TO: Consensus Standards Approval Committee (CSAC)

FR: Sarah Sampsel, Suzanne Theberge, Kirsten Reed

RE: Person and Family Centered Care Phase 3 Member Voting Results

DA: October 4, 2016

The CSAC will review recommendations from the Person and Family Centered Care Phase 3 project at its October 11 conference call.

This memo includes a summary of the project, recommended measures, and themes identified from and responses to the public and member comments.

Member voting on these recommended measures ended on September 30, 2016.

Accompanying this memo are the following documents:

1. **PFCC Draft Report.** The draft report has been updated to reflect the changes made following the Standing Committee discussion of public and member comments. The complete draft report and supplemental materials are available on the project page.
2. **Comment table**. Staff has identified themes within the comments received. This table lists 21 comments received and the NQF/Standing Committee responses.

BACKGROUND

Ensuring person and family centered care is a core concept embedded in the National Quality Strategy priority of ensuring that each person and family is engaged as partners in their care. Person and family-centered care encompasses patient and family engagement in care, including shared decision-making and preparation and activation for self-care management, and the outcomes of interest to patients receiving healthcare services, including health-related quality of life, functional status, symptoms and symptom burden, and experience with care. Due to the large number of person- and family-centered care measures, maintenance review of endorsed measures and consideration of new measures is taking place over several phases. The 2014-2015 phase focused on reviewing experience with care measures, and the 2015-2016 phase examined clinician and patient-assessed measures of functional status. This third phase of work offered an opportunity to review and consider measures across the spectrum of person and family centered care, including experience of care, functional status, shared decision making and symptoms and symptom management.

DRAFT REPORT

The PFCC Draft Report presents the results of the evaluation of 13 measures considered under the CDP. (Note: Measure #2967: CAHPS Home and Community Based Services Experience of Care Measures, includes 19 measures grouped under one NQF number.) All 13 are recommended for endorsement as voluntary consensus standards suitable for accountability and quality improvement. The measures were evaluated against the 2015 version of the measure evaluation criteria.

	MAINTENANCE	NEW	TOTAL
Measures considered	1	12	13
Withdrawn from consideration	-	-	-
Recommended	1	12	13
Not recommended	-	-	-
Reasons not Recommended	-	-	-

CSAC ACTION REQUIRED

Pursuant to the CDP, the CSAC may consider approval of 13 candidate consensus standards (details of the evaluation are available via the links).

PFCC Measures Recommended for Endorsement:

- 0420: Pain Assessment and Follow Up, Centers for Medicare & Medicaid Services (CMS)
  Overall Suitability for Endorsement: Y-20; N-1
- 2614: CoreQ: Short Stay Discharge Measure, American Health Care Association (AHCA)
  Overall Suitability for Endorsement: Y-16; N-1
- 2615: CoreQ: Long-Stay Resident Measure, AHCA
  Overall Suitability for Endorsement: Y-17; N-1
- 2616: CoreQ: Long-Stay Family Measure, AHCA
  Overall Suitability for Endorsement: Y-17; N-1
- 2769: Functional Change: Change in Self Care Score for Skilled Nursing Facilities, Uniform Data System for Medical Rehabilitation (UDSMR)
  Overall Suitability for Endorsement: Y-16; N-3
- 2774: Functional Change: Change in Mobility Score for Skilled Nursing Facilities, UDSMR
  Overall Suitability for Endorsement: Y-15; N-4
- 2775: Functional Change: Change in Motor Score for Skilled Nursing Facilities, UDSMR
  Overall Suitability for Endorsement: Y-15; N-4
- 2776: Functional Change: Change in Motor Score for Long Term Acute Care Facilities, UDSMR
  Overall Suitability for Endorsement: Y-16; N-0
- 2777: Functional Change: Change in Self Care Score for Long Term Acute Care Facilities, UDSMR
  Overall Suitability for Endorsement: Y-16; N-0
- 2778: Functional Change: Change in Mobility Score for Long Term Acute Care Facilities, UDSMR
  Overall Suitability for Endorsement: Y-16; N-0
- 2958: Informed, Patient Centered Hip and Knee Replacement Surgery, Massachusetts General Hospital
  Overall Suitability for Endorsement: Y-16; N-0
- 2962: Shared Decision Making, Healthwise
  Overall Suitability for Endorsement: Y-19; N-0
- 2967: CAHPS Home and Community Based Services (HCBS) Survey Measures, CMS

Overall Suitability for Endorsement: Y-15; N-1

COMMENTS AND THEIR DISPOSITION

NQF received 21 comments from 11 organizations (including 3 member organizations) and individuals pertaining to the general draft report and to the measures under consideration.

A table of comments submitted during the comment period, with the responses to each comment and the actions taken by the Standing Committee and measure developers, is posted to the PFCC project page under the Public and Member Comment section.

Comment Themes and Committee Responses

Comments about specific measure specifications and rationale were forwarded to the developers, who were invited to respond.

During the review of all comments, the Standing Committee had the benefit of developer responses. Committee members focused their discussion on measures or topic areas with the most significant and recurring issues.

Theme 1 – Support for Measures & Committee's Work

Six comments were received that supported the endorsement of measures the Committee recommended, or with the general focus of the Committee's work. One commenter submitted comments on #2769: Functional Change: Change in Self Care Score for Skilled Nursing Facilities and #2774: Functional Change: Change in Mobility Score for Skilled Nursing Facilities noting the importance of including toileting in these measures.

An additional set of comments (three comments total) focused on the preference for the functional status change measures based on the FIM tool for use in Long Term Acute Care Facilities (LTACs). The commenter noted "*We support the use of the existing metrics related to FIM scoring from UDS/CMS. Our infra structures are built accordingly and our decision making is focused on this relevant data. The negative impact on the care delivery and cost of care will be overwhelming while offering little to no value for the catastrophic/traumatic population of patients we serve.*"

Committee Response: Thank you for your comment.

Theme 2 – Competing Measures

One comment noted a preference for the CARE tool-based measure, stating "*USDmR Functional Measures based on the FIM Tool has multiple problems including a) data is problematic as it's based on LTAC with little variation shown, b) there is overlap in the data with other tools (e.g., CARE) and c) would be are burdensome for clinicians, particularly nurses who collect these data. For example, there is overlap with the CARE Tool data which is already being collected and measures already validated. For PAC settings choosing to use the FIM there is also overlap.*"

Developer Response: During the public comment period, the developers were asked to respond to the question of competing measures. Both developers, AHCA and UDSMR, submitted responses. The full responses are too extensive to include in this memo, so they are posted on SharePoint and are summarized below:

- AHCA notes that the measures are based on different data collection tools and they state their measures are more feasible and more usable for several reasons. They summarize differences in the numerator and risk model development and specifications. (*Information is included in the related & competing section of the measure worksheet.*)
- UDSMR summarizes the similarity between the measures and also provides detail on the differences. They also note that the UDSMR measures have more data available to substantiate their strength based on the long history of use of the instrument in a variety of post-acute care settings. Additionally, the developer summarizes differences in how the assessments are conducted and can be used across patient populations and facilities. Because the Committee had previously questioned feasibility and use and usability of the metrics, UDSMR provides information on the benefits and payment model for their subscription services, but note that use of the FIM instrument is free. (*Information is included in the letter posted to the measure folder on SharePoint*.)

Committee Response: Given the complex nature of these measures and the need to ensure that everything be addressed fairly, the related/competing measures discussion will be deferred to an off-cycle review, most likely to occur in 2017. The Committee does not have enough time in the current phase of work to fully and fairly discuss and make a decision; therefore, measures #2769 and #2774 are recommended for endorsement and measures #2612 and #2613 will continue to be endorsed and the discussion and decision on harmonization and best in class will be deferred to 2017.

Theme 3 – Request for Reconsideration: HCBS Measure

The majority of the comments received focused on the Home and Community Based Services Survey measures, many urging the Committee to reconsider their recommendation due to the importance of the measure topic. This measure submission includes 19 measures within 5 topic areas. Two of the measure sets did not pass performance gap and the remaining 3 sets did not pass reliability. Commenters noted the need for outcome measures (particularly patient-reported) and the lack of measures for home and community based services.

Developer Response: The developer has requested reconsideration. To support this request, they have submitted additional information supporting the survey and measures, as requested by the Committee, as well as additional testing data.

