons individuals are discharged from a hospital to a SNF is to improve their functional abilities so they can return to their prior living situation. The Medicare SNF Part A benefit is designed principally for the purpose of either providing therapy to individuals who were hospitalized or to complete their nursing care in a less intensive setting. Over 85% of all Medicare Beneficiaries admitted to SNFs receive therapy services. The level of improvement in function (self care and mobility) is a strong predictor of a person's ability to reside in the community independently (Granger et al., 1979).

Currently, there is no standard assessment tool used across all post-acute care (PAC) providers. Many of the current therapy companies have developed their own or use existing therapy related assessment tools; none of these assessments are standardized. This lack of consistency prevents industry wide comparison. Consumers need a standard measure in order to make educated choices in selecting high quality care providers to maximize outcomes. Educated consumers who make decisions regarding care based on validated measures will also promote accountability and quality among PAC providers. PAC providers, in turn, need standardized assessments to evaluate the quality of services provided, identify areas where improvement is needed, and present data to demonstrate the value of therapy services. Research by Standardized outcome assessments, questionnaires or tools are a vital part of evidence-based practice. Selecting the most appropriate outcomes assessment, questionnaire or tool enhances clinical practice by (1) identifying and quantifying areas related to mobility; (2) formulating the evaluation, diagnosis, and prognosis; (3) designing the plan of care; and (4) helping to evaluate the success of physical and occupational therapy interventions.

Potter, K., Fulk, G.D., Salem, Y., Sullivan, J. "Outcome Measures in Neurological Physical Therapy Practice: Part I. Making Sound Decisions". Journal of Neurologic Physical Therapy (2011): 57-64.

Grander, C.V., Dewis, L.S., Peters, N.C., Sherwood, C.C. & Barrett, J.E. (1979). "Stroke rehabilitation: Analysis of repeated Barthel index measures". Arch Phys Med Rehabil 60.1 (1979): 14-7.

S.4. **Numerator Statement:** The measure assesses the change in mobility. The numerator is the risk adjusted sum of the change in the CARE Tool mobility subscale items between admission and discharge for each individual admitted from a hospital or another post acute care setting regardless of payor status and are receiving therapy (PT or OT) for any reason in a skilled nursing center.
S.7. **Denominator Statement:** The denominator includes all residents admitted to a SNF from a hospital or another post-acute care setting who receive either PT or OT therapy for any reason during their stay regardless of payor status, have a completed mobility CARE tool assessment at admission and discharge and do not meet any of the exclusion criteria. The mobility items used from the CARE tool are listed below and rated on a 1-6 scale (see Appendix for copy of the mobility CARE tool assessment).

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 lessthan-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
Mobility_Evidence_Submission_Form_2612.docx

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., the benefits or improvements in quality envisioned by use of this measure)*
Therapies in skilled nursing facilities (SNFs) serve a critical role in helping individuals achieve and maintain maximum physical and functional well-being following hospitalization. One of the principle reasons individuals are discharged from a hospital to a SNF is to improve their functional abilities so they can return to their prior living situation. The Medicare SNF Part A benefit is designed principally for the purpose of either providing therapy to individuals who were hospitalized or to complete their nursing care in a

The items included in the CARE Tool Mobility subscale include:

- B1. Lying to Sitting on Side of Bed
- B2. Sit to Stand
- B3. Chair/Bed to Chair Transfer
- B4. Toilet Transfer
- B5a & B5b. Walking or Wheelchair Mobility
- C3. Roll left / right
- C4. Sit to Lying
- C5. Picking up object
- C7a. One Step Curb
- C7b. Walk 50 ft. with Two Turns
- C7c. Walk 12 Steps.
- C7d. Walk Four Steps
- C7e. Walking 10 ft. on Uneven Surface
- C7f. Car Transfer

S.10. **Denominator Exclusions:** Patients are excluded for two broad reasons:

1. if they have conditions where improvement in mobility is very unlikely,
OR
2. have missing data necessary to calculate the measure

Additionally, facilities with denominator size of fewer than 30 patients during a 12 month period are excluded from reporting their data.

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Electronic Clinical Data, Other
S.26. **Level of Analysis:** Facility

IF Endorsement Maintenance – Original Endorsement Date: Jul 23, 2015 **Most Recent Endorsement Date:** Jul 23, 2015

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

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

De.4. **IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** Not Applicable

less intensive setting. Over 85% of all Medicare Beneficiaries admitted to SNFs receive therapy services. The level of improvement in function (self care and mobility) is a strong predictor of a person's ability to reside in the community independently (Granger et al., 1979).

Currently, there is no standard assessment tool used across all post-acute care (PAC) providers. Many of the current therapy companies have developed their own or use existing therapy related assessment tools; none of these assessments are standardized. This lack of consistency prevents industry wide comparison. Consumers need a standard measure in order to make educated choices in selecting high quality care providers to maximize outcomes. Educated consumers who make decisions regarding care based on validated measures will also promote accountability and quality among PAC providers. PAC providers, in turn, need standardized assessments to evaluate the quality of services provided, identify areas where improvement is needed, and present data to demonstrate the value of therapy services. Research by Standardized outcome assessments, questionnaires or tools are a vital part of evidence-based practice. Selecting the most appropriate outcomes assessment, questionnaire or tool enhances clinical practice by (1) identifying and quantifying areas related to mobility; (2) formulating the evaluation, diagnosis, and prognosis; (3) designing the plan of care; and (4) helping to evaluate the success of physical and occupational therapy interventions.

Potter, K., Fulk, G.D., Salem, Y., Sullivan, J. "Outcome Measures in Neurological Physical Therapy Practice: Part I. Making Sound Decisions". Journal of Neurologic Physical Therapy (2011): 57-64.

Grander, C.V., Dewis, L.S., Peters, N.C., Sherwood, C.C. & Barrett, J.E. (1979). "Stroke rehabilitation: Analysis of repeated Barthel index measures". Arch Phys Med Rehabil 60.1 (1979): 14-7.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. Include mean, 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Using data from Jan 2012 to Dec 2012 from two large therapy companies representing 360 SNFs and 60,146 matched (admissiondischarge) patient assessments. The frequency distribution of the SNFs' average mobility change scores are shown in Figure A3 in the Appendix. The mean = 26.8, std dev=2.1, min=18.4, max=31.0, 1st quartile = 25.4, 3rd quartile = 28.4, 1st decile = 23.9, 2nd decile = 25.1, 3rd decile = 25.7, 4th decile = 26.5, 5th decile = 27.1, 6th decile = 27.6, 7th decile = 28.1, 8th decile = 28.7, 9th decile = 29.2.

Additionally, table A1 in the Appendix shows the distribution of facilities and risk adjusted mobility change scores by facility bed count, ownership type, and by urban/rural location.

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 endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.* Specific data on disparities was not included in this measure as specified by current NQF requirements; however, due to the potential for widespread use of this measure we have included distribution of SNF admissions by gender, age and ethnicity in SNFs from the second quarter of 2014, based on the MDS 3.0 (see Appendix tables A2-A4). Nationally, 66% of all admissions to SNFs are female. Approximately three-quarters are between the ages of 65 and 95 years. Based on the MDS, the majority are considered white (76%) with 14% African American, 5% Hispanic, 2% Asian and less than 1% each native Hawaiian or other pacific islander and American Indian or Alaska native. A state by state break down is provide in the appendix. This makes stratification at a facility level extremely difficult because sample sizes for ethnic groups within a facility are small and frequently below the minimum denominator size of 30.

We are not able to present information on insurance status based on the MDS, as it is not reliable due to the accuracy of the information submitted by providers, the ambiguity of payer status at admission, the number of patients with multiple payers and patient's whose payor status changes during the course of care in the SNF.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.

A PubMed search on disparities related to therapy outcomes in skilled nursing facility residents did not produce any meaningful results. There is some evidence that suggest differences in access and utilization of post-acute rehabilitation care by ethnicity but none on differences in the quality of care delivered or outcomes within a single provider. A patient can receive post-acute rehabilitation care in inpatient rehabilitation facilities, skilled nursing homes, or through home health care; the former provides more hours of care than the latter when viewed on a continuum. Freburger et al. (2012) found that minorities and those with lower socioeconomic statuses receive lower volumes of rehabilitation care. These individuals are more likely to be discharged home and receive care in a SNF verses an inpatient rehabilitation facility. This finding is also supported by earlier research showing that racial minorities, women, older individuals, and those with lower incomes are more likely to receive care in SNFs or home health (Freburger et al., 2011). However, no studies looked at difference in volume of therapy services by ethnicity within SNFs just across different types of PAC providers.

There is evidence that racial disparities exist in care provided in different nursing homes. An article by Smith et al. (2007) suggests that racial segregation in nursing homes mirrors that which occurs in metropolitan areas. Black nursing home residents are 1.41 times more likely to be in a facility cited with a deficiency causing actual harm or immediate jeopardy to residents. Forty percent of African American patients are in lower tier facilities, those with higher number of Medicaid patients and limited resources, compared to nine percent of white residents. The lower tier facilities are shown to have fewer nurses, lower occupancy rates, and more health-related deficiencies (Mor et al., 2004). However, the outcomes of these individuals did not differ from other residents in the same facility. Thus, suggesting differences are related to the facility's location and practice not differences related to ethnicity or social economic status of the residents. A 2013 study also found that the differences in quality between SNFs with higher proportion of African American residents was mediated by the overall financial health of the facility and overall quality in the facility, rather than the racial mix (Chisholm et al., 2103). In summary, the literature suggests that ethnic and social economic status differences are related to inter-facility differences not to intra-facility differences in care. Therefore, it is unclear based on the literature if social economic status should be risk adjusted.

Chisholm, L., Weech-Maldonado, R., Laberge, A., Lin, F. & Hyer, K. "Nursing home quality and financial performance: Does the racial composition of residents matter?" Health Services Research. (2013): 2060- 2080.

Fennell, M. L., Miller, S. C., & Mor, V. "Facility effects on racial differences in nursing home quality of care". American Journal of Medical Quality. 15.4 (2000): 174-181.

Freburger, J.K., Holmes, G.M., & Ku, L.J. "Postacute rehabilitation care for hip fracture: Who gets the most care?". J Am Geriatr Soc. 60.10 (2012):1929-1935.

Freburger, J.K., Holmes, G.M., Ku, L.J., Cutchin, M.P., Heatwole-Shank, K. & Edwards, L.J. "Disparities in postacute rehabilitation care for stroke: An analysis of the state inpatient databases". Arch Phys Med Rehabil. 92.8 (2011): 1220-1229.

Grabowski, D.C. "The admission of blacks to high-deficiency nursing homes". Medical Care. 42.5 (2004): 456-464.

Harada, N.D., Chun, A., Chiu, V. & Pakalniskis, A. "Patterns of rehabilitation utilization after hip fracture in acute hospitals and skilled nursing facilities". Medical Care. 38.11 (2000): 1119-1130.

Holmes, G.M., Freburger, J.K. & Ku, L.J. "Decomposing racial and ethnic disparities in the use of postacute rehabilitation care". Health Serv Res. 47.3 (2012): 1158-1178.

McCallum, C.A. "Access to physical therapy services among medically underserved adults: A mixed-method study". Phys Ther. 90.5 (2010):735-747.

Mor, V., Zinn, J., Angelelli, J., Teno, J.M., & Miller, S.C. "Driven to tiers: Socioeconomic and racial disparities in the quality of nursing home care". The Milbank Quarterly. 82.2 (2004): 227-156.

Smith, D.B., Fang, Z., Fennell, M.L, Zinn, J.S. & Mor, V. "Separate and unequal: Racial segregation and disparities in quality across U.S. nursing homes". Health Affairs. 26.5 (2007): 1448-1458.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare

Affects large numbers, Patient/societal consequences of poor quality

1c.2. If Other:

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.

Nearly 2.5 million Medicare beneficiaries in Fee for Service receive physical and/or occupational therapy each year in the skilled nursing setting. According to the 2012 AHCA Annual Report, the skilled nursing setting received over 2.3 million FFS Medicare admissions and over half a million non-Medicare admissions in 2012. Of those Medicare admissions, 86.4% received occupational therapy and 89.6% received physical therapy. Of the non-Medicare admissions, 54.4% received occupational therapy and 58.5% received physical therapy.

Medicare FFS spending on SNF services was $28.7 billion in 2012. On a per user basis, spending per nursing home resident averaged $31,735 in 2010 (MedPac, 2014). Discharges for rehabilitation assume that patients will improve in functionality, however there is no consistent measurement showing whether or not patients improve. It is vital to be able to show the benefits provided to patients through these services.

On average, about 57% of all individuals admitted to a skilled nursing facility for rehabilitation services are discharged to home. Rehabilitation services, including the intensity and appropriateness of services, have strong relationships with patient functional gains, community discharge and survival. These services are vital to the quality of life of the patients receiving them. The level of improvement in function (Self-care and mobility) is a strong predictor of a person's ability to reside in the community independently (Grander et al., 1979).

The CARE Tool was developed by CMS based on a modification of the existing Barthel Index (Gage et al., 2012). The Barthel Index was developed in the 1950's, has been modified and adapted by numerous researchers. It is one of the most widely used measures of functional status and as a result, has been repeatedly tested for reliability, validity and sensitivity. The Barthel Index has been shown be highly reliable with lower administrative burden than other widely used tools. The Barthel Index has shown high interrater reliability when performed by a therapist versus a nurse assessor. The CARE tool was found to have equally high reliability and validity when tested by CMS in SNFs and other post-acute care settings (Gage et al., 2012).

1c.4. Citations for data demonstrating high priority provided in 1a.3

American Health Care Association. "2013 Quality Report". 2013. http://www.ahcancal.org/quality_improvement/Documents/AHCA%20Quality%20Report%20FINAL.pdf.

Cohen, M.E., Marino, R.J. "The Tools of Disability Outcomes Research Functional Status Measures". Arch Phys Med Rehabil. 81 (2000): S21-S29.

Gage, G., Morley, M., Smith, L., Ingber, M.J., Deutsch, A., Kline, T., … Manning, J. "Post-acute care payment reform demonstration: Final report". RTI International. 2 (2012).

Grander, C.V., Dewis, L.S., Peters, N.C., Sherwood, C.C. & Barrett, J.E. "Stroke rehabilitation: Analysis of repeated Barthel index measures". Arch Phys Med Rehabil. 60.1 (1979): 14-7.

Jette, D.U., Warren, R.L., Wirtalla, C. "The Relation Between Therapy Intensity and Outcomes of Rehabilitation in Skilled Nursing Facilities". Arch Phys Med Rehabil. 86 (2005): 373-379.

MedPac. "Report to Congress: Payment policy". 2014. http://www.medpac.gov/documents/reports/mar14_entirereport.pdf?sfvrsn=0.

Murray, P.K., Singer, M., Dawson, N.V., Thomas, C.L., & Cebul, R.D. "Outcomes of Rehabilitation Services for Nursing Home Residents". Arch Phys Med Rehabil. 84 (2003): 1129-1136.

Richards, S.H., Peters, T.J., Coast, J., Gunnel, D.J., Darlow, M., Pounsford, J. "Inter-Rater Reliability of the Barthel ADL Index: How Does a Researcher Compare to a Nurse?". Clinical Rehabilitation. 2000 (1999): 72-28.

Sangha, H., Lipson, D., Foley, N., Salter, K., Bhogal, S., Pohani, G., Teasell, R.W. "A comparison of the Barthel Index and the Functional Independence Measure as outcome measures in stroke rehabilitation: patterns of disability scale usage in clinical trials". International Journal of Rehabilitation Research. 28.2 (2005): 135-139.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)* Not Applicable

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 subcriteria 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. **Cross Cutting Areas** *(check all the areas that apply):*
Functional Status, Health and Functional Status : Functional Status

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.)*
Not Applicable

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.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons. Not Applicable

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
The measure assesses the change in mobility. The numerator is the risk adjusted sum of the change in the CARE Tool mobility subscale items between admission and discharge for each individual admitted from a hospital or another post acute care setting regardless of payor status and are receiving therapy (PT or OT) for any reason in a skilled nursing center.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)* Rolling 12 month average, updated quarterly.

S.6. **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, 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.*

The numerator includes all residents admitted from a hospital or another post acute care setting that receive any PT or OT therapy for any reason in a SNF that have a completed mobility CARE tool assessment at admission and discharge (see denominator definition below). The mobility items used from the CARE tool are listed below and rated on a 1-6 scale (see Appendix for copy of the CARE Tool assessment).

The items included in the CARE Tool Mobility subscale include:

- B1. Lying to Sitting on Side of Bed
- B2. Sit to Stand
- B3. Chair/Bed to Chair Transfer
- B4. Toilet Transfer
- B5a & B5b. Walking or Wheelchair Mobility
- C3. Roll left / right
- C4. Sit to Lying
- C5. Picking up object
- C7a. One Step Curb
- C7b. Walk 50 ft. with Two Turns
- C7c. Walk 12 Steps.
- C7d. Walk Four Steps
- C7e. Walking 10 ft. on Uneven Surface
- C7f. Car Transfer

The numerator is a facility's average risk adjusted change score on the mobility component of the CARE tool. The risk adjusted average change score is calculated in several steps:

Step 1: Each individual's admission and discharge mobility scale score is calculated. Items rated as S. Not attempted due to safety concerns, A. Task attempted but not completed, N. Not applicable and P. Patient Refused were recoded to one. For each individual, the ratings for all the mobility items on the CARE tool at admission are summed and transformed to a 0-100 scale. The same is done for the discharge assessment.

Step 2: Each individual's unadjusted change score is calculated by taking the admission score minus the discharge score.

Step 3: The individual's unadjusted change score is risk adjusted (see risk adjustment section)

Step 4: The facilities risk adjusted change score is the sum of all the individual's risk adjusted change scores divided by the denominator.

S.7. Denominator Statement *(Brief, narrative description of the target population being measured)*

The denominator includes all residents admitted to a SNF from a hospital or another post-acute care setting who receive either PT or OT therapy for any reason during their stay regardless of payor status, have a completed mobility CARE tool assessment at admission and discharge and do not meet any of the exclusion criteria. The mobility items used from the CARE tool are listed below and rated on a 1-6 scale (see Appendix for copy of the mobility CARE tool assessment).

The items included in the CARE Tool Mobility subscale include:

- B1. Lying to Sitting on Side of Bed
- B2. Sit to Stand

- B3. Chair/Bed to Chair Transfer
- B4. Toilet Transfer
- B5a & B5b. Walking or Wheelchair Mobility
- C3. Roll left / right
- C4. Sit to Lying
- C5. Picking up object
- C7a. One Step Curb
- C7b. Walk 50 ft. with Two Turns
- C7c. Walk 12 Steps.
- C7d. Walk Four Steps
- C7e. Walking 10 ft. on Uneven Surface
- C7f. Car Transfer

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk : Dual eligible beneficiaries, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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 denominator includes all residents admitted to a SNF who are receiving any PT or OT therapy for any reason.

The denominator is based on admission from any hospital or post-acute care setting and is determined using information from MDS item "A1800 Entered From" coded as "03 Acute Care Hospital" or "02 Another nursing home or swing bed" or "05 inpatient rehabilitation facility" or "09 Long Term Care Hospital" regardless of payor status. They must receive either PT or OT therapy during their stay. A resident's stay is defined as an episode of care from admissions to discharge from the facility or discharge from therapy services (defined as completing a discharge CARE tool assessment). Overall, approximately 85% of all admissions from a hospital receive either PT or OT therapy based on SNF Part A claims (or MDS 3.0 data).

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)* Patients are excluded for two broad reasons:

1. if they have conditions where improvement in mobility is very unlikely,
OR
2. have missing data necessary to calculate the measure

Additionally, facilities with denominator size of fewer than 30 patients during a 12 month period are excluded from reporting their data.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
Individuals with conditions where improvement in mobility (as determined by a panel of expert therapists) is very unlikely were excluded based on information from the admission MDS 3.0 assessment. Individuals with one of the following MDS 3.0 items marked as yes were excluded:

- Ventilator (O0100F1 =1 or O0100F2 =1)
- Coma (B0100 =1)
- Quadriplegic (I5100=1)
- Hospice (O0100K1 = 1)

In addition, we also excluded individuals whose age is less than 18 years.
Overall, these exclusions resulted in 1.1% of all admissions being excluded.

Missing data also resulted in individuals being excluded

- Missing a discharge CARE Tool assessment (this resulted when individuals died or were hospitalized during their SNF stay)

resulted in patients being excluded since one could not calculate a change from admission. Nationally approximately 21.6% of admissions to a SNF will be hospitalized during their therapy stay and 4.5% will die (based on analysis of SNF part A claims from 2009-2011).

- Missing data on individual CARE Tool mobility assessment items on at least one item occurred 27.2% of the time. Approximately a third of all missing data related to just three items C7c walking 12 steps; C7d walking 4 steps and C7f car transfer but did not differ significantly between admission and discharge assessments. We did not impute any missing data for mobility items.

S.12. Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
Not Applicable

S.13. Risk Adjustment Type (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15) Statistical risk model If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Each individuals change score was risk adjusted based on the following formula:
Risk Adjusted Score for individual = (National Average Change Score – Predicted Change Score) + Actual Change Score.

The National Average Change Score was calculated as a population average change score for all patients in all SNFs who had a CARE Tool mobility subscale assessment completed at admission and discharge. The change score is the difference in the aggregate of each individuals scale score from admission to discharge transformed to 0 to 100 scale.

The Predicted Change Score is calculated based on logistic regression using the process outlined in 2b4.

The Actual Change Score is the difference between the individual person's admission mobility score transformed to 0 to 100 scale and their discharge mobility score transformed to a 0 to 100 scale.

S.15. Detailed risk model specifications *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Provided in response box S.15a

S.15a. Detailed risk model specifications (*if not provided in excel or csv file at S.2b*)
The variables used in the risk adjustment model all came from admission MDS 3.0 and include:

- Age ->85 years of age (calculated from date birth on admission MDS)
- Diabetes (MDS item I2900 = 01)
- Male (MDS item 00100K2 = 1)
- Admitted from a SNF (MDS item A180=02)
- Oxygen while a patient (MDS item –O0100C1 = 1 and O0100C2 =1)
- Catheterizations/ostomy (MDS Item H0100C = 1)
- Unhealed Pressure ulcer (MDS Item M0210= 1)

- Poor Mental Status defined as either a BIMS score for mental status of less than 12 or Impaired cognitive skills for decision making (MDS item C1000 = 3)

- Altered Resident Mood defined as either score for resident mood self-assessed of greater than 10 or summary score by staff assessment of resident mood of greater than 10

- Psychiatric Conditions defined as entered from psychiatric hospital (MDS item A1800= 04) or Behavioral symptoms of psychosis (MDS item E0100 = A or B)

- Requiring use of Feeding / IV defined as either having an IV Feeding within past 7 days (MDS item K0500 = A) or using a feeding Tube within past 7 days (MDS item K0500 = B)

- Requiring Suctioning or Tracheostomy defined as either Tracheostomy care (MDS item O0100 = E) or Suctioning (MDS item O0100 = D)

- Any infections/problems with feet defined as any of following: Infection of foot (MDS item M1040 = A); Other open lesions on the foot (MDS item M1040 = C) or Diabetic foot ulcers (MDS item M1040 = B).

S.16. Type of score:
Continuous variable, e.g. average
If other:

S.17. 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.18. Calculation Algorithm/Measure Logic *(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; aggregating data; risk adjustment; etc.)*
The facility-level mobility improvement scores are calculated using the following 15 steps.

Step 1. Choose the 12 month window for which we will select episodes. This is the four consecutive calendar quarters ending with the most recent calendar quarter for which both MDS data and CARE Tool data are available for use in the measure.