Committee Response: During their post-comment call, the Committee reviewed and discussed the additional information that was submitted by the developer and re-voted on all NQF criteria (with the exception of evidence, which passed during the in-person meeting). In re-voting on the criteria, this measure passed and was recommended by the Committee. Committee members also noted that the public commenters did not appear to have a full understanding of the NQF process of measure endorsement. Committee members agreed that the process had been followed and stressed the need to fully evaluate measures on each evaluation criterion in order to make a sound recommendation.

Theme 4 – Measure Gaps

Commenters noted some additional gap areas in the PFCC portfolio, including:

- Self-care measures to help families as they take on the caregiving tasks associated with aging/recovering at home
- Measures that specifically address eliciting and aligning patient goals with their plan of care
- Inclusion of the palliative care population in shared decision making measures

Committee Response: Thank you for your comment. These items have been added to the measure gaps list in the report.

NQF MEMBER VOTING RESULTS

All 13 of the recommended measures were approved with 100 % approval. Complete voting results are detailed in Appendix A.

Representatives of 13 member organizations voted; no votes were received from the Public/Community Health Agency and Supplier/Industry Councils. (Links are provided to the full measure summary evaluation tables in Appendix B.)

Appendix A: NQF Member Voting Results

Measure #0420 PAIN ASSESSMENT AND FOLLOW-UP

Measure Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer	1	0	2	3	100%
Health Plan	1	0	0	1	100%
Health Professional	2	0	0	2	100%
Provider Organizations	1	0	0	1	100%
Public/Community Health Agency	0	0	0	0	
Purchaser	2	0	0	2	100%
QMRI	3	0	1	4	100%
Supplier/Industry	0	0	0	0	
All Councils	**10**	**0**	**3**	**13**	**100%**
Percentage of councils approving (>60%)					100%
Average council percentage approval					100%

*equation: Yes/ (Total - Abstain)

Voting Comment:

- American Occupational Therapy Association: Occupational therapists are eligible providers and report this measure as a part of PQRS for outpatient therapy. While this measure is indeed important for understanding and addressing pain, the measure does not capture how or if the pain is limiting function of the client or patient. The NQF PFCC project has identified numerous areas where additional measure development is needed. One of these areas includes "The next level of functional measures … that focus on functional restoration, becoming independent and non-medical outcomes". AOTA would encourage the inclusion of pain's effect on function in this category of measures.

Measure #2614 COREQ SHORT STAY DISCHARGE MEASURE

Measure Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer	1	0	2	3	100%
Health Plan	1	0	0	1	100%
Health Professional	2	0	0	2	100%
Provider Organizations	1	0	0	1	100%
Public/Community Health Agency	0	0	0	0	
Purchaser	2	0	0	2	100%
QMRI	2	0	2	4	100%
Supplier/Industry	0	0	0	0	
All Councils	**9**	**0**	**4**	**13**	**100%**
Percentage of councils approving (>60%)					100%

Average council percentage approval	100%

*equation: Yes/ (Total - Abstain)

Voting Comment:

- American Occupational Therapy Association: AOTA appreciates that AHCA has appropriately differentiated between stays over 100 days and those 100 days or less.

Measure # 2615 COREQ LONG-STAY RESIDENT MEASURE

Measure Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer	1	0	2	3	100%
Health Plan	1	0	0	1	100%
Health Professional	2	0	0	2	100%
Provider Organizations	1	0	0	1	100%
Public/Community Health Agency	0	0	0	0	
Purchaser	2	0	0	2	100%
QMRI	2	0	2	4	100%
Supplier/Industry	0	0	0	0	
All Councils	**9**	**0**	**4**	**13**	**100%**
Percentage of councils approving (>60%)					100%
Average council percentage approval					100%

*equation: Yes/ (Total - Abstain)

Voting Comment:

- American Occupational Therapy Association: AOTA appreciates that AHCA has appropriately differentiated between stays over 100 days and those 100 days or less. While the measure itself is very similar to measure #2614, the population of interest can be quite different.

Measure #2616 COREQ LONG-STAY FAMILY MEASURE

Measure Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer	1	0	2	3	100%
Health Plan	1	0	0	1	100%
Health Professional	2	0	0	2	100%
Provider Organizations	1	0	0	1	100%
Public/Community Health Agency	0	0	0	0	
Purchaser	2	0	0	2	100%
QMRI	2	0	2	4	100%
Supplier/Industry	0	0	0	0	
All Councils	**9**	**0**	**4**	**13**	**100%**
Percentage of councils approving (>60%)					100%

Average council percentage approval	100%

*equation: Yes/ (Total - Abstain)

Measure # 2769 FUNCTIONAL CHANGE: CHANGE IN SELF CARE SCORE FOR SKILLED NURSING FACILITIES

Measure Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer	1	0	2	3	100%
Health Plan	1	0	0	1	100%
Health Professional	2	0	0	2	100%
Provider Organizations	1	0	0	1	100%
Public/Community Health Agency	0	0	0	0	
Purchaser	2	0	0	2	100%
QMRI	2	0	2	4	100%
Supplier/Industry	0	0	0	0	
All Councils	**9**	**0**	**4**	**13**	**100%**
Percentage of councils approving (>60%)					100%
Average council percentage approval					100%

*equation: Yes/ (Total - Abstain)

Voting Comment

- American Occupational Therapy Association: During the comment period prior to the initial review, a commenter discussed the importance of the Jimmo v. Sebelius case which reaffirmed that that functional improvement cannot be the determining factor in Medicare reimbursement of services. These six measures (as well as other functional measures) do not necessarily limit the post-acute care settings in providing services to maintain or slow the deterioration of conditions. However, improper interpretation of implementation of these measures into Medicare assessment tools may lead to denial of care or may result in an interpretation that maintenance therapy is of less value. While AOTA supports the endorsement of these measures, we also encourage agencies to consider the implications of functional maintenance in their use and interpretation of them. The committee identified long term acute care (LTAC) as an especially difficult practice area in which improvement may not be expected. While this is likely true of LTAC, it is also true of a subset of patients that may receive skilled services including occupational therapy in a SNF, especially patients who have severely debilitating chronic conditions.

  AOTA appreciates the PFCC committee's decision to delay the best in class conversation until it can be more fully discussed and considered.

Measure #2774 FUNCTIONAL CHANGE: CHANGE IN MOBILITY SCORE FOR SKILLED NURSING FACILITIES

Measure Council	Yes	No	Abstain	Total	%

				Votes	Approval*
Consumer	1	0	2	3	100%
Health Plan	1	0	0	1	100%
Health Professional	2	0	0	2	100%
Provider Organizations	1	0	0	1	100%
Public/Community Health Agency	0	0	0	0	
Purchaser	2	0	0	2	100%
QMRI	2	0	2	4	100%
Supplier/Industry	0	0	0	0	
All Councils	**9**	**0**	**4**	**13**	**100%**
Percentage of councils approving (>60%)					100%
Average council percentage approval					100%

*equation: Yes/ (Total - Abstain)

Voting Comments

- American Occupational Therapy Association: See comments for Measure #2769.

Measure #2775 FUNCTIONAL CHANGE: CHANGE IN MOTOR SCORE FOR SKILLED NURSING FACILITIES

Measure Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer	1	0	2	3	100%
Health Plan	1	0	0	1	100%
Health Professional	2	0	0	2	100%
Provider Organizations	1	0	0	1	100%
Public/Community Health Agency	0	0	0	0	
Purchaser	2	0	0	2	100%
QMRI	2	0	2	4	100%
Supplier/Industry	0	0	0	0	
All Councils	**9**	**0**	**4**	**13**	**100%**
Percentage of councils approving (>60%)					100%
Average council percentage approval					100%

*equation: Yes/ (Total - Abstain)

Voting Comments

- American Occupational Therapy Association: See comments for Measure #2769.

Measure #2776 FUNCTIONAL CHANGE: CHANGE IN MOTOR SCORE FOR LONG TERM ACUTE CARE FACILITIES

Measure Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer	1	0	2	3	100%
Health Plan	1	0	0	1	100%
Health Professional	2	0	0	2	100%
Provider Organizations	1	0	0	1	100%
Public/Community Health Agency	0	0	0	0	
Purchaser	2	0	0	2	100%
QMRI	2	0	2	4	100%
Supplier/Industry	0	0	0	0	
All Councils	**9**	**0**	**4**	**13**	**100%**
Percentage of councils approving (>60%)					100%
Average council percentage approval					100%

*equation: Yes/ (Total - Abstain)

Voting Comment:

- American Occupational Therapy Association: See comments for Measure #2769.