Step 2. Identify all MDS discharge assessments (in which we understand the CARE Tool items will be embedded) with a discharge date that fell within the 12 month window identified in Step 1.

Step 3. For each MDS tool discharge assessment identified in Step 2, identify the corresponding MDS admission assessment (in which we understand the CARE Tool items will be embedded). An MDS assessment is identified as an admission assessment if A0310F == "01" (entry record). Note that the admission date may lie before the 12 month window defined in Step 1. The period of time from the admission date (corresponding with the MDS admission assessment) through to the discharge date (corresponding with the MDS discharge assessment) is called an "episode". If no MDS admission assessment was found, discard the discharge assessment from all subsequent steps.

Step 4. Identify all MDS admission assessments that indicate the admission to the SNF was from the hospital, another SNF or IRF. An MDS admission assessment indicates that the SNF admission was from a hospital when MDS item "A1800 Entered From" coded as "03 Acute Care Hospital" or "02 Another nursing home or swing bed" or "05 inpatient rehabilitation facility" or "09 Long Term Care Hospital". The MDS item A1600 indicates the date of entry to the SNF.

Step 5. For any admission or discharge CARE Tool item (that enters the calculation of the mobility improvement scores) with letter code "S" (activity not attempted due to safety concerns), A. Task attempted but not completed, N. Not applicable and P. Patient Refused were recoded to "1" on a six point rating scale (indicating full functional dependence).

Step 6. Apply the mobility improvement measure's exclusions (see s.11), and exclude any episode that did not involve either physical or occupational therapy. The clinical measure exclusions are detailed in S.11 (Denominator exclusion details and codes). The exclusion of episodes not involving either occupational or physical therapy is as follows:
We identify the patient as having received occupational therapy if on the MDS discharge assessment:
The total number of minutes of occupational therapy in the last 7 days (O0400B1) is greater than zero; or
The most recent occupational therapy regimen (starting on the date recorded in O0400B5, and ending on the date recorded in O0400B6) intersects the episode (beginning with the CARE admission assessment's admission date and ending with the CARE discharge assessment's discharge date).
We identify the patient as having received physical therapy if on the MDS discharge assessment:
The total number of minutes of physical therapy in the last 7 days (O0400C1) is greater than zero; or
The most recent physical therapy regimen (starting on the date recorded in O0400C5, and ending on the date recorded in

O0400C6) intersects the episode (beginning with the CARE admission assessment's admission date and ending with the CARE discharge assessment's discharge date).
If the episode involves neither occupational nor physical therapy, as identified above, then exclude it from all subsequent steps in the calculation.

Step 7. Map the CARE Tool B5a (walking) and B5b (wheeling) items to obtain a harmonious 1-6 score for all assessments, and recode walking items C7b, C7c, C7d and C7e to 1=dependent if resident cannot walk. First, consolidate the four sub-items B5a1, B5a2, B5a3 and B5a4 corresponding to different distances the resident can walk (if the patient can walk); and the four sub-items B5b1, B5b2, B5b3 and B5b4 corresponding to different distances the resident can wheel (if the patient cannot walk). To do this, use the crosswalk presented in Figure A1 in the Appendix. Call the resulting two items B5a and B5b.
Second, consolidate the B5a and B5b items into a harmonious summary item called B5. To do this use the crosswalk presented in Figure A1 in the Appendix. This is the item used in the calculation of mobility outcome scores in the subsequent steps.
Finally, if the patient is unable to walk (i.e., no values for the B5a and C7 items), recode each item C7a, C7b, C7d and C7e to 1 = dependent.

Step 8. For each episode remaining after Step 6, using the CARE Tool items as transformed in Step 7, calculate a preliminary admission score and a discharge score as the sum of the values for the following CARE tool mobility items B1 (Lying to sitting on side of bed), B2 (Sit to stand), B3 (Chair/bed-to-chair transfer), B4 (Toilet transfer), B5 (Walking/wheeling), C3 (Roll left and right), C4 (Sit to lying), C7a (One step (curb)), C7b (Walking 50 feet with two turns), C7c (Walking 12 steps), C7d (Walking four steps), C7e (Walking 10 feet on uneven surfaces).
Each of those 12 CARE Tool items takes an integer value of 1, 2, 3, 4, 5 or 6, and so the preliminary admission score will be an integer between 12 and 72, and the preliminary discharge score will be an integer between 12 and 72.

Step 9. For each episode, linearly transform the preliminary admission score and preliminary discharge score so that it lies in the range 1-100 using the following equation:
["transformed mobility admission score"]=1.65×["preliminary mobility admission score"]-18.8
["transformed mobility discharge score"]=1.65×["preliminary mobility discharge score"]-18.8

Step 10. For each episode, calculate the episode-level change score by subtracting the transformed discharge score from the transformed admission score. Each score will lie between -99 and 99.

Step 11. Calculate the national average change score as the simple mean of all episode-level change scores calculated in Step 10.

Step 12. For each episode, calculate the predicted change score using the risk adjustment methodology detailed in S.15a. That is, having prepared the risk adjustment variables in the way described in S.15a, apply the equation: [predicted change score] = 33.61 1.56×[patient is 85 years or older] -9.11×[dialysis while a resident] -5.08×[entered from SNF] -2.81×[oxygen while a patient] 4.23×[unhealed pressure ulcers] -8.85×[mental status] -4.75×[resident mood] -9.30×[psychiatric conditions] -6.91×[feeding tube or IV feeding] -4.10×[suctioning or tracheotomy] -3.98×[infections of the foot].

Step 13. For each episode, calculate the risk adjusted change score using the actual change score calculated in Step 10, the national average change score calculated in Step 11, and the predicted change score calculated in Step 12. The risk adjusted change score is: [risk adjusted change score] = ([national average change score] - [predicted change score]) + [actual change score].

Step 14. Exclude any facility that has fewer than 30 episodes for which we could calculate a risk adjusted change score.

Step 15. For each facility remaining after Step 14, calculate its mobility improvement score as the simple mean of the risk adjusted change scores calculated in Step 13.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without

Not Applicable

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not Applicable

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Missing data resulted in the individual being excluded from the calculation. We note, however, that we applied three rules for recoding certain CARE Tool mobility items that would otherwise either be coded missing or using a letter code (indicating, for example, inability of a resident to perform a test). The first rule was to recode any mobility item with a letter code of S (not attempted due to safety concerns), A (task attempted but not completed), N (not applicable), and P (patient refused) to 1 (completely dependent). The second rule consolidated the eight items B5a1-B5a4 and B5b1-B5b4 to get a single summary ambulation score; these items identify whether a resident can walk or can only wheel, the maximum distance that they can walk or wheel, and their functional independence at doing so. The second rule occurred in three steps: first we consolidated items B5a1-4 into a single summary walking score B5a, second we consolidated items and B5b1-4 into a single summary wheeling score B5b, and third consolidated the consolidated walking and wheeling scores B5a and B5b into a single summary ambulation score B5. These were according to Figures A1 and A2 in the Appendix. Once these rules are applied, the frequency of CARE Tool assessments with missing data (and which are thus discarded) is substantially reduced. When the CARE Tool is implemented as part of the IMPACT Act of 2014; the mobility section will be incorporated into MDS 3.0 and we anticipated missing data to be very low as is currently the case for ADL items in MDS and what we found across our four companies.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Other

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Resident Assessment Instrument Minimum Data Set (MDS) version 3.0
Continuity Assessment and Record Evaluation (CARE) Tool; Mobility subscale

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)* Post Acute/Long Term Care Facility : Nursing Home/Skilled Nursing Facility If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)* Not Applicable

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
Mobility_Testing_Submission_Form_2612-635509010775453804.docx

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.**

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score) 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)* 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.**

The CARE Tool is currently available in a PDF document, which can be completed in written format. However, a number of software vendors for rehabilitation therapy providers have begun the process of adding the CARE item set to their online documentation systems. Two of the largest therapy software companies now provide the CARE Tool in electronic format. Currently over 48 organizations representing 1016 SNFs have begun to use the CARE tool. One software therapy company has also developed on online secure portal where providers can submit their data to a larger database and receive confidential, secure outcome performance reports.

Skilled nursing care centers encode and electronically transmit the MDS 3.0 data set, as required by the federal government.

The IMPACT act of 2014 recently passed by congressed and signed by the President; requires the incorporation of standardized assessments for mobility and self-care into the MDS by October 2018 (fiscal year 2019). This will make the data collection for this measure extremely feasible as it will be universally collected on all admissions and discharges to all SNFs in the country. The IMPACT act also requires the public reporting of functional outcome measures for SNFs.

United States. Cong. House of Representatives. IMPACT Act of 2014. 113th Cong., HR 4994. Washington: GPO, 2014.

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measurespecific URL.**

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. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

In developing this measure, 89 skilled nursing centers agreed to complete and collect the mobility subscale of the Continuity Assessment and Record Evaluation (CARE) tool. A basic training program to the CARE tool was established, requiring therapists to pass a post-test. 425 therapists were trained with a 95% pass rate on the post-test. A key challenge identified was the CARE tools method of rating the patient's usual level of performance, rather than the lowest level of performance which is commonly used in various proprietary tools (e.g.,MBI, FOM and ROM) as well as in the MDS 3.0 scale. This was addressed by focusing the training and design of the clinical vignettes used in the training and test to highlight the difference in rating between the usual level of performance versus the most dependent/independent level of performance. Overall, therapists did not report any significant issues in understanding or assessing patients using the CARE scale, nor did they report that the CARE scale was more burdensome than the proprietary scales currently in use. The two commonly reported data collection issues were the inconvenience of completing a written scale and the need to complete all items on the scale. The first issue, completing a written scale, is currently being addressed by software vendors (see 3b.1.) and per the IMPACT act of 2014 will be incorporated into the MDS in the near future. The second issue, completing all items on the scale, must become an industry standard. When the CARE core mobility items are incorporated into the MDS, per the IMPACT act of 2014, we anticipate that all items will be required as this is the current standard of the MDS 3.0 tool. Common practices among many therapy providers are to complete only items which are the focus of care. However, moving to a practice of completing all items on the CARE assessment form is integral to quality improvement and measurement efforts. This was supported by the therapy providers that participated in this study. Currently the therapy companies that do not require therapists to complete all items on their proprietary scales are also not able to generate an overall quality measure score.

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.

Planned	Current Use (for current use provide URL)
Public Reporting	
Quality Improvement with Benchmarking (external benchmarking to multiple organizations)	
Quality Improvement (Internal to the specific organization)	

One section on the original CARE scale that caused therapists confusion was the Mode of Mobility question in section C7 (see Appendix). This question asks if the patient primarily walks or uses a wheelchair, then has the therapist test a series of walking or wheelchair tasks dependent on their response. Confusion arose around how to respond if the patient both walks and uses a wheelchair or the therapy plan or care included goals for both walking and wheelchair locomotion. The form was changed to add a response option denoting "both" and allowing therapists to rate patients on all walking or wheeling tasks. This also maintains consistency with question B5 (see Appendix).

In addition, one letter code from the original CARE scale developed by CMS was dropped. This code identified "M. Not attempted due to medical condition." It was determined that this letter code was unnecessary because the codes "1. Dependent"and "S. Not attempted due to safety concerns" would replace the code M in any situation.

We used the core mobility items and scoring on the functional status section of the CARE tool and when implemented in the MDS this measure will be able to be calculated from these items.

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)*.

No fees or licensing requirements for use of CARE Tool or the use of the submitted quality measure. The CARE Tool is currently available in the public domain. http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-CareQuality-Initiatives/Downloads/CARE-Institutional-Admission-Assessment-Tool.pdf.

Additionally, no fees are required for the utilization of the MDS 3.0, it is publicly available at http://www.resdac.org/cmsdata/files/mds-3.0.