Measure #2777 FUNCTIONAL CHANGE: CHANGE IN SELF CARE SCORE FOR LONG TERM ACUTE CARE FACILITIES

Measure Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer	1	0	2	3	100%
Health Plan	1	0	0	1	100%
Health Professional	2	0	0	2	100%
Provider Organizations	1	0	0	1	100%
Public/Community Health Agency	0	0	0	0	
Purchaser	2	0	0	2	100%
QMRI	2	0	2	4	100%
Supplier/Industry	0	0	0	0	
All Councils	**9**	**0**	**4**	**13**	**100%**
Percentage of councils approving (>60%)					100%
Average council percentage approval					100%

*equation: Yes/ (Total - Abstain)

Voting Comment:

- American Occupational Therapy Association: See comments for Measure #2769.

Measure #2778 FUNCTIONAL CHANGE: CHANGE IN MOBILITY SCORE FOR LONG TERM ACUTE CARE FACILITIES

Measure Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer	1	0	2	3	100%
Health Plan	1	0	0	1	100%
Health Professional	1	0	1	2	100%
Provider Organizations	1	0	0	1	100%
Public/Community Health Agency	0	0	0	0	
Purchaser	2	0	0	2	100%
QMRI	2	0	2	4	100%
Supplier/Industry	0	0	0	0	
All Councils	**8**	**0**	**5**	**13**	**100%**
Percentage of councils approving (>60%)					100%
Average council percentage approval					100%

*equation: Yes/ (Total - Abstain)

Voting Comment:

- American Occupational Therapy Association: See comments for Measure #2769.

Measure #2958 INFORMED, PATIENT CENTERED (IPC) HIP AND KNEE REPLACEMENT SURGERY

Measure Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer	1	0	2	3	100%
Health Plan	1	0	0	1	100%
Health Professional	2	0	0	2	100%
Provider Organizations	1	0	0	1	100%
Public/Community Health Agency	0	0	0	0	
Purchaser	1	0	1	2	100%
QMRI	2	0	2	4	100%
Supplier/Industry	0	0	0	0	
All Councils	**8**	**0**	**5**	**13**	**100%**
Percentage of councils approving (>60%)					100%
Average council percentage approval					100%

*equation: Yes/ (Total - Abstain)

Measure # 2962 SHARED DECISION MAKING

Measure Council	Yes	No	Abstain	Total Votes	% Approval*

Consumer	1	0	2	3	100%
Health Plan	1	0	0	1	100%
Health Professional	2	0	0	2	100%
Provider Organizations	1	0	0	1	100%
Public/Community Health Agency	0	0	0	0	
Purchaser	2	0	0	2	100%
QMRI	3	0	1	4	100%
Supplier/Industry	0	0	0	0	
All Councils	**10**	**0**	**3**	**13**	**100%**
Percentage of councils approving (>60%)					100%
Average council percentage approval					100%

*equation: Yes/ (Total - Abstain)

Measure #2967 CAHPS® HOME AND COMMUNITY BASED SERVICES (HCBS) MEASURES

Measure Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer	3	0	0	3	100%
Health Plan	1	0	0	1	100%
Health Professional	2	0	0	2	100%
Provider Organizations	1	0	0	1	100%
Public/Community Health Agency	0	0	0	0	
Purchaser	2	0	0	2	100%
QMRI	4	0	0	4	100%
Supplier/Industry	0	0	0	0	
All Councils	**13**	**0**	**0**	**13**	**100%**
Percentage of councils approving (>60%)					100%
Average council percentage approval					100%

*equation: Yes/ (Total - Abstain)

Voting Comment:

- American Occupational Therapy Association: There is some evidence that the population served by Medicaid home and community based services may be unable to access important and medically necessary services. This can introduce a particular gap in quality of care. In fact, with appropriate services such as occupational therapy, beneficiaries may require less assistance for items identified in one of these topics. The report for voting highlights HCBS services as a priority gap area identified by the Dual Eligible project. AOTA would encourage exploration of measures in this gap, especially for persons who receive long-term support services through Medicaid.

Appendix B- Measure Evaluation Summary Tables

LEGEND: Y = Yes; N = No; H = High; M = Moderate; L = Low; I = Insufficient

Measures Recommended

0420 Pain Assessment and Follow-Up
Submission | Specifications
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 **Numerator Statement**: 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 **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 **Adjustment/Stratification**: No risk adjustment or risk stratification **Level of Analysis:** Clinician : Group/Practice, Clinician : Individual **Setting of Care:** Ambulatory Care : Clinician Office/Clinic, Behavioral Health/Psychiatric : Outpatient, Ambulatory Care : Outpatient Rehabilitation **Type of Measure**: Process **Data Source**: Administrative claims, Paper Medical Records **Measure Steward**: Centers for Medicare & Medicaid Services
STANDING COMMITTEE MEETING [06/06/2016] **1. Importance to Measure and Report: The measure meets the Importance criteria** (1a. Evidence, 1b. Performance Gap) 1a. Evidence: **H-0**; **M-7**; **L-5**; **I-9**; 1b. Performance Gap: **H-9**; **M-12**; **L-0**; **I-0;** Evidence Exception: **Y-19; N-2** Rationale: • Committee members noted that assessing pain is crucial in order to treat it, but the literature that demonstrates better outcomes after such assessment is limited. The developer agreed that all of the published studies that look at the effectiveness of pain assessment are low quality, both those reporting a difference and those reporting no difference; the developer recommends further study. • Committee members with expertise in palliative care also noted limitations of the current pain scales used to do these assessments, both in terms of providing meaningful data (particularly since the FACES scale was designed for children and the evidence for it was on low back pain) and because the assessments are relatively easy to game; patients who report higher numbers get stronger medications. The developer noted the measure does not require any particular pain assessment tool. • Patient advocates on the Committee strongly supported the need for pain assessment. • The measure was originally developed for use by physical and occupational therapists. • This is a process measure, but the Committee agreed it is one step closer to an outcome measure since it includes both the assessment of pain and the development of a plan to address it. In response to

questions, the developer noted that the intent of the measure is not to specify treatment, but to create a care plan, which could include non-pharmacological interventions.
- The Committee discussed concerns around over-prescription of opioids and the opioid epidemic, noting that much research still needs to be done on how to best manage pain, and that providers are currently being encouraged to limit opioid prescriptions.
- The developer clarified that pain needed to be assessed by a valid pain tool, not just a simple question or two.
- The Committee struggled with the lack of direct evidence linking better outcomes to pain assessment. The developer noted part of the reason there is a lack of data are because it is very difficult to do a controlled study on this particular topic since obtaining a patient history and developing a treatment plan is the standard of care; therefore, to not do an assessment in order to study outcomes would be unethical. Committee members noted there is general evidence supporting the practice of monitoring symptoms and then altering practice based on that monitoring.
- NQF staff noted, in response to questions, that other endorsed pain measures have passed the evidence criteria by using insufficient evidence with exception option. In the vote on evidence, the Committee did not reach consensus. However, the Committee did reach consensus on the evidence exception, and the measure moved forward.
- There are differences in assessment and treatment rates by race/ethnicity (Asian 76.2%, Black 68.2%, Hispanic 79.1%, Native 73.6%, White 84.2%, Other 79.6%, Unknown 86.1%), and this was highlighted as a gap area demonstrating the need for continuing endorsement of this measure.

2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-5**; **M-14**; **L-1**; **I-1** 2b. Validity: **H-2**; **M-11**; **L-6**; **I-2**

Rationale:

- Committee members requested information on why patients under 18 were excluded, given that there are good tools for measuring pain in children. The developer explained that the measure was developed for use in adults and hasn't been updated, and agreed that was a concern.
- One Committee member had questions about the reliability testing at the provider level and how well the measure demonstrates variability between providers; another noted that 90% of the providers reporting are in the 25th percentile. The developer responded that only 10% of eligible providers are reporting and so they believe the scores are skewed towards high performance, especially since this is typical of voluntary measures; however, they cannot confirm this.
- During the validity discussion, the Committee noted that while most providers (over 90%) are reporting very high scores, the mean is 82%; this means a small group of providers are reporting very poor scores. The measure also passed the validity criteria.