4a.1. For each CURRENT use, checked above, provide:

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

Not Applicable

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

The CARE Tool currently is not required by CMS or others and is not part of the MDS. As a result, data on all the SNFs in the country are not available for public reporting. However, the IMPACT Act of 2014 passed by Congress and signed into law by the President in October 2014 requires the adoption of a standardized functional assessment tool in all post-acute care settings and public reporting of self care and mobility quality measures. The CARE tool (including self care and mobility assessment scales), developed by CMS, is the only assessment tool validated across all the PAC settings (Skilled Nursing Facilities (SNFs), Home Health Agencies (HHAs), Inpatient Rehab Facilities (IRFs) and LTACHs Long Term Acute Care Hospitals).

4a.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.*)

American Health Care Association (AHCA) and National Association of the Support of Long Term Care (NASL) who jointly sponsored the development of this measure have actively supported the IMPACT Act. They also participated in the development of the CARE Tool and support its use and adoption by their respective members as a means for standardized assessment of our patients. AHCA represents nearly 10,000 of the approximately 15,000 SNFs in the country. NASL represents therapy companies as well as software vendors supporting therapy services. The AHCA Board of Governors and NASL have both endorsed the mobility quality measure. AHCA and NASL have met with CMS asking that the CARE Tool mobility assessment be adopted in SNFs so that outcome measures can be developed for public reporting and eventually incorporated into new payment models.

MedPAC has also called on CMS to adopt the CARE Tool to measure outcome measures (MedPAC annual report to congress 2014; page 174 Chapter 7 Post Acute Care Provider: Steps toward broad payment reforms. http://www.medpac.gov/documents/reports/chapter-7-post-acute-care-providers-steps-toward-broad-payment-reforms-(march2014-report).pdf?sfvrsn=2).

As a result of these efforts, NASL members, including one of the largest software companies that supports therapy services in SNFs, has already incorporated the CARE Tool into their software and is working with AHCA to incorporate this quality measure into their software. This has resulted in 48 organizations representing 1,016 SNFs adopting the use of the CARE tool. To date, they have completed CARE Tool assessments on over 48,971 of patients. AHCA is also reaching out to other software companies that support therapy in SNFs to adopt the CARE Tool mobility assessment and provide the necessary information to calculate the quality measures. In addition, two large NF chains will adopt the CARE tool mobility assessment starting in early 2015.

AHCA also plans to incorporate this quality measure into its web-based reporting/benchmarking tool – Long Term Care Trend Tracker. This tool allows SNFs to calculate and trend a wide array of quality metrics over time and benchmark to peers – see http://www.ahcancal.org/research_data/trendtracker/Pages/default.aspx. We are currently working on a portal to allow SNFs of therapy companies to upload their mobility quality measure scores and benchmark to peers. The AHCA Board of Governors has approved funds to build this portal in 2015. This information will help individuals SNFs with their internal quality improvement efforts as they look at their trends over time in improvement in mobility. In addition, one of the large rehabilitation software vendors is developing similar ability for providers not using their software to upload their mobility change score for quality improvement tracking and bench-marking against others. We are in the process of providing the specifications and algorithms necessary for these software vendors and NF companies to calculate the mobility quality measure.

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

Not Applicable

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Not Applicable

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
Not Applicable

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)**

2774: Functional Change_ Change in Mobility Score for Skilled Nursing Facilities

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

5a. Harmonization

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 completely harmonized?

No

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

While there are small differences between the functional characteristics used to measure improvement, AHCA agrees with NQF that the AHCA and UDSMR measures capture essentially the same quality domain. The key substantive differences between the two measures are:

1. The measures are based on different data collection tools:
    a. The AHCA measure is based on the CARE tool, and the UDSMR measure is based on the FIM – two distinct assessment instruments that cannot be harmonized. While there are several CARE and FIM items that collect similar information, the wording of the questions when similar and the rating scales used are different between the CARE and FIM. CARE items are rated on a 1-6 scale, and FIM are on a 1-7 scale. That is, while there are similarities between the two instruments, the differences are sufficient that the two measures could not be harmonized.
2. Feasibility and usability:
    a. AHCA's measure in simple to implement on national data available from CMS for all SNFs in the country, once Section GG is added to the MDS in October 2016 (and one minor change to the numerator definition is made to the measure to align with the final set of MDS Section GG items), and will be calculated and published freely by AHCA for all SNFs in the nation. This means the AHCA measure can be used for benchmarking against other SNFs, which is critical for SNF Quality Assurance/Performance Improvement work. In the measure's current form, a number of vendors have implemented specifications in our NQF application, and are providing AHCA mobility and self care improvement measure rates, with benchmarking for their clients using CARE assessments. We do not have numbers on how many SNFs are currently using the measure, however are aware of 128 organizations representing 2,511 SNFs that have undergone CARE tool training, and therefore are likely to be doing CARE tool assessments within their organizations. Starting October 2016, all SNFs in the country will be completing the MDS Section GG assessments needed to calculate the AHCA measure, and those data will be available through the standard Research Identifiable File CMS assessment data dissemination route.
    b. While UDSMR has recently made the FIM instrument available freely for SNFs to use, SNFs, even if they were able to implement the measure (which would be difficult, for reasons described next), they would only have access to their own performance – without any ability to compare against SNFs outside their organization.
    c. UDSMR's numerator is built on a special case of a very complex statistical method called multilevel Rasch analysis, and UDSMR has not provided the full specifications of this special case needed to implement their methodology, making implementation of their numerator calculation impossible using the NQF application. While the advanced and very specialist statistical programming environment "R" has an implementation of this methodology under an "open source" license (i.e., could be used for other purposes), adapting the algorithm to implement in a SNF's own data systems would require a sophisticated software engineering firm with advanced statisticians available to do the project. This is prohibitively expensive for a SNF to do, and means the measure is not accessible without a paid subscription to UDSMR's services.

d. UDSMR has employed IRF Case Mix Groups (IRF-CMGs) to risk stratify the measure (though they appear to have renamed them as SNF Case Mix Groups in the NQF application). In order to implement CMS's IRF-CMG grouper algorithm, available on the CMS IRF-PPS website, the SNF would either need to implement the Visual Basic/C++ software libraries on their own system, or implement the algorithm CMS lays out in their software documentation; both of these approaches require either very specialist in-house software development resources, or hiring a software engineering firm to do it for the SNF. Therefore, the risk stratification approach further reduces the feasibility and usability of the measure, unless, again, the SNF pays UDSMR for their subscription services.

3. Numerator definition:
    a. Mobility
        i. The AHCA and UDSMR mobility and improvement measures capture essentially the same concept, though the AHCA measure uses a much expanded set of items than the UDSMR measure does. Also, as described above the wording of the items and rating scales differ.
        ii. The AHCA mobility improvement measure includes the following CARE tool items in the numerator:
            1. Lying to Sitting on Side of Bed
            2. Sit to Stand
            3. Chair/Bed to Chair Transfer
            4. Toilet Transfer
            5. Walking or Wheelchair Mobility
            6. Roll left / right
            7. Sit to Lying
            8. Picking up object
            9. One Step Curb
            10. Walk 50 ft. with Two Turns
            11. Walk 12 Steps
            12. Walk Four Steps
            13. Walking 10 ft. on
        iii. The UDSMR mobility improvement measure includes the following FIM tool items in the numerator:
            1. Transfer Bed/Chair/Wheelchair
            2. Transfer Toilet
            3. Locomotion and Stairs
    b. The AHCA measure sums functional independence scores (ranging 1=fully dependent to 6 = fully independent) on the included items at admission and discharge, and calculates the difference. The FIM uses a different approach relying on Rasch model. During development, we explored using a multilevel Rasch model, similar to UDSMR. However, we found the facility-level performance scores for the measure using Rasch correlated almost exactly with those from simply adding the independence scores together, with a correlation coefficient of 0.99. Therefore, because employing a Rash model approach significantly damaged the feasibility and usability of the measure, as well as the user's ability to interpret the meaning of the performance scores, we adopted the simpler approach.
    c. UDSMR, as stated, has used this very complex approach without any data comparing to a simpler approach that is easier and more feasible to implement for providers. Additionally, UDSMR has not provided full specifications in the NQF application of how to calculate the measure using this approach limiting our ability to fully comment on the harmonization. .

4. Risk model development and specifications
    a. The AHCA measure risk model was developed by a workgroup of clinicians and statisticians, choosing clinically appropriate risk factors for the measure in the SNF setting, who then iteratively analyzed

patterns and reviewed the clinical meaning of the to develop and test the final statistical model. This was a rigorous process, and we laid the process and testing of the model out in our answers to NQF questions 2b4.4b, 2b4.5, 2b4.9 and 2b4.10 to evaluate the adequacy of the risk factors and the overall adequacy of the risk stratification approach.

b. UDSMR used IRF case mix groups to risk stratify the measure, but renamed them SNF-CMGs. . UDSMR application does not provide information on how they developed their approach, their testing of it, and the effect of the risk stratification approach on the measure. This makes it difficult for us to comment on harmonization of our approach with theirs. They claim that "Patients within the same CMG are expected to have similar resource utilization needs and similar outcomes" but do not provide any data or references supporting the use of IRF CMGs applied to SNF population. We are unfamiliar with SNF_CMGs and therefore cannot comment on harmonization with our risk adjustment approach.
c. We are not aware of any common practice of applying IRF case mix groups to risk adjust SNF quality measures by CMS or skilled nursing centers in general. We are also not aware of any testing of the appropriateness of applying IRF case mix groups to SNF patients or to the UDSMR SNF measure. SNF benefit eligibility is simply that the patient had a 3-day hospital stay. IRF benefit eligibility is that the patient can handle 3 hours of intensive rehabilitative services for 5 consecutive days, and the IRF must have 60% of its patients in 13 specified clinical groups. These differences in the IRF and SNF benefit rules mean the IRF and SNF patient populations and functional needs are fundamentally different, which raises the question about whether this approach is appropriate for the SNF population or for their measure, particularly given UDSMR declined to complete this section of the application.
d. Additionally, the IRF-CMG grouper algorithm that UDSMR has employed in their measure relies on a mixture of assessment information (which can be taken from the FIM), but also relies heavily on diagnosis coding. It is well known that SNF diagnosis coding on SNF-PPS claims and on the MDS is extremely sparse and unreliable, and reporting practices vary by SNF and by organization. SNF diagnosis codes, except for HIV, are not currently used for payment or any other purpose. Some SNFs, for example, copy the hospital diagnosis codes presented on the hospital's transfer documents onto the MDS assessments and SNF claims; others will record their own, but because SNFs typically do not have physicians on staff, the accuracy of the diagnosis coding is questionable. The unreliability of SNF diagnosis coding, therefore, translates into unreliability of the measure's risk stratification approach. Having not tested their risk stratification approach, however, it is not possible to know the extent of this problem.

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.)
NQF should not endorse both sets of measures, as they are measuring the same thing. We believe that NQF should endorse AHCA's measures because they are more feasible and usable on the CMS's national assessments infrastructure, and they will be published freely for all SNFs in the country, creating a complete and live national Quality Assurance/Performance Improvement (QAPI) discussion about functional outcomes. UDSMR's measures cannot accomplish the same thing, for all the reasons mentioned above.

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:** Mobility_Appendix_2612.docx

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** American Health Care Association
Co.2 **Point of Contact:** Urvi, Patel, ushah@ahca.org, 202-842-4444-
Co.3 **Measure Developer if different from Measure Steward:** The Moran Company
Co.4 **Point of Contact:** Rachel, Feldman, rlfeldman@themorancompany.com, 703-841-8405-

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.
Mary Ousley, Ousley and Associates- Expert Panel Co-chair Howie
Groff, Tealwood Care Centers- Expert Panel Co-chair

Members of the Expert Panel Include:
Cynthia Morton, NASL
Martha Schram, Aegis Therapies
Bill Goulding, Aegis Therapies
Mary Van De Kamp, Kindred/Rehab Care
Matt Sivret, Kindred/Rehab Care
Phil Fogg, Marquis Companies
Tracy Fritts, Consonus Healthcare
Garry Pezzano, Genesis Rehab Services
Felicia Chew, Genesis Rehab Services
Mike Morris, Genesis Rehab Services
John Barber, White Oak Manor
Rick Black, HCR Manor Care
Leigh Ann Frick, Heritage Health Care
Doug Burr, Health Care Navigator
Katarika Lewis, Halcyon Rehabilitation
Victoria Cruce Hollar, Halcyon Rehabilitation
Chris Castel, Accelerated Care Plus
Ellen Strunk, Rehab, Resources and Consulting

The expert panel met regularly and provided guidance on risk adjustment, exclusions, measure specifications and use of the

measure.