3. Feasibility: H-14; M-6; L-1; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)

Rationale:

- The measure uses administrative data and has been in use for several years, so the Committee had no concerns with the feasibility.

4. Usability and Use: H-10; M-11; L-0; I-0

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

Rationale:

- The measure has been in use for several years and the Committee did not have major concerns with the usability. However, they did note the potential unintended consequence of narcotics overuse.
- Committee members noted that patients with chronic complex conditions are actually more likely to under report pain. Ultimately the Committee agreed that the potential unintended consequences did

not outweigh the importance of the measure.

5. Related and Competing Measures

This measure is related, but not competing, with a number of 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

Standing Committee Recommendation for Endorsement: **Y-20**; **N-1**

Rationale

- This measure did not pass Evidence but moved forward on the Evidence Exception.

6. Public and Member Comment

- No comments were received on this measure.
- The developer submitted additional evidence in support of the measure, but since the measure was already recommended, the Committee did not make any changes to this recommendation.

7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X

8. Board of Directors Vote: Y-X; N-X

9. Appeals

2614 CoreQ: Short Stay Discharge Measure

Submission | Specifications

Description: The measure calculates the percentage of individuals discharged in a six month time period from a SNF, within 100 days of admission, who are satisfied (see: S.5 for details of the time-frame). This patient reported outcome measure is based on the CoreQ: Short Stay Discharge questionnaire that utilizes four items.

Numerator Statement: The measure assesses the number of patients who are discharged from a SNF, within 100 days of admission, who are satisfied. The numerator is the sum of the individuals in the facility that have an average satisfaction score of =>3 for the four questions on the CoreQ: Short Stay Discharge questionnaire.

Denominator Statement: The denominator includes all of the patients that are admitted to the SNF, regardless of payor source, for post-acute care, that are discharged within 100 days; who receive the survey (e.g. people meeting exclusions do not receive a questionnaire) and who respond to the CoreQ: Short Stay Discharge questionnaire within the time window (See: S.5).

Exclusions: Exclusions used are made at the time of sample selection and include:

(1) Patients who died during their SNF stay;

(2) Patients discharged to a hospital, another SNF, psychiatric facility, inpatient rehabilitation facility or long term care hospital;

(3) Patients with court appointed legal guardian for all decisions;

(4) Patients discharged on hospice;

(5) Patients who left the nursing facility against medical advice (AMA);

(6) Patients who have dementia impairing their ability to answer the questionnaire defined as having a BIMS score on the MDS as 7 or lower. [Note: we understand that some SNCCs may not have information on cognitive function available to help with sample selection. In that case, we suggest administering the survey to all residents and assume that those with cognitive impairment will not complete the survey or have someone else complete on

their behalf which in either case will exclude them from the analysis.]

(7) Patients who responded after the two month response period; and

(8) Patients whose responses were filled out by someone else.

Adjustment/Stratification: No risk adjustment or risk stratification

Level of Analysis: Facility

Setting of Care: Post Acute/Long Term Care Facility: Nursing Home/Skilled Nursing Facility

Type of Measure: PRO

Data Source: Healthcare Provider Survey**Measure Steward**: American Health Care Association

STANDING COMMITTEE MEETING [06/06/2016]

1. Importance to Measure and Report: <u>The measure meets the Importance criteria</u>

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Y-17; N-1**; 1b. Performance Gap: **H-7**; **M-10**; **L-1**; **I-0**

<u>Rationale</u>:

- Committee members noted that this is a very significant measure for those who go into a nursing home or a SNF who will not stay indefinitely or for a long period of time. Measuring patient satisfaction and the rate of discharges back into the community is very important to measurement as including the patient and their preferences is becoming an integral part of healthcare's changing landscape. Additionally, measuring and reporting satisfaction with care helps patients and their families choose and trust a healthcare facility and can help facilities improve the quality of the care they provide.
- One Committee member had a question about the scale being used for this measure and felt that the choice of the response scale (poor, average, good, very good, and excellent) seemed heavily weighted towards positive responses. The developer explained that they did focus groups and cognitive testing of different response scales from ten points down to four point Likert scales and found that no matter how they captured responses, they had different satisfaction scores but the relative ranking remained the same.
- Overall, Committee members liked that there was a conceptual framework at the beginning of the measure submission form that linked the measure with information on additional improvement programs, organizational change initiatives, and policies that are going on both at the federal level and the facility level.

2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-6**; **M-8**; **L-4**; **I-0** 2b. Validity: **H-6**; **M-9**; **L-3**; **I-0**

<u>Rationale</u>:

- One Committee member felt that the exclusions may limit the generalizability to a small proportion of facility nursing home patients.
- There was additional concern around the consistency of implementation across facilities and the possibility that scores could be compromised by the low response rate.
- Committee members also questioned the test/retest reliability at the patient level and sample size. The developer explained that the data elements were tested using a test-retest methodology: the survey was sent out and responses received from 853 patients; 100 were re-surveyed one month later. The developer responded to these concerns by saying that while morbidity does occur, and may affect the data, there is an emphasis on making sure that both the voice of the patient and the voice of the family are heard.
- There was also discussion around cognitive impairment and the effect this has on the survey's overall responses. The developer agreed that cognitive impairment does have an effect in this setting and that by having everyone use the BIMs score, which is used to get a snapshot of how well someone is functioning cognitively at a given moment, allows for a more consistent approach across all nursing home residents. A standardized approach helps reduce the incidence of gaming.

- One Committee member had a question on the methodology used to reduce the number of items in the tool and how they got from 22 to 4 items without losing some precision. The developer responded that the process was extremely iterative and was done hundreds of times. The purpose of this was to try and get to the items that were capturing the most satisfaction information that did not overlap with other items and if two items correlated very highly, it made sense to drop one of them.
- All members agreed with the decision not to risk adjust as it is inappropriate to control out differences based on sociodemographic factors.
- Cognitive testing was done with family members, residents, and with short stay residents. The developers collected more than 100 responses from each population at facilities in Pittsburgh. This testing was conducted by reading questions and having the testing groups respond back based on what they thought was being asked and if they felt it could be asked differently. The Committee indicated providing the results of this testing, although supplemental, would have been useful information.

3. Feasibility: H-5; M-13; L-0; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)

Rationale:

- The Committee agreed that this tool is timely as there is currently no required experience of care reporting or measurement in the SNF population.
- Members appreciated that this tool is brief especially since the staffing in this area tends to be very sparse.

4. Usability and Use: H-5; M-11; L-2; I-0

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

Rationale:

- The Committee did not have any concerns or questions about the use and usability.

5. Related and Competing Measures

- This measure was identified as related with #2615: CoreQ: Long-Stay Resident Measure and #2616: CoreQ: Long-Stay Family Measure, submitted by the same developer.

Standing Committee Recommendation for Endorsement: **Y-16**; **N-1**

6. Public and Member Comment

- No comments were received on this measure.

7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X

8. Board of Directors Vote: Y-X; N-X

9. Appeals

2615 CoreQ: Long-Stay Resident Measure

Submission | Specifications

Description: The measure calculates the percentage of long-stay residents, those living in the facility for 100 days or more, who are satisfied (see: S.5 for details of the time-frame). This patient reported outcome measure is based on the CoreQ: Long-Stay Resident questionnaire that is a three item questionnaire.

Numerator Statement: The numerator is the sum of the individuals in the facility that have an average satisfaction score of =>3 for the three questions on the CoreQ: Long -Stay Resident questionnaire.

Denominator Statement: The denominator includes all of the residents that have been in the SNF for 100 days or

more regardless of payer status; who received the CoreQ: Long-Stay Resident questionnaire (e.g. people meeting exclusions do not receive the questionnaire), who responded to the questionnaire within the two month time window, who did not have the questionnaire completed by somebody other than the resident, and who did not have more than one item missing.

Exclusions: Exclusions made at the time of sample selection are the following: (1) Residents who have poor cognition defined by the BIMS score; (2) residents receiving hospice; (3) residents with a legal court appointed guardian; and (4) residents who have lived in the SNF for less than 100 days.