Observers:
Mary Pratt, CMS
Tara McMullen, CMS

Members of measure steward, American Healthcare Association:
David Gifford
Courtney Bishnoi
Urvi Patel
James Muller

Members of the project contractor, The Moran Company:
Iara Woody
Chris Young
Rachel Feldman
Peter Gruhn

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2014

Ad.3 **Month and Year of most recent revision:** 06, 2014

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

Ad.5 **When is the next scheduled review/update for this measure?** 12, 2015

Ad.6 **Copyright statement:** None

Ad.7 **Disclaimers:** None

Ad.8 **Additional Information/Comments:** None

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 subcriterion 1b).

Brief Measure Information

NQF #: 2613
De.2. **Measure Title:** CARE: Improvement in Self Care
Co.1.1. **Measure Steward:** American Health Care Association
De.3. **Brief Description of Measure:** The measure calculates a skilled nursing facility's (SNFs) average change in self care for patients admitted from a hospital who are receiving therapy. The measure calculates the average change in self care score between admission and discharge for all residents admitted to a SNF from a hospital or another post-acute care setting for therapy (i.e., PT or OT) regardless of payor status. This is a risk adjusted outcome measure, based on the self care subscale of the Continuity Assessment and Record Evaluation (CARE) Tool and information from the admission MDS 3.0 assessment. The measure is calculated on a rolling 12 month, average updated quarterly.
1b.1. **Developer Rationale:** Therapies in SNFs serve a critical role in helping individuals achieve and maintain maximum physical and functional well-being following hospitalization. One of the principle reasons individuals are discharged from a hospital to a SNF is to improve their functional abilities so they can return to their prior living situation. The Medicare SNF Part A benefit is designed principally for the purpose of either providing therapy to individuals who were hospitalized or to complete their nursing care in a less intensive setting. Over 85% of all Medicare Beneficiaries admitted to SNFs receive therapy services. The level of improvement in function (self care and mobility) is a strong predictor of a person's ability to reside in the community independently (Granger et al., 1979).

Currently, there is no quality measure to assess how well SNFs improve a person's functional ability, particularly for self-care. This is in part due to the lack of a standard assessment tool used across all post-acute care (PAC) providers. Many of the current therapy companies have developed their own or use existing therapy related assessment tools; none of these assessment tools are standardized. This lack of consistency prevents industry wide comparison. Consumers need a standard measure in order to make educated choices in selecting high quality care providers to maximize outcomes. Educated consumers who make decisions regarding care based on validated measures will also promote accountability and quality among PAC providers. PAC providers, in turn, need standardized assessments to evaluate the quality of services provided, and identify areas where improvement is needed. Policy makers and payors need such a measure to evaluate the value of therapy services they are paying for or to incorporate into new payment models. Research by standardized outcome assessments, questionnaires or tools are a vital part of evidence-based practice. Selecting the most appropriate outcomes assessment, questionnaire or tool enhances clinical practice by identifying and quantifying areas related to self care; formulating the evaluation, diagnosis, and prognosis; designing the plan of care; and helping to evaluate the success of physical and occupational therapy interventions. Policy makers and payors also need standardized measures of improved function to evaluate whether the purpose for paying for post-acute rehab is met as well as to incorporate into value based payment models.

Potter, K., Fulk, G.D., Salem, Y., Sullivan, J. "Outcome Measures in Neurological Physical Therapy Practice: Part I. Making Sound Decisions". Journal of Neurologic Physical Therapy. (2011): 57-64.

Grander, C.V., Dewis, L.S., Peters, N.C., Sherwood, C.C. & Barrett, J.E. (1979). "Stroke rehabilitation: Analysis of repeated Barthel index measures". Arch Phys Med Rehabil. 60.1 (1979): 14-7.

S.4. **Numerator Statement:** This outcome measure assesses the change in self-care. The numerator is the risk adjusted sum of the change in the CARE Tool self care subscale items between admission and discharge for each individual admitted from a hospital or another post-acute care setting regardless of payor status and are receiving therapy (PT or OT) for any reason in a skilled nursing center.
S.7. **Denominator Statement:** The denominator includes all residents admitted to a SNF from a hospital or another post-acute care

setting who receive either PT or OT therapy for any reason during their stay regardless of payor status, have a completed self care subscale of the CARE Tool at admission and discharge and do not meet any of the exclusion criteria and do not have missing data. The self care items used from the CARE tool are listed below and rated on a 1-6 scale (see Appendix for CARE Tool).

The items included in the CARE Tool self care subscale include:

- A1. Eating
- A3. Oral Hygiene
- A4. Toilet Hygiene
- A5. Upper Body Dressing
- A6. Lower Body Dressing
- C1. Wash Upper Body
- C2. Shower / Bathe
- C6. Putting on / taking off footwear

S.10. **Denominator Exclusions:** Individual patients are excluded for two broad reasons:

1. if they have conditions where improvement in self-care is very unlikely,
OR
2. have missing data necessary to calculate the measure

Additionally, facilities with denominator size of fewer than 30 patients during a 12 month period are excluded from reporting of their data.

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Electronic Clinical Data, Other
S.26. **Level of Analysis:** Facility

IF Endorsement Maintenance – Original Endorsement Date: Jul 23, 2015 **Most Recent Endorsement Date:** Jul 23, 2015

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

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

De.4. **IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** Not Applicable

1. Evidence, 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
Self_Care_Evidence_Submission_Form_2613.docx

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., the benefits or improvements in quality envisioned by use of this measure)*
Therapies in SNFs serve a critical role in helping individuals achieve and maintain maximum physical and functional well-being following hospitalization. One of the principle reasons individuals are discharged from a hospital to a SNF is to improve their functional abilities so they can return to their prior living situation. The Medicare SNF Part A benefit is designed principally for the purpose of either providing therapy to individuals who were hospitalized or to complete their nursing care in a less intensive setting. Over 85% of all Medicare Beneficiaries admitted to SNFs receive therapy services. The level of improvement in function (self care and

mobility) is a strong predictor of a person's ability to reside in the community independently (Granger et al., 1979).

Currently, there is no quality measure to assess how well SNFs improve a person's functional ability, particularly for self-care. This is in part due to the lack of a standard assessment tool used across all post-acute care (PAC) providers. Many of the current therapy companies have developed their own or use existing therapy related assessment tools; none of these assessment tools are standardized. This lack of consistency prevents industry wide comparison. Consumers need a standard measure in order to make educated choices in selecting high quality care providers to maximize outcomes. Educated consumers who make decisions regarding care based on validated measures will also promote accountability and quality among PAC providers. PAC providers, in turn, need standardized assessments to evaluate the quality of services provided, and identify areas where improvement is needed. Policy makers and payors need such a measure to evaluate the value of therapy services they are paying for or to incorporate into new payment models. Research by standardized outcome assessments, questionnaires or tools are a vital part of evidence-based practice. Selecting the most appropriate outcomes assessment, questionnaire or tool enhances clinical practice by identifying and quantifying areas related to self care; formulating the evaluation, diagnosis, and prognosis; designing the plan of care; and helping to evaluate the success of physical and occupational therapy interventions. Policy makers and payors also need standardized measures of improved function to evaluate whether the purpose for paying for post-acute rehab is met as well as to incorporate into value based payment models.

Potter, K., Fulk, G.D., Salem, Y., Sullivan, J. "Outcome Measures in Neurological Physical Therapy Practice: Part I. Making Sound Decisions". Journal of Neurologic Physical Therapy. (2011): 57-64.

Grander, C.V., Dewis, L.S., Peters, N.C., Sherwood, C.C. & Barrett, J.E. (1979). "Stroke rehabilitation: Analysis of repeated Barthel index measures". Arch Phys Med Rehabil. 60.1 (1979): 14-7.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Using data from Jan 2012 to Dec 2012 from three large therapy companies representing 607 SNFs and 142,125 matched (admission-discharge) patient assessments. The frequency distribution of the SNFs' average self-care change scores are shown in Figure A3 in the Appendix. The mean = 21.2, std dev=6.28, min=8.1, max=54.7, 1st quartile = 16.7, 3rd quartile = 25.7, 1st decile= 13.4, 2nd decile = 15.8, 3rd decile = 17.8, 4th decile = 18.9, 5th decile = 20.4, 6th decile = 22.4, 7th decile = 24.4, 8th decile = 26.8, 9th decile = 29.2.

Additionally, Table A1 in the Appendix shows the distribution of facilities and risk adjusted self care change scores by facility bed count, ownership type, and by urban/rural location.

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 endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Specific data on disparities was not included in this measure as specified by current NQF requirements; however, due to the potential for widespread use of this measure we have included distribution of SNF admissions by gender, age and ethnicity in SNFs from the second quarter of 2014, based on the MDS 3.0 (see Table A2-A4 Appendix). Nationally, 66% of all admissions to SNFs are female. Approximately three-quarters are between the ages of 65 and 95 years. Based on the MDS, the majority are considered white (76%) with 14% African American, 5% Hispanic, 2% Asian and less than 1% each native Hawaiian or other pacific islander and American Indian or Alaska native. A state by state break down is provided in the appendix. This makes stratification at a facility level extremely difficult because sample sizes for ethnic groups within a facility are small and frequently below the minimum denominator size of 30 assuming the average nursing home bed size of approximately 110.

We are not able to present information on insurance status based on the MDS, as it is not reliable due to the accuracy of the information submitted by providers, the ambiguity of payer status at admission, the number of patients with multiple payers and

patient's whose payor status changes during the course of care in the SNF.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.

A PubMed search on disparities related to therapy outcomes in skilled nursing facility residents did not produce any meaningful results. There is some evidence that suggest differences in access and utilization of post-acute rehabilitation care by ethnicity but none on differences in the quality of care delivered or outcomes within a single provider. A patient can receive post-acute rehabilitation care in inpatient rehabilitation facilities, skilled nursing homes, or through home health care; the former provides more hours of care than the latter when viewed on a continuum. Freburger et al. (2012) found that minorities and those with lower socioeconomic statuses receive lower volumes of rehabilitation care. These individuals are more likely to be discharged home and receive care in a SNF verses an inpatient rehabilitation facility. This finding is also supported by earlier research showing that racial minorities, women, older individuals, and those with lower incomes are more likely to receive care in SNFs or home health (Freburger et al., 2011). However, no studies looked at difference in volume of therapy services by ethnicity within SNFs, just across different types of PAC providers.

There is evidence that racial disparities exist in care provided in different nursing homes (though not specific to therapy services). An article by Smith et al. (2007) suggests that racial segregation in nursing homes mirrors that which occurs in metropolitan areas. Black nursing home residents are 1.41 times more likely to be in a facility cited with a deficiency causing actual harm or immediate jeopardy to residents. Forty percent of African American patients are in lower tier facilities, those with higher number of Medicaid patients and limited resources, compared to nine percent of white residents. The lower tier facilities are shown to have fewer nurses, lower occupancy rates, and more health-related deficiencies (Mor et al., 2004). However, the outcomes of these individuals did not differ from other residents in the same facility. Thus, suggesting differences are related to the facility's location and overall practices not differences related to ethnicity or social economic status of the residents. A 2013 study also found that the differences in quality between SNFs with higher proportion of African American residents was mediated by the overall financial health of the facility and overall quality in the facility, rather than the racial mix (Chisholm et al., 2103). In summary, the literature suggests that ethnic and social economic status differences are related to inter-facility differences not to intra-facility differences in care. Therefore, it is unclear based on the literature if social economic status should be risk adjusted but no studies have looked specifically at racial differences in therapy services or outcomes in the SNF setting.