Additionally, once the survey is administered, the following exclusions are applied: a) surveys received outside of the time window (two months after the administration date) b) surveys that have more than one questionnaire item missing c) surveys from residents who indicate that someone else answered the questions for the resident. (Note this does not include cases where the resident solely had help such as reading the questions or writing down their responses.)

Adjustment/Stratification: No risk adjustment or risk stratification

Level of Analysis: Facility

Setting of Care: Post Acute/Long Term Care Facility: Nursing Home/Skilled Nursing Facility

Type of Measure: PRO

Data Source: Healthcare Provider Survey

Measure Steward: American Health Care Association

STANDING COMMITTEE MEETING [06/06/2016]

1. Importance to Measure and Report: <u>The measure meets the Importance criteria</u>

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Y-17; N-1**; 1b. Performance Gap: **H-7**; **M-10**; **L-1**; **I-0**

<u>Rationale</u>:

- The Committee agreed that this measure was very similar to #2614 and did not require additional discussion or voting. They agreed to carry the votes on evidence and gap from the previous measure.

2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-6**; **M-8**; **L-4**; **I-0** 2b. Validity: **H-6**; **M-9**; **L-3**; **I-0**

<u>Rationale</u>:

- One Committee member had questions around validity and whether staff members were allowed to fill out the surveys on patients' behalf. The developer responded that while there is no way to stop them from filling it out on the patient's behalf, if they do indicate as such, their data will be excluded.
- The Committee agreed that this measure was very similar to #2614 and did not require additional discussion or voting. They agreed to carry the votes on reliability and validity from the previous measure.

3. Feasibility: H-5; M-13; L-0; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)

<u>Rationale</u>:

- The Committee agreed that this measure was very similar to #2614 and did not require additional discussion or voting. They agreed to carry the votes on feasibility from the previous measure.

4. Usability and Use: H-5; M-11; L-2; I-0

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

<u>Rationale</u>:

- The Committee agreed that this measure was very similar to #2614 and did not require additional discussion or voting. They agreed to carry the votes on usability and use from the previous measure.

5. Related and Competing Measures

- This measure was identified as related with #2614: CoreQ: Short-Stay Discharge Measure and #2616: CoreQ: Long-Stay Family Measure, submitted by the same developer.

Standing Committee Recommendation for Endorsement: **Y-17**; **N-1**

- **Although the Committee carried the discussions and votes through to each of these SNF experience of care measures, they voted separately for Recommendation for Endorsement.**

6. Public and Member Comment

- No comments were received on this measure.

7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X

8. Board of Directors Vote: Y-X; N-X

9. Appeals

2616 CoreQ: Long-Stay Family Measure

Submission | Specifications

Description: The measure calculates the percentage of family or designated responsible party for long stay residents (i.e., residents living in the facility for 100 days or more), who are satisfied (see: S.5 for details of the timeframe). This consumer reported outcome measure is based on the CoreQ: Long-Stay Family questionnaire that has three items.

Numerator Statement: The numerator assesses the number of family or designated responsible party for long stay residents that are satisfied. Specifically, the numerator is the sum of the family or designated responsible party members for long stay residents that have an average satisfaction score of =>3 for the three questions on the CoreQ: Long-Stay Family questionnaire.

Denominator Statement: The target population is family or designated responsible party members of a resident residing in a SNF for at least 100 days. The denominator includes all of the individuals in the target population who respond to the CoreQ: Long-Stay Family questionnaire within the two month time window (see S.5) who do not meet the exclusion criteria (see S.10).

Exclusions: Please note, the resident representative for each current resident is initially eligible regardless of their being a family member or not. Only one primary contact per resident should be selected.

Exclusions made at the time of sample selection include: (1) family or designated responsible party for residents with hospice; (2) family or designated responsible party for residents with a legal court appointed guardian; (3) representatives of residents who have lived in the SNF for less than 100 days; and (4) representatives who reside in another country.

Additionally, once the survey is administered, the following exclusions are applied: a) surveys received outside of the time window (more than two months after the administration date) and b) surveys that have more than one questionnaire item missing.

Adjustment/Stratification: No risk adjustment or risk stratification

Level of Analysis: Facility

Setting of Care: Post Acute/Long Term Care Facility: Nursing Home/Skilled Nursing Facility

Type of Measure: PRO

Data Source: Healthcare Provider Survey

Measure Steward: American Health Care Association

STANDING COMMITTEE MEETING [06/06/2016]

1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Y-17; N-1**; 1b. Performance Gap: **H-7**; **M-10**; **L-1**; **I-0;**

Rationale:

- The Committee agreed that this measure was very similar to #2614 and did not require additional discussion or voting. They agreed to carry the votes on evidence and gap from the previous measure.

2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-6**; **M-8**; **L-4**; **I-0** 2b. Validity: **H-6**; **M-9**; **L-3**; **I-0**

Rationale:

- The Committee agreed that this measure was very similar to #2614 and did not require additional discussion or voting. They agreed to carry the votes on reliability and validity from the previous measure.

3. Feasibility: H-5; M-13; L-0; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)

Rationale:

- The Committee agreed that this measure was very similar to #2614 and did not require additional discussion or voting. They agreed to carry the votes on feasibility from the previous measure.

4. Usability and Use: H-5; M-11; L-2; I-0

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

Rationale:

- One Committee member had a question about other languages that this survey was available in. The developer responded and said that it is currently only available in English but they are exploring other options for the future.
- The Committee agreed that this measure was very similar to #2614 and did not require additional discussion or voting. They agreed to carry the votes on usability and use from the previous measure.

5. Related and Competing Measures

- This measure was identified as related with #2614: CoreQ: Short-Stay Discharge Measure and #2615: CoreQ: Long-Stay Resident Measure, submitted by the same developer.

Standing Committee Recommendation for Endorsement: **Y-17**; **N-1**

- **Although the Committee carried the discussions and votes through to each of these SNF experience of care measures, they voted separately for Recommendation for Endorsement.**

6. Public and Member Comment

- No comments were received on this measure.

7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X

8. Board of Directors Vote: Y-X; N-X

9. Appeals

2769 Functional Change: Change in Self Care Score for Skilled Nursing Facilities

Submission | Specifications

Description: 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. The time frame for the measure is 12 months. The measure includes the following 8 items: Eating, Grooming, Dressing

Upper Body, Dressing Lower Body, Toileting, Bowel, Expression, and Memory.

Numerator Statement: Average change in rasch derived self-care functional score from admission to discharge at the facility level, including items: Eating, Grooming, Dressing Upper Body, Dressing Lower Body, Toileting, Bowel, Expression, and Memory.

Denominator Statement: Facility adjusted expected change in rasch derived values, adjusted for SNF-CMG (Skilled Nursing Facility Case Mix Group), based on impairment type, admission functional status, and age

Exclusions: Excluded in the measure are patients who died in the SNF or patients less than 18 years old.

Adjustment/Stratification: Stratification by risk category/subgroup

Level of Analysis: Facility

Setting of Care: Post Acute/Long Term Care Facility: Nursing Home/Skilled Nursing Facility

Type of Measure: Outcome

Data Source: Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Paper Medical Records

Measure Steward: Uniform Data System for Medical Rehabilitation, a division of UB Foundation Activities, Inc. and its successor in interest, UDSMR, LLC.

STANDING COMMITTEE MEETING [06/06/2016]

1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Y-19; N-0**; 1b. Performance Gap: **H-3**; **M-13**; **L-1**; **I-2**

Rationale:

- This measure uses the FIM tool, and is similar to measures endorsed in the PFCC Phase 2 project; those measures were for inpatient rehabilitation facilities while this measure is set in SNFs.
- The Committee was concerned about the overlap and potential burden of data collection between this measure, which uses the FIM tool, and the mobility and self-care functional status changes measures that are derived from the CARE tool as well as data collected through the Minimum Data Set (MDS). The developer explained this measure includes self-care items of both cognitive and physical function, while the CARE measure for self-care only covers physical function. They also noted that data shows a change over time when using the FIM-based measures but the change is not shown for reports using the MDS, which leads the developer to conclude they are measuring different functional domains.
- The submission form for this measure focuses on restoration and improvement of function as a goal of rehabilitation, which is a component of skilled nursing. Committee members expressed concern about this, noting that for some patients, the goal may be to maintain function and thus facilities would be able to use these measures to potentially "cherry pick" patients and only choose those that have the opportunity to improve. They also brought up Jimmo v. Sebelius, the Medicare law requiring SNFs to provide services to maintain or slow deterioration of function, even for patients that cannot improve. The developer agreed their measure submission placed a heavy emphasis on improvement, but they are amenable to adding language that clarified the measures can not only identify improvement, but those patients who are maintaining or declining in function. They also indicated the performance measure is an aggregated population measure, and thus was looking more globally at performance of a facility versus singling out individuals.
- The developer explained how the expected performance range was developed; since as the Committee noted, almost half of the facilities reporting were below expectations in 2014. Using rasch modeling, the developer calculated the average patient's function for each measure and compared each facility to that number; expected performance therefore is a statistical value rather than a benchmark.
- The Committee requested a distribution of the facility level scores to better assess the performance gaps. Committee members also requested information on whether functional performance has changed over time in response to the efforts made to improve quality in this area. The developer noted that differences are clear when the data are stratified.

2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-3**; **M-9**; **L-2**; **I-5** 2b. Validity: **H-4**; **M-13**; **L-1**; **I-1**

Rationale:

- In response to questions on the exclusions, the developer explained they have another tool, the WeeFIM, for children under 18 that accounts for differences between adults and children; the developer thought it would make the measure simpler to exclude children from this measure. Patients who died in care are excluded due to the lack of discharge scores, which would make it impossible to measure change.
- The developer also noted that missing data are not an issue because their system requires all the information needed to calculate the measure. However, they are not able to track the percentage of patients that data was not collected on.
- The Committee had questions about the 12-month window, since stays at SNFs are less than 12 months, and the developer explained that it was intended to allow smaller facilities to collect enough data (at least 30 cases). They also explained that facilities receive internal quarterly reports.
- After the submission, NQF suggested that the developer perform inter-class correlation testing at the facility level to provide additional reliability data. The results from this testing were submitted prior to the Committee meeting. The intra-class correlation (ICC) between facilities was -0.03 with a P value of 0.59; according to the developer this is a poor score which demonstrates a good amount of variability between facilities. The within-facility ICC was 0.87 with a P value of less than 0.001, demonstrating consistency in ratings within a particular facility.
- Committee members questioned this interpretation and indicated the results demonstrate a lot variation within a single facility but not a lot of variation between facilities; lots of difference at the patient level makes it challenging to understand whether there are facility variations. It was noted by the Committee that while this type of testing is important for identifying variation within a facility and reliability, understanding the reliability of the performance measure *between* facilities requires different testing. The Committee was asked to vote and make their recommendations with the data provided, and the developers are being provided the opportunity to assess if they have data to support the additional analyses for consideration. The Committee specifically suggested the developers could do generalized estimation equations; and then perform the ICC.
- During the validity discussion, Committee members asked about the response rate. The measure is currently voluntary, and the developers do not know the exact response rate but they believe it is the majority of patients.

3. Feasibility: H-5; M-11; L-3; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)

Rationale:

- The Committee had no major concerns with the feasibility of the measure.

4. Usability and Use: H-3; M-11; L-2; I-3

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

Rationale:

- The developer clarified that the FIM tool is free to use, but is not in the public domain, as the developer wants to maintain the integrity of the instrument through uniform use. Use of the tool requires training, and the developer does offer certification training to subscribers. Committee members noted concerns around burden for facilities that have not trained their staff. They noted that several groups of providers will need to be involved and there will need to be periodic retraining in response to staff turnover. The developer responded that there is free training available, and that it is important that the staff collecting the data understand what they are measuring to ensure the data are good. The Committee agreed that training to ensure accurate data collection is especially important for measures that may be used for payment.

- Non-subscriber facilities have access to the instrument and the published training guide, but not the data repository. The developer clarified that if the measures are endorsed and adopted for use in federal programs, CMS will be able to use them royalty-free in any venue they choose.
- Committee members reiterated the potential unintended consequences of this measure in relation to Jimmo vs. Sebelius, with the possibility of making patients who cannot improve "less desirable", but they noted this could be an issue for many measures and was not enough of a reason to not endorse this measure.
- Committee members also warned that this measure should not be used to make comparisons to other levels of care (IRF vs. SNF for example) as they are not comparable (in terms of patient complexity, levels of care, etc.), even though the measures are very similar. The developer stated that they agree, but others do not, and that collecting the same data across venues will provide data to prove that point.
- The developer also noted the IMPACT Act requires common measures that can be used across settings of care.
- Committee members who use the FIM-based measures in the IRF setting noted that they receive results at a facility, regional, and national level, so that they can compare themselves to other providers. The developer added that they provide reports for facilities that take into account the average patient's change as well as the discharge dispositions, adjusting for case mix.
- In response to questions about potential manipulation of data, the developer added that they do not usually see major drastic changes in performance over short times without other significant changes at the facility such as a change in administration.

5. Related and Competing Measures

- This measure was identified as competing with measure #2613: CARE: Improvement in Self Care. The Committee has decided to delay the related/competing discussion until 2017, when additional consideration of the complexities of measuring functional status can be discussed and data from previously endorsed measures will be available.

Standing Committee Recommendation for Endorsement: **Y-16**; **N-3**

6. Public and Member Comment

- This measure received one supportive comment, commending the developer for the inclusion of toileting.

7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X

8. Board of Directors Vote: Y-X; N-X

9. Appeals

2774 Functional Change: Change in Mobility Score for Skilled Nursing Facilities

Submission | Specifications

Description: 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. 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.

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.

Denominator Statement: Facility adjusted adjusted expected change in rasch derived values, adjusted at the Skilled Nursing Facility Case Mix Group level.

Exclusions: Excluded in the measure are patients who died in the SNF or patients less than 18 years old.

Adjustment/Stratification: Stratification by risk category/subgroup

Level of Analysis: Facility

Setting of Care: Post Acute/Long Term Care Facility : Nursing Home/Skilled Nursing Facility

Type of Measure: Outcome

Data Source: Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Registry

Measure Steward: Uniform Data System for Medical Rehabilitation

STANDING COMMITTEE MEETING [06/06/2016]

1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Y-19; N-0**; 1b. Performance Gap: **H-3**; **M-13**; **L-1**; **I-2**

Rationale:

- This measure is very similar to #2769: Functional Change in Self Care. The Committee questioned why there is also a Functional Change in Motor Skills measure, which includes both the self-care and mobility domains. The developer explained that there are patients who may have restricted mobility, but still be able to do self-care; the different measures are intended to provide different levels of functional measurement for different facilities and different patients. It was further clarified that the composite score would not require duplicate data collection since it is the same data.
- The developer reported that they did not see differences in performance by sociodemographic factors.
- The Committee agreed that this measure was very similar to #2769 and did not require additional discussion or voting. They agreed to carry the votes on evidence and gap from the previous measure.

2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-3**; **M-9**; **L-2**; **I-5** 2b. Validity: **H-4**; **M-13**; **L-1**; **I-1**

Rationale:

- The Committee agreed that this measure was very similar to #2769 and did not require additional discussion or voting. They agreed to carry the votes on reliability and validity from the previous measure.

3. Feasibility: H-5; M-11; L-3; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)

Rationale:

- The Committee agreed that this measure was very similar to #2769 and did not require additional discussion or voting. They agreed to carry the vote on feasibility from the previous measure.

4. Usability and Use: H-3; M-11; L-2; I-3

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

Rationale:

- The Committee agreed that this measure was very similar to #2769 and did not require additional discussion or voting. They agreed to carry the vote on usability from the previous measure.

5. Related and Competing Measures

- This measure was identified as competing with measure #2612: CARE: Improvement in Mobility. The Committee has decided to delay the related/competing discussion until 2017, when additional consideration of the complexities of measuring functional status can be discussed and data from previously endorsed measures will be available.