Chisholm, L., Weech-Maldonado, R., Laberge, A., Lin, F. & Hyer, K. "Nursing home quality and financial performance: Does the racial composition of residents matter?" Health Services Research. (2013): 2060- 2080.

Fennell, M. L., Miller, S. C., & Mor, V. "Facility effects on racial differences in nursing home quality of care". American Journal of Medical Quality. 15.4 (2000): 174-181.

Freburger, J.K., Holmes, G.M., & Ku, L.J. "Postacute rehabilitation care for hip fracture: Who gets the most care?". J Am Geriatr Soc. 60.10 (2012):1929-1935.

Freburger, J.K., Holmes, G.M., Ku, L.J., Cutchin, M.P., Heatwole-Shank, K. & Edwards, L.J. "Disparities in postacute rehabilitation care for stroke: An analysis of the state inpatient databases". Arch Phys Med Rehabil. 92.8 (2011): 1220-1229.

Grabowski, D.C. "The admission of blacks to high-deficiency nursing homes". Medical Care. 42.5 (2004): 456-464.

Harada, N.D., Chun, A., Chiu, V. & Pakalniskis, A. "Patterns of rehabilitation utilization after hip fracture in acute hospitals and skilled nursing facilities". Medical Care. 38.11 (2000): 1119-1130.

Holmes, G.M., Freburger, J.K. & Ku, L.J. "Decomposing racial and ethnic disparities in the use of postacute rehabilitation care". Health Serv Res. 47.3 (2012): 1158-1178.

McCallum, C.A. "Access to physical therapy services among medically underserved adults: A mixed-method study". Phys Ther. 90.5 (2010):735-747.

Mor, V., Zinn, J., Angelelli, J., Teno, J.M., & Miller, S.C. "Driven to tiers: Socioeconomic and racial disparities in the quality of nursing home care". The Milbank Quarterly. 82.2 (2004): 227-156.

Smith, D.B., Fang, Z., Fennell, M.L, Zinn, J.S. & Mor, V. "Separate and unequal: Racial segregation and disparities in quality across U.S.

nursing homes". Health Affairs. 26.5 (2007): 1448-1458.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**

Affects large numbers, Patient/societal consequences of poor quality

1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

Nearly 2.5 million Medicare beneficiaries in Fee for Service receive physical and/or occupational therapy each year in the skilled nursing setting. According to the 2012 AHCA Annual Report, the skilled nursing setting received over 2.3 million FFS Medicare admissions and over half a million non-Medicare admissions in 2012. Of those Medicare admissions, 86.4% received occupational therapy and 89.6% received physical therapy. Of the non-Medicare admissions, 54.4% received occupational therapy and 58.5% received physical therapy.

Medicare FFS spending on SNF services was $28.7 billion in 2012. On a per user basis, spending per nursing home resident averaged $31,735 in 2010 (MedPac, 2014). Discharges for rehabilitation assume that patients will improve in functionality, however there is no consistent measurement showing whether or not patients improve. It is vital to be able to show the benefits provided to patients through these services.

On average, about 57% of all individuals admitted to a skilled nursing facility for rehabilitation services are discharged to home. Rehabilitation services, including the intensity and appropriateness of services, have strong relationships with patient functional gains, community discharge and survival. These services are vital to the quality of life of the patients receiving them. The level of improvement in function (Self-care and mobility) is a strong predictor of a person's ability to reside in the community independently (Grander et al., 1979).

The CARE assessment tool was developed by CMS based on a modification of the existing Barthel Index (Gage et al., 2012). The Barthel Index was developed in the 1950's, has been modified and adapted by numerous researchers. It is one of the most widely used measures of functional status and as a result, has been repeatedly tested for reliability, validity and sensitivity. The Barthel Index has been shown be highly reliable with lower administrative burden than other widely used tools. The Barthel Index has shown high inter-rater reliability when performed by a therapist versus a nurse assessor. Not surprising, the CARE tool was found to have equally high reliability and validity when tested by CMS in SNFs and other post-acute care settings (Gage et al., 2012).

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

American Health Care Association. "2013 Quality Report". 2013. http://www.ahcancal.org/quality_improvement/Documents/AHCA%20Quality%20Report%20FINAL.pdf.

Cohen, M.E., Marino, R.J. "The Tools of Disability Outcomes Research Functional Status Measures". Arch Phys Med Rehabil. 81 (2000): S21-S29.

Gage, G., Morley, M., Smith, L., Ingber, M.J., Deutsch, A., Kline, T., … Manning, J. "Post-acute care payment reform demonstration: Final report". RTI International. 2 (2012).

Grander, C.V., Dewis, L.S., Peters, N.C., Sherwood, C.C. & Barrett, J.E. "Stroke rehabilitation: Analysis of repeated Barthel index measures". Arch Phys Med Rehabil. 60.1 (1979): 14-7.

Jette, D.U., Warren, R.L., Wirtalla, C. "The Relation Between Therapy Intensity and Outcomes of Rehabilitation in Skilled Nursing Facilities". Arch Phys Med Rehabil. 86 (2005): 373-379.

MedPac. "Report to Congress: Payment policy". 2014.
http://www.medpac.gov/documents/reports/mar14_entirereport.pdf?sfvrsn=0.

Murray, P.K., Singer, M., Dawson, N.V., Thomas, C.L., & Cebul, R.D. "Outcomes of Rehabilitation Services for Nursing Home Residents". Arch Phys Med Rehabil. 84 (2003): 1129-1136.

Richards, S.H., Peters, T.J., Coast, J., Gunnel, D.J., Darlow, M., Pounsford, J. "Inter-Rater Reliability of the Barthel ADL Index: How Does a Researcher Compare to a Nurse?". Clinical Rehabilitation. 2000 (1999): 72-28.

Sangha, H., Lipson, D., Foley, N., Salter, K., Bhogal, S., Pohani, G., Teasell, R.W. "A comparison of the Barthel Index and the Functional Independence Measure as outcome measures in stroke rehabilitation: patterns of disability scale usage in clinical trials". International Journal of Rehabilitation Research. 28.2 (2005): 135-139.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
Not Applicable

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 subcriteria 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. **Cross Cutting Areas** *(check all the areas that apply):*
Functional Status, Health and Functional Status : Functional Status

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.)*
Not Applicable

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.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Not Applicable

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.

This outcome measure assesses the change in self-care. The numerator is the risk adjusted sum of the change in the CARE Tool self care subscale items between admission and discharge for each individual admitted from a hospital or another post-acute care setting regardless of payor status and are receiving therapy (PT or OT) for any reason in a skilled nursing center.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Rolling 12 month average, updated quarterly.

S.6. **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, 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.
The numerator includes all residents admitted from a hospital or another post-acute care setting that receive any PT or OT therapy for any reason in a SNF that have a completed CARE Tool self care subscale assessment at admission and discharge (see denominator definition below). The self care items used from the CARE tool are listed below and rated on a 1-6 scale (see Appendix for CARE Tool).

The items included in the CARE Tool self care subscale include:
- A1. Eating
- A3. Oral Hygiene
- A4. Toilet Hygiene
- A5. Upper Body Dressing
- A6. Lower Body Dressing
- C1. Wash Upper Body
- C2. Shower / Bathe
- C6. Putting on / taking off footwear

The numerator is facility's average risk adjusted change score on the self care subscale of the CARE tool. The risk adjusted average change score is calculated in several steps:

Step 1: Each individual's admission and discharge self care subscale score is calculated. Items rated as S. Not attempted due to safety concerns, A. Task attempted but not completed, N. Not applicable and P. Patient Refused were recoded to one on a six point rating scale (e.g. dependent). For each individual, the ratings for all the self care items on the CARE tool at admission are summed and transformed to a 0-100 scale. The same is done for the discharge assessment.

Step 2: Each individual's unadjusted change score is calculated by taking the admission score minus the discharge score.

Step 3: The individual's unadjusted change score is risk adjusted (see S.14)

Step 4: The facility's risk adjusted change score is the sum of all the individual's risk adjusted change scores divided by the denominator.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The denominator includes all residents admitted to a SNF from a hospital or another post-acute care setting who receive either PT or OT therapy for any reason during their stay regardless of payor status, have a completed self care subscale of the CARE Tool at admission and discharge and do not meet any of the exclusion criteria and do not have missing data. The self care items used from the CARE tool are listed below and rated on a 1-6 scale (see Appendix for CARE Tool).

The items included in the CARE Tool self care subscale include:
- A1. Eating
- A3. Oral Hygiene
- A4. Toilet Hygiene
- A5. Upper Body Dressing
- A6. Lower Body Dressing

- C1. Wash Upper Body
- C2. Shower / Bathe
- C6. Putting on / taking off footwear

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk : Dual eligible beneficiaries, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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 denominator includes all residents admitted to a SNF who are receiving any PT or OT therapy for any reason. The denominator is based on admission from any hospital or post-acute care setting and is determined using information from MDS item "A1800 Entered From" coded as "03 Acute Care Hospital" or "02 Another nursing home or swing bed" or "05 inpatient rehabilitation facility" or "09 Long Term Care Hospital (LTCH)", regardless of payor status. They must receive either PT or OT therapy during their stay. A resident's stay is defined as an episode of care from admissions to discharge from the facility or discharge from therapy services (defined as completing a discharge CARE Tool assessment).

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Individual patients are excluded for two broad reasons:

1. if they have conditions where improvement in self-care is very unlikely,
OR
2. have missing data necessary to calculate the measure

Additionally, facilities with denominator size of fewer than 30 patients during a 12 month period are excluded from reporting of their data.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
Individuals with conditions where improvement in self care (as determined by a panel of expert therapists) is very unlikely were excluded based on information from the admission MDS 3.0 assessment. Individuals with one of the following MDS 3.0 items marked as yes were excluded:

- Ventilator (O0100F1 =1 or O0100F2 =1)
- Coma (B0100 =1)
- Quadriplegic (I5100=1)
- Hospice (O0100K1 = 1)

In addition, we also excluded individuals whose age is less than 18 years.
Overall, these exclusions resulted in 1.1% of all admissions being excluded.

Missing data also resulted in individuals being excluded, details are as follows:

- Missing a discharge CARE Tool assessment (this resulted when individuals died or were hospitalized during their SNF stay) resulted in patients being excluded since one could not calculate a change from admission. Nationally approximately 21.6% of admissions to a SNF will be hospitalized during their therapy stay and 4.5% will die (based on analysis of SNF part A claims from 2009-2011).
- Missing data on individual items on either the admission or discharge CARE Tool assessment resulted in the individual being excluded from calculation. For self care items, this occurred 4.4% of the time. We did not impute any missing data for self care items.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
Not Applicable

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Each individual's change score was risk adjusted based on the following formula:

Risk Adjusted Score for individual = (National Average Change Score – Predicted Change Score) + Actual Change Score.

The National Average Change Score was calculated as a population average change score for all patients in all SNFs who had a CARE Tool self care subscale assessment completed at admission and discharge. The change score is the difference in the aggregate of each individuals scale score from admission to discharge transformed to 0 to 100 scale.

The Predicted Change Score is calculated based on logistic regression using the process outlined in 2b4.

The Actual Change Score is the difference between the individual person's admission self care score transformed to 0 to 100 scale and their discharge self care score transformed to a 0 to 100 scale.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Provided in response box S.15a

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
The predicted change score used in S.14 is the sum of risk adjustment variables times each risk adjustment variable's respective coefficient. The risk adjustment variables and their coefficients are:

- Intercept (25.98),
- Patient is 85 years or older (-0.28),
- Dialysis while a patient (-4.43),
- Entered from a SNF (-3.83),
- Oxygen while a patient (-2.37),
- Catheterization/ostomy (-1.06),
- Unhealed pressure ulcers (-2.87),
- Mental status (-7.12),
- Resident mood (-3.33),
- Psychiatric conditions (-8.11),
- Feeding tube or IV feeding (-4.05),
- Suctioning or tracheotomy (-5.43), and
- Infections of the foot (-2.76).

Note that in the risk adjustment model all variables are non-missing even if the MDS items on which they are based are missing. This is consistent with MDS coding instructions. For example, if the MDS item O0100J2 (dialysis while a patient) was missing, then the condition (O0100J2 = "1") will not have been met, and the risk adjustment variable will simply be zero. Thus, any missing data is treated as if the person does not have that risk variable.
The full derivation of the risk adjustment variables from MDS items is specified in Section 2b4.4.

S.16. **Type of score:**
Continuous variable, e.g. average
If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
The facility-level self care improvement scores are calculated using the following 14 steps.

Step 1. Choose the 12 month window for which we will select episodes. This is the four consecutive calendar quarters ending with the most recent calendar quarter for which both MDS data and CARE tool data are available for use in the measure.

Step 2. Identify all MDS discharge assessments (in which we understand the CARE tool items will be embedded) with a discharge date that fell within the 12 month window identified in Step 1.

Step 3. For each MDS tool discharge assessment identified in Step 2, identify the corresponding MDS admission assessment (in which we understand the CARE tool items will be embedded). An MDS assessment is identified as an admission assessment if A0310F == "01" (entry record). Note that the admission date may lie before the 12 month window defined in Step 1. The period of time from the admission date (corresponding with the MDS admission assessment) through to the discharge date (corresponding with the MDS discharge assessment) is called an "episode". If no MDS admission assessment was found, discard the discharge assessment from all subsequent steps.

Step 4. Identify all MDS admission assessments that indicate the admission to the SNF was from the hospital, another SNF or IRF. An MDS admission assessment indicates that the SNF admission was from a hospital when MDS item "A1800 Entered From" coded as "03 Acute Care Hospital" or "02 Another nursing home or swing bed" or "05 inpatient rehabilitation facility" of "09 Long Term Care Hospital". The MDS item A1600 indicates the date of entry to the SNF.

Step 5. For any admission or discharge CARE tool item (that enters the calculation of the self-care improvement scores) with letter code "S" (activity not attempted due to safety concerns), A. Task attempted but not completed, N. Not applicable and P. Patient Refused were recoded to "1" on a six point rating scale (indicating full functional dependence).

Step 6. Apply the self care improvement measure's exclusions (see s.11), and exclude any episode that did not involve either physical or occupational therapy. The clinical measure exclusions are detailed in S.11 (Denominator exclusion details and codes). The exclusion of episodes not involving either occupational or physical therapy is as follows:

We identify the patient as having received occupational therapy if on the MDS discharge assessment:
The total number of minutes of occupational therapy in the last 7 days (O0400B1) is greater than zero; or
The most recent occupational therapy regimen (starting on the date recorded in O0400B5, and ending on the date recorded in O0400B6) intersects the episode (beginning with the CARE admission assessment's date and ending with the CARE discharge assessment's date).
We identify the patient as having received physical therapy if on the MDS discharge assessment:
The total number of minutes of physical therapy in the last 7 days (O0400C1) is greater than zero; or
The most recent physical therapy regimen (starting on the date recorded in O0400C5, and ending on the date recorded in O0400C6) intersects the episode (beginning with the CARE admission assessment's admission date and ending with the CARE discharge assessment's discharge date).
If the episode involves neither occupational nor physical therapy, as identified above, then exclude it from all subsequent steps in the calculation.

Step 7. For each episode remaining after Step 6, calculate a preliminary admission score and a discharge score as the sum of the values for the following CARE tool self care items A1 (Eating), A3 (Oral Hygiene), A4 (Toilet Hygiene), A5 (Upper Body Dressing), A6 (Lower Body Dressing), C1 (Wash Upper Body), C2 (Shower/Bath Self), C6 (Putting on/Taking off Footwear).

Each of those 8 CARE tool items takes an integer value of 1, 2, 3, 4, 5 or 6, and so the preliminary admission score will be an integer between 8 and 48, and the preliminary discharge score will be an integer between 8 and 48.

Step 8. For each episode, linearly transform the preliminary admission score and preliminary discharge score so that it lies in the range 1-100 using the following equation:
["transformed self-care admission score"]=2.475×["preliminary self-care admission score"]-18.8
["transformed self-care discharge score"]=2.475×["preliminary self-care discharge score"]-18.8

Step 9. For each episode, calculate the episode-level change score by subtracting the transformed discharge score from the transformed admission score. Each score will lie between -99 and 99.

Step 10. Calculate the national average change score as the simple mean of all episode-level change scores calculated in Step 9.

Step 11. For each episode, calculate the predicted change score using the risk adjustment methodology detailed in S.15a. That is, having prepared the risk adjustment variables in the way described in S.15a, apply the equation: [predicted change score] = 25.98 - 0.28×[patient is 85 years or older] -4.43×[dialysis while a patient] -3.83×[entered from SNF] -2.37×[oxygen while a patient] -1.06×[catheterization/ostomy] -2.87×[unhealed pressure ulcers] -7.12×[mental status] -3.33×[resident mood] -8.11×[psychiatric conditions] -4.05×[feeding tube or IV feeding] -5.43×[suctioning or tracheotomy] -2.76×[infections of the foot].

Step 12. For each episode, calculate the risk adjusted change score using the actual change score calculated in Step 9, the national average change score calculated in Step 10, and the predicted change score calculated in Step 11. The risk adjusted change score is:
["risk adjusted change score"]=(["national average change score"]-["predicted change score"])+["actual change score"]

Step 13. Exclude any facility that has fewer than 30 episodes for which we could calculate a risk adjusted change score.

Step 14. For each facility remaining after Step 13, calculate its self care improvement score as the simple mean of the risk adjusted change scores calculated in Step 12.

S.19. Calculation Algorithm/Measure Logic Diagram URL or Attachment *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not Applicable

S.21. Survey/Patient-reported data *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not Applicable

S.22. Missing data (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Missing data resulted in the individual being excluded from the calculation. We did not perform imputation of missing data for individual self care items. When the CARE Tool is implemented as part of the IMPACT Act of 2014; the self care section will be incorporated into MDS 3.0 and we anticipated missing data to be very low as is currently the case for ADL items in MDS and what we found across our four companies. The majority of missing data related to having missing discharge assessments to calculate a change from admission. This occurred due to rehospitalization and death. However, individuals who are rehospitalized and came back to the SNF would have a new admission and discharge assessment completed

S.23. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Other

S.24. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*

IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Resident Assessment Instrument Minimum Data Set (MDS) version 3.0

Continuity Assessment and Record Evaluation (CARE) tool; Self Care subscale

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Post Acute/Long Term Care Facility : Nursing Home/Skilled Nursing Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not Applicable

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
Self_Care_Measure_Testing_Submission_Form_2613.docx

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.**
Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)
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)*
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.**
The CARE Tool is currently available in a PDF document, which can be completed in written format. However, a number of software vendors for rehabilitation therapy providers have begun the process of adding the CARE item set to their online documentation systems. Two of the largest therapy software companies now provide the CARE Tool in electronic format. Currently over 48 organizations representing 1,016 SNFs have begun to use the CARE tool. One software therapy company has also developed on online secure portal where providers can submit their data to a larger database and receive confidential, secure outcome performance reports.

Skilled nursing care centers encode and electronically transmit the MDS 3.0 data set, as required by the federal government.

The IMPACT act of 2014 recently passed by Congressand signed by the President; requires the incorporation of standardized assessments for mobility and self care into the MDS by October 2018 (fiscal year 2019). This will make the data collection for this measure extremely feasible as it will be universally collected on all admissions and discharges to all SNFs in the country. The IMPACT act also requires the public reporting of functional outcome measures for SNFs.

United States. Cong. House of Representatives. IMPACT Act of 2014. 113th Cong., HR 4994. Washington: GPO, 2014.

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.**
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. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
In developing this measure, 87 skilled nursing centers agreed to complete and collect the self-care subscale of the CARE tool. A basic training program to the CARE tool was established, requiring therapists to pass a post-test. 425 therapists were trained with a 95% pass rate on the post-test. A key challenge identified was the CARE tools method of rating the patient's usual level of performance, rather than the lowest level of performance which is commonly used in various proprietary tools (e.g.,MBI, FOM and ROM) as well as in the MDS 3.0 scale. This was addressed by focusing the training and design of the clinical vignettes used in the training and test to highlight the difference in rating between the usual level of performance versus the most dependent/independent level of performance. Overall, therapists did not report any significant issues in understanding or assessing patients using the CARE scale, nor did they report that the CARE scale was more burdensome than the proprietary scales currently in use. The two commonly reported data collection issues were the inconvenience of completing a written scale and the need to complete all items on the scale. The first issue, completing a written scale, is currently being addressed by software vendors (see 3b.1.) and per the IMPACT act of 2014 will be incorporated into the MDS in the near future. The second issue, completing all items on the scale, must become an industry standard. When the CARE core self care items are incorporated into the MDS, per the IMPACT act of 2014, we anticipate that all items will be required as this is the current standard of the MDS 3.0 tool. A frequent practice among many therapy providers are to complete only items which are the focus of care. However, moving to a practice of completing all items on the CARE assessment form is integral to quality improvement and measurement efforts. This was supported by the therapy providers that participated in this study. Currently the therapy companies that do not require therapists to complete all items on their proprietary scales are also not able to generate an overall quality measure score. We were restricted to using data from therapy companies that required their therapist to complete all the items on their self care assessment tools.

In addition, one letter code from the original CARE scale developed by CMS was dropped. This code identified "M. Not attempted due to medical condition." It was determined that this letter code was unnecessary because the codes "1. Dependent" and "S. Not attempted due to safety concerns" would replace the code M in any situation. This was incorporated into the training and instructions provided to all the therapists.

We used the core self care items and scoring on the functional status section of the CARE tool and when implemented in the MDS this measure will be able to be calculated from these items.

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)*.
There are no fees or licensing requirements for use of CARE Tool or the use of the submitted quality measure. The Continuity Assessment and Record Evaluation (CARE) Tool is currently available in the public domain: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/CARE-Institutional-Admission-Assessment-Tool.pdf.

Additionally, no fees are required for the utilization of the MDS 3.0, it is publically available at http://www.resdac.org/cms-data/files/mds-3.0.

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.

Planned	Current Use (for current use provide URL)
Public Reporting Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Quality Improvement (Internal to the specific organization)	

4a.1. For each CURRENT use, checked above, provide:

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

Not Applicable

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

The CARE Tool currently is not required by CMS or others and is not part of the MDS. As a result, data on all the SNFs in the country are not available for public reporting. However, the IMPACT Act of 2014 passed by Congress and signed into law by the President in October 2014 requires the adoption of a standardized functional assessment tool in all post-acute care settings and public reporting of mobility and self care quality measures. The CARE tool (including self care and mobility assessment scales), developed by CMS, is the only assessment tool validated across all the PAC settings (Skilled Nursing Facilities (SNFs), Home Health Agencies (HHAs), Inpatient Rehab Facilities (IRFs) and LTACHs Long Term Acute Care Hospitals).

United States. Cong. House of Representatives. IMPACT Act of 2014. 113th Cong., HR 4994. Washington: GPO, 2014.

4a.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.*)

American Health Care Association (AHCA) and National Association of the Support of Long Term Care (NASL) who jointly sponsored the development of this measure have actively supported the IMPACT Act. They also participated in the development of the CARE

Tool and support its use and adoption by their respective members as a means for standardized assessment of our patients. AHCA represents nearly 10,000 of the approximately 15,000 SNFs in the country. NASL represents therapy companies as well as software vendors supporting therapy services. The AHCA Board of Governors and NASL have both endorsed the self care quality measure. AHCA and NASL have met with CMS asking that the CARE Tool self care assessment be adopted in SNFs so that outcome measures can be developed for public reporting and eventually incorporated into new payment models.
MedPAC has also called on CMS to adopt the CARE Tool to measure outcome measures (MedPAC annual report to congress 2014; page 174 Chapter 7 Post Acute Care Provider: Steps toward broad payment reforms. http://www.medpac.gov/documents/reports/chapter-7-post-acute-care-providers-steps-toward-broad-payment-reforms-(march-2014-report).pdf?sfvrsn=2).