Standing Committee Recommendation for Endorsement: **Y-15**; **N-4**

Rationale • Although the committee decided to carry both the discussions and voting across the UDSMR SNF measures, they voted on overall recommendation for endorsement for each individually.
6. Public and Member Comment • This measure received one supportive comment, commending the developer for the inclusion of toileting.
7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X
8. Board of Directors Vote: Y-X; N-X
9. Appeals

2775 Functional Change: Change in Motor Score for Skilled Nursing Facilities
Submission | Specifications
Description: Change in rasch derived values of motor function from admission to discharge among adult short term rehabilitation skilled nursing facility patients aged 18 years and older who were discharged alive. The time frame 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. **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 SNF or patients who died within the SNF are excluded. **Denominator Statement**: Facility adjusted expected change in rasch derived values, adjusted for SNF-CMG (Skilled Nursing Facility Case Mix Group), based on impairment type, admission functional status, and age. **Exclusions**: Patients age at admission less than 18 years old Patients who died in the SNF. **Adjustment/Stratification**: Stratification by risk category/subgroup **Level of Analysis:** Facility **Setting of Care:** Post Acute/Long Term Care Facility : Nursing Home/Skilled Nursing Facility **Type of Measure**: Outcome **Data Source**: Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Paper Medical Records **Measure Steward**: Uniform Data System for Medical Rehabilitation, a division of UB Foundation Activities, Inc. and its successor in interest, UDSMR, LLC.
STANDING COMMITTEE MEETING [06/06/2016] **1. Importance to Measure and Report: The measure meets the Importance criteria** (1a. Evidence, 1b. Performance Gap) 1a. Evidence: **Y-19; N-0**; 1b. Performance Gap: **H-3**; **M-13**; **L-1**; **I-2** Rationale: • The Committee agreed that this measure was very similar to #2769 and did not require additional discussion or voting. They agreed to carry the votes on evidence and gap from the previous measure.

2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-3**; **M-9**; **L-2**; **I-5** 2b. Validity: **H-4**; **M-13**; **L-1**; **I-1**

Rationale:

- The Committee agreed that this measure was very similar to #2769 and did not require additional discussion or voting. They agreed to carry the votes on reliability and validity from the previous measure.

3. Feasibility: H-5; M-11; L-3; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)

Rationale:

- The Committee agreed that this measure was very similar to #2769 and did not require additional discussion or voting. They agreed to carry the vote on feasibility from the previous measure.

4. Usability and Use: H-3; M-11; L-2; I-3

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

Rationale:

- The Committee agreed that this measure was very similar to #2769 and did not require additional discussion or voting. They agreed to carry the vote on usability from the previous measure.

5. Related and Competing Measures

- This measure is the "parent" to the mobility and self-care measures that have been identified as competing with measures #2612: CARE Improvement in Mobility and #2613: Care Improvement in Self-Care. The Committee has decided to delay the related/competing discussion until 2017, when additional consideration of the complexities of measuring functional status can be discussed and data from previously endorsed measures will be available.

Standing Committee Recommendation for Endorsement: **Y-15**; **N-4**

Rationale

- Although the committee decided to carry both the discussions and voting across the UDSMR SNF measures, they voted on overall recommendation for endorsement for each individually.

6. Public and Member Comment

- No comments were received on this measure.

7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X

8. Board of Directors Vote: Y-X; N-X

9. Appeals

2962 Shared Decision Making Process

Submission | Specifications

Description: This measure assesses the extent to which health care providers actually involve patients in a decision-making process when there is more than one reasonable option. This proposal is to focus on patients who have undergone any one of 7 common, important surgical procedures: total replacement of the knee or hip, lower back surgery for spinal stenosis of herniated disc, radical prostatectomy for prostate cancer, mastectomy for early stage breast cancer or percutaneous coronary intervention (PCI) for stable angina. Patients answer four questions (scored 0 to 4) about their interactions with providers about the decision to have the procedure, and the measure of the extent to which a provider or provider group is practicing shared decision making for a particular procedure

is the average score from their responding patients who had the procedure.

Numerator Statement: Patient answers to four questions about whether not 4 essential elements of shared decision making (laying out options, discussing the reasons to have the intervention and not to have the intervention, and asking for patient input) were part of the interactions with providers when the decision was made to have the procedure.

Denominator Statement: All responding patients who have undergone one of the following 7 surgical procedures: back surgery for a herniated disc; back surgery for spinal stenosis; knee replacement for osteoarthritis of the knee; hip replacement for osteoarthritis of the hip; radical prostatectomy for prostate cancer; percutaneous coronary intervention (PCI) for stable angina, and mastectomy for early stage breast cancer.

Exclusions: For back, hip, knee, and prostate surgery patients, there are no exclusions, so long as the surgery is for the designated condition.

PCI patients who had a heart attack within 4 weeks of the PCI procedure are excluded, as are those who have had previous coronary artery procedures (either PCI or CABG).

For patients who have mastectomy, patients who had had a prior lumpectomy for breast cancer in the same breast and patients who have not been diagnosed with breast cancer (who are having prophylactic mastectomies) are excluded.

Adjustment/Stratification: No risk adjustment or risk stratification

Level of Analysis: Clinician : Group/Practice

Setting of Care: Ambulatory Care : Clinician Office/Clinic

Type of Measure: PRO

Data Source: Patient Reported Data/Survey

Measure Steward: Informed Medical Decisions Foundation, a division of Healthwise

STANDING COMMITTEE MEETING [06/07/2016]

1. Importance to Measure and Report: <u>The measure meets the Importance criteria</u>

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Y-19; N-0**; 1b. Performance Gap: **H-10**; **M-8**; **L-1**; **I-0**

<u>Rationale</u>:

- The Committee agreed that this PRO-PM demonstrated the value of the shared decision making approach and the 4 items within the questionnaire are based on the 3 essential concepts it was designed to address (ensuring that patients were informed and understood their issues; ensuring there was meaningful interaction between provider and patient to provide the opportunity for the patient's voice to be heard during the decision making process; and aligning the patient's goals, concerns and priorities by the end of the process).
- The developer noted that this measure works best when applied to a specific kind of decision (e.g. decision to have surgery for herniated disc).
- The Committee voted to pass the evidence criteria for this measure.
- The Committee noted a lack of diversity in the testing population and voiced their concerns about whether the developer had looked at health literacy and how that was accounted for in the tool, as health literacy level has been shown to impact people's ability to participate in the decision making process.
- The developer agreed the testing population was less heterogeneous than they would have liked, but said they reviewed the research carefully and were unable to find evidence that any groups (i.e., older or low educated patients) are resistant to being involved in decision making.
- Committee members noted the gap was smaller for back surgery patients. The developer explained they thought it was that back pain is often not fixable by surgery so back surgeons work particularly hard to ensure patients are aware of the pros and cons.
- Committee members wanted to know if there were some procedures not included because there is less of a choice in whether to have the procedure. The developer noted that shared decision making is appropriate for all medical care, but the procedures in the measure were selected because they thought

they could both reliably sample the people who had made a decision at a given point, and they had the data.
- In response to Committee questions, the developer noted that discussing the patient's goals and concerns is an essential part of real shared decision making, but they wanted to keep the questionnaire as short as they could. They hope to expand it in the future.
- The measure passed performance gap.

2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-2**; **M-14**; **L-3**; **I-0** 2b. Validity: **H-2**; **M-15**; **L-2**; **I-0**

Rationale:

- The Committee discussed the challenge of reliably identifying people who are faced with a decision and decided not to do something (i.e., surgery), and agreed that there needs to be a reliable way of getting the same population of patients who have had the same experiences. They also understood the limitations the lack of such data places on measurement.
- The developer addressed this concern by stating the goals of the measure are to be able to identify a set of people that should actually have had a choice and to ensure that the same kind of patients can reliably be identified and compared across multiple clinical sites.
- The Committee agreed that although the numbers in the testing population were small, there would likely be more variability with larger numbers and hospitals involved in the shared decision making process.
- Committee members requested more information about response rates, particularly the rate needed to ensure a valid sample (and whether that was feasible), and whether the homogeneity of the sample impacted the response rate. The developer noted that the way the survey is presented affects response rates, particularly when the clinical site follows up to ensure it is returned. The developer noted they are working on shared decision making on pregnancy and childbirth related care, but didn't currently have the data to include them. Their research thus far indicates the questions would not only apply to white men or to orthopedic decisions.
- The developer provided additional information on the cognitive testing performed.
- In response to questions, the developer explained they had randomized practices (not within practices) to ensure the samples were not contaminated.
- General consensus was reached that this measure met the reliability and validity criteria.