As a result of these efforts, NASL members, including one of the largest software companies that supports therapy services in SNFs, has already incorporated the CARE Tool into their software and is working with AHCA to incorporate this quality measure into their software. This has resulted in 48 organizations representing 1,016 SNFs adopting the use of the CARE tool. To date, they have completed CARE Tool assessments on over 48,971 of patients. AHCA is also reaching out to other software companies that support therapy in SNFs to adopt the CARE Tool self care assessment and provide the necessary information to calculate the quality measures. In addition, two large NF chains will adopt the CARE tool self care assessment starting in early 2015.

AHCA also plans to incorporate this quality measure into its web-based reporting/benchmarking tool – Long Term Care Trend Tracker. This tool allows SNFs to calculate and trend a wide array of quality metrics over time and benchmark to peers – see http://www.ahcancal.org/research_data/trendtracker/Pages/default.aspx. We are currently working on a portal to allow SNFs of therapy companies to upload their self care quality measure scores and benchmark to peers. The AHCA Board of Governors has approved funds to build this portal in 2015. This information will help individuals SNFs with their internal quality improvement efforts as they look at their trends over time in improvement in self care. In addition, one of the large rehabilitation software vendors is developing similar ability for providers not using their software to upload their self care change score for quality improvement tracking and benchmarking against others. We are in the process of providing the specifications and algorithms necessary for these software vendors and NF companies to calculate the self care quality measure.

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

Not Applicable

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Not Applicable

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
Not Applicable

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)
2769: Functional Change in Self Care Score for Skilled Nursing Facilities

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

5a. Harmonization

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 completely harmonized?
No

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

While there are small differences between the functional characteristics used to measure improvement, AHCA agrees with NQF that the AHCA and UDSMR measures capture essentially the same quality domain. The key substantive differences between the two measures are:

1. The measures are based on different data collection tools:
    a. The AHCA measure is based on the CARE tool, and the UDSMR measure is based on the FIM – two distinct assessment instruments that cannot be harmonized. While there are several CARE and FIM items that collect similar information, the wording of the questions when similar and the rating scales used are different between the CARE and FIM. CARE items are rated on a 1-6 scale, and FIM are on a 1-7 scale. That is, while there are similarities between the two instruments, the differences are sufficient that the two measures could not be harmonized.
2. Feasibility and usability:
    a. AHCA's measure in simple to implement on national data available from CMS for all SNFs in the country, once Section GG is added to the MDS in October 2016 (and one minor change to the numerator definition is made to the measure to align with the final set of MDS Section GG items), and will be calculated and published freely by AHCA for all SNFs in the nation. This means the AHCA measure can be used for benchmarking against other SNFs, which is critical for SNF Quality Assurance/Performance Improvement work. In the measure's current form, a number of vendors have implemented specifications in our NQF application, and are providing AHCA mobility and self care improvement measure rates, with benchmarking for their clients using CARE assessments. We do not have numbers on how many SNFs are currently using the measure, however are aware of 128 organizations representing 2,511 SNFs that have undergone CARE tool training, and therefore are likely to be doing CARE tool assessments within their organizations. Starting October 2016, all SNFs in the country will be completing the MDS Section GG assessments needed to calculate the AHCA measure, and those data will be available

through the standard Research Identifiable File CMS assessment data dissemination route.

b. While UDSMR has recently made the FIM instrument available freely for SNFs to use, SNFs, even if they were able to implement the measure (which would be difficult, for reasons described next), they would only have access to their own performance – without any ability to compare against SNFs outside their organization.
c. UDSMR's numerator is built on a special case of a very complex statistical method called multilevel Rasch analysis, and UDSMR has not provided the full specifications of this special case needed to implement their methodology, making implementation of their numerator calculation impossible using the NQF application. While the advanced and very specialist statistical programming environment "R" has an implementation of this methodology under an "open source" license (i.e., could be used for other purposes), adapting the algorithm to implement in a SNF's own data systems would require a sophisticated software engineering firm with advanced statisticians available to do the project. This is prohibitively expensive for a SNF to do, and means the measure is not accessible without a paid subscription to UDSMR's services.
d. UDSMR has employed IRF Case Mix Groups (IRF-CMGs) to risk stratify the measure (though they appear to have renamed them as SNF Case Mix Groups in the NQF application). In order to implement CMS's IRF-CMG grouper algorithm, available on the CMS IRF-PPS website, the SNF would either need to implement the Visual Basic/C++ software libraries on their own system, or implement the algorithm CMS lays out in their software documentation; both of these approaches require either very specialist in-house software development resources, or hiring a software engineering firm to do it for the SNF. Therefore, the risk stratification approach further reduces the feasibility and usability of the measure, unless, again, the SNF pays UDSMR for their subscription services.

3. Numerator definition:
    a. Self care items:
        i. The items included in the AHCA and UDSMR self care improvement measures are similar (though the wording and rating scales differ – see above), though UDSMR adds two cognition items to their measure. Our group of clinical experts felt cognition was a different domain than self-care and should be included in a separate measure.
        ii. The AHCA self care improvement measures include the following CARE tool items in the numerator:
            1. Eating
            2. Oral Hygiene
            3. Toilet Hygiene
            4. Upper Body Dressing
            5. Lower Body Dressing
            6. Wash Upper Body
            7. Shower / Bathe
            8. Putting on / taking off footwear
        iii. The UDSMR self care improvement measure includes the following FIM tool items in the numerator:
            1. Eating
            2. Grooming
            3. Dressing Upper Body
            4. Dressing Lower Body
            5. Toileting, Bowel
            6. Expression

7. Memory

b. The AHCA measure sums functional independence scores (ranging 1=fully dependent to 6 = fully independent) on the included items at admission and discharge, and calculates the difference. The FIM uses a different approach relying on Rasch model. During development, we explored using a multilevel Rasch model, similar to UDSMR. However, we found the facility-level performance scores for the measure using Rasch correlated almost exactly with those from simply adding the independence scores together, with a correlation coefficient of 0.99. Therefore, because employing a Rash model approach significantly damaged the feasibility and usability of the measure, as well as the user's ability to interpret the meaning of the performance scores, we adopted the simpler approach.
c. UDSMR, as stated, has used this very complex approach without any data comparing to a simpler approach that is easier and more feasible to implement for providers. Additionally, UDSMR has not provided full specifications in the NQF application of how to calculate the measure using this approach limiting our ability to fully comment on the harmonization. .

4. Risk model development and specifications
   a. The AHCA measure risk model was developed by a workgroup of clinicians and statisticians, choosing clinically appropriate risk factors for the measure in the SNF setting, who then iteratively analyzed patterns and reviewed the clinical meaning of the to develop and test the final statistical model. This was a rigorous process, and we laid the process and testing of the model out in our answers to NQF questions 2b4.4b, 2b4.5, 2b4.9 and 2b4.10 to evaluate the adequacy of the risk factors and the overall adequacy of the risk stratification approach.
   b. UDSMR used IRF case mix groups to risk stratify the measure, but renamed them SNF-CMGs. . UDSMR application does not provide information on how they developed their approach, their testing of it, and the effect of the risk stratification approach on the measure. This makes it difficult for us to comment on harmonization of our approach with theirs. They claim that "Patients within the same CMG are expected to have similar resource utilization needs and similar outcomes" but do not provide any data or references supporting the use of IRF CMGs applied to SNF population. We are unfamiliar with SNF_CMGs and therefore cannot comment on harmonization with our risk adjustment approach.
   c. We are not aware of any common practice of applying IRF case mix groups to risk adjust SNF quality measures by CMS or skilled nursing centers in general. We are also not aware of any testing of the appropriateness of applying IRF case mix groups to SNF patients or to the UDSMR SNF measure. SNF benefit eligibility is simply that the patient had a 3-day hospital stay. IRF benefit eligibility is that the patient can handle 3 hours of intensive rehabilitative services for 5 consecutive days, and the IRF must have 60% of its patients in 13 specified clinical groups. These differences in the IRF and SNF benefit rules mean the IRF and SNF patient populations and functional needs are fundamentally different, which raises the question about whether this approach is appropriate for the SNF population or for their measure, particularly given UDSMR declined to complete this section of the application.
   d. Additionally, the IRF-CMG grouper algorithm that UDSMR has employed in their measure relies on a mixture of assessment information (which can be taken from the FIM), but also relies heavily on diagnosis coding. It is well known that SNF diagnosis coding on SNF-PPS claims and on the MDS is extremely sparse and unreliable, and reporting practices vary by SNF and by organization. SNF diagnosis codes, except for HIV, are not currently used for payment or any other purpose. Some SNFs, for example, copy the hospital diagnosis codes presented on the hospital's transfer documents onto the MDS assessments and SNF claims; others will record their own, but because SNFs typically do not have physicians on staff, the accuracy of the

diagnosis coding is questionable. The unreliability of SNF diagnosis coding, therefore, translates into unreliability of the measure's risk stratification approach. Having not tested their risk stratification approach, however, it is not possible to know the extent of this problem.

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.)
NQF should not endorse both sets of measures, as they are measuring the same thing. We believe that NQF should endorse AHCA's measures because they are more feasible and usable on the CMS's national assessments infrastructure, and they will be published freely for all SNFs in the country, creating a complete and live national Quality Assurance/Performance Improvement (QAPI) discussion about functional outcomes. UDSMR's measures cannot accomplish the same thing, for all the reasons mentioned above.

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:** Self_Care_Appendix_2613.docx

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** American Health Care Association
Co.2 **Point of Contact:** Urvi, Patel, ushah@ahca.org, 202-842-4444-
Co.3 **Measure Developer if different from Measure Steward:** The Moran Company
Co.4 **Point of Contact:** Rachel, Feldman, rlfeldman@themorancompany.com, 703-841-8405-

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.
Mary Ousley, Ousley and Associates- Expert Panel Co-chair
Howie Groff, Tealwood Care Centers- Expert Panel Co-chair

Members of the Expert Panel Include:
Cynthia Morton, NASL
Martha Schram, Aegis Therapies
Bill Goulding, Aegis Therapies
Mary Van De Kamp, Kindred/Rehab Care
Matt Sivret, Kindred/Rehab Care
Phil Fogg, Marquis Companies
Tracy Fritts, Consonus Healthcare
Garry Pezzano, Genesis Rehab Services
Felicia Chew, Genesis Rehab Services
Mike Morris, Genesis Rehab Services
John Barber, White Oak Manor
Rick Black, HCR Manor Care

Leigh Ann Frick, Heritage Health Care
Doug Burr, Health Care Navigator
Katarika Lewis, Halcyon Rehabilitation
Victoria Cruce Hollar, Halcyon Rehabilitation
Chris Castel, Accelerated Care Plus
Ellen Strunk, Rehab, Resources and Consulting

The expert panel met regularly and provided guidance on risk adjustment, exclusions, measure specifications and use of the measure.

Observers:
Mary Pratt, CMS
Tara McMullen, CMS

Members of measure steward, American Healthcare Association:
David Gifford
Courtney Bishnoi
Urvi Patel
James Muller

Members of the project contractor, The Moran Company:
Iara Woody
Chris Young
Rachel Feldman
Peter Gruhn

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2014
Ad.3 **Month and Year of most recent revision:** 06, 2014
Ad.4 **What is your frequency for review/update of this measure?** Two Years
Ad.5 **When is the next scheduled review/update for this measure?** 12, 2015

Ad.6 **Copyright statement:** None
Ad.7 **Disclaimers:** None

Ad.8 **Additional Information/Comments:** None