3. Feasibility: H-0; M-12; L-7; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)

Rationale:

- The Committee noted that mailed surveys and follow up calls are expensive and asked if there were IT ways to make gathering data easier. The developer explained that currently, the response rates were much lower with online surveys but it might be more feasible in other populations.
- The developer also noted this would not be performed all the time, but might be collected on back patients one year and hip patients the next, reducing the burden on any particular group.
- Despite some concerns, the measure did pass feasibility.

4. Usability and Use: H-6; M-11; L-2; I-0

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

Rationale:

- In response to questions, the developer noted that getting shared decision making right involves more providers than just physicians, and there are training programs available to teach providers how to incorporate shared decision making into their care.
- Although this measure is not currently in use for public reporting (and the developer indicated that while they support public reporting, they cannot have a direct role in implementing it), the Committee noted

that accountable care organization evaluations could find shared decision making useful within quality improvement.

5. Related and Competing Measures

- The developer identified measure #1741: Patient Experience with Surgical Care Based on the Consumer Assessment of Healthcare Providers and Systems (CAHPS)® Surgical Care Survey, as a related measure and stated that the approved PCMH and ACO CAHPS measures of shared decision making were adaptations of the measures they developed and are proposing. The Committee agreed they are similar but not competing.
- The developer mentioned that the measures were used for respondents who reported they had discussed starting or stopping a prescription medication (for PCMH) and for patients who reported discussion a prescription medication or a procedure with a provider (ACO). The shared decision making measure focuses measuring the process of patient and provider interaction and the extent it meets the process of shared-decision making.

Standing Committee Recommendation for Endorsement: **Y-19**; **N-0**

6. Public and Member Comment

- This measure received two comments. One noted the continuing gap area of measures that "specifically address eliciting and aligning patient goals with their plan of care"; this was added to the measure gaps list.
- The second comment supported both the concept of shared decision making and the measure, as well as the Committee's consensus that shared decision making is appropriate for all patients, but suggested the measure needs to go further to include more patients. This comment was referred to the developer for a response:
    - **Developer response:** *This is a response to the public comment by Mark Dann from Compassion and Choices about the proposed measure of Shared Decision Making (SDM) Process. We proposed that the measure would be used to assess the extent to which patients reported they had an interaction with their providers that reflected shared decision making when they had decided to have any one of 7 surgical interventions: knee or hip replacement, surgery for herniated disc or spinal stenosis, PCI for stable angina, mastectomy for early stage breast cancer, or prostatectomy for prostate cancer. Mr. Dann comments that he would hope that the measure would be used to assess decision making for a much broader set of decisions for which there is more than one reasonable treatment approach. We could not agree more.*
    Our proposal to NQF focused on those 7 decisions because we could reliably identify patients who had made those decisions and because we had data that supported the validity of the measure to distinguish those clinical practices making a special effort to do shared decision making from "usual care". However, we have used those questions in survey studies of patients who have made decisions about taking new long-term medications and about screening for cancer, as well as surgical procedures other than the 7 listed. We are confident that the measure does provide valid information about the decision making process, and we are very hopeful that we and others can collect data that help make the case for the value of extending the use of these questions to a wide variety of decisions beyond the 7 targeted in our proposal.

7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X

8. Board of Directors Vote: Y-X; N-X

9. Appeals

2776 Functional Change: Change in Motor Score in Long Term Acute Care Facilities

Submission | Specifications

Description: 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.

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.

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.

Exclusions: Patients age at admission less than 18 years old

Patients who died in the LTAC.

Adjustment/Stratification: Stratification by risk category/subgroup

Level of Analysis: Facility

Setting of Care: Post Acute/Long Term Care Facility: Long Term Acute Care Hospital

Type of Measure: Outcome

Data Source: Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Paper Medical Records

Measure Steward: Uniform Data System for Medical Rehabilitation, a division of UB Foundation Activities, Inc. and its successor in interest, UDSMR, LLC.

STANDING COMMITTEE MEETING [06/06/2016]

1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Y-18; N-1**; 1b. Performance Gap: **H-2**; **M-7**; **L-4**; **I-6 UPDATED GAP: H-0; M-15; L-1; I-0**

Rationale:

- The Committee agreed that many of the issues discussed for measure #2769 would be applicable, as the main difference for this measure is the setting: LTAC instead of SNF. However, Committee members pointed out that LTACs are a new setting for the FIM tool, and the data on their use are limited thus far: the reliability testing was performed using data from 6 facilities and ICC testing was performed using 16 LTAC facilities, as compared to almost 200 SNFs and more than 800 IRFs using the measure. Similar to SNFs, this measure is voluntary for LTACs.
- The developer noted that the same drastic level of functional improvement is not expected or seen in LTACs, but that a slight improvement can be possible. The measure can be used to find patients who are starting to decline and need readmission to acute or intensive care. Patients at the lowest level – complete dependence – are also captured. In addition, the developer said that LTACs have not traditionally measured function, and they believe that asking questions about function can improve the quality of care by reminding providers of the importance of mobility and overall function.
- The measure also assesses the burden of care a patient needs by quantifying the help needed, therefore providing information needed by providers and families if patients are projected to go home.
- In response to questions, Committee members explained that patients in LTACs are medically debilitated and require serious care such intravenous or respiratory therapy, or are dependent on ventilators; patients may have spinal cord or traumatic brain injuries.
- The data presented only reflect through 2011, but Committee members noted a shrinking gap in care; the developer indicated they believe it is an artifact of the small sample size. Committee members noted that

some premiere LTACs are providing significant rehabilitation services, but were uncomfortable with agreeing there was a gap based on testing in 6 facilities, especially since 3 were in 1 state (Massachusetts). The developer explained they now had more data on more facilities and could provide it if requested.
- At the in-person meeting, the measure passed evidence but did not reach consensus on performance gap.
- During the comment period, the developer provided additional data from 39 facilities. They submitted additional information on performance gap demonstrating opportunity for improvement. At the post-comment call, the Committee revoted on gap and the measure passed this subcriterion.

2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-2**; **M-8**; **L-3**; **I-6** 2b. Validity: **H-1**; **M-10**; **L-4**; **I-4**

UPDATED Reliability: H-1; M-13; L-2; I-0 Validity: H-1; M-14; L-1; I-0

Rationale:

- The Committee agreed that aside from the new setting and limited data, the issues for this measure were very similar to #2769 and did not require additional discussion on the reliability and validity. The Committee did not reach consensus on reliability or validity.
- The developer completed and submitted additional testing during the comment period, including the intra-class correlation scores. The ICC score for this measure was 0.905, p<.001, which they indicated demonstrates very high consistency. Concurrent and predictive validity scores were also provided and were high. The Committee agreed the new testing data was sufficient to ensure the reliability and validity of the measure, and it passed both during the voting on the post-comment call.

3. Feasibility: H-4; M-11; L-3; I-1

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)

Rationale:

- The Committee had no major concerns around the feasibility for this measure.

4. Usability and Use: H-2; M-9; L-3; I-5 UPDATED H-2; M-14; L-0; I-0

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

Rationale:

- The Committee agreed that the issues for this measure were very similar to #2769 and did not require additional discussion on the usability. The Committee did not reach consensus on usability.

5. Related and Competing Measures

- No related or competing measures noted.

Standing Committee Recommendation for Endorsement: **Y-11**; **N-8 UPDATED: Y-16; N-0**

Rationale

- After reviewing the additional data submitted by the developer, the Committee voted to recommend the measure during the post-comment call.

6. Public and Member Comment

- This measure received one comment supporting endorsement of the measure.

7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X

8. Board of Directors Vote: Y-X; N-X

9. Appeals

2777 Functional Change: Change in Self Care Score for Long Term Acute Care Facilities

Submission | Specifications

Description: Change in rasch derived values of self-care function from admission to discharge among adult patients treated in a long term acute care facility who were discharged alive. The time frame for the measure is 12 months. The measure includes the following 8 items: Eating, Grooming, Dressing Upper Body, Dressing Lower Body, Toileting, Bowel, Expression, and Memory.

Numerator Statement: Average change in rasch derived self-care functional score from admission to discharge at the facility level, including items: Eating, Grooming, Dressing Upper Body, Dressing Lower Body, Toileting, Bowel, Expression, and Memory.

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

Exclusions: Excluded in the measure are patients who died in the LTAC or patients less than 18 years old.

Adjustment/Stratification: Stratification by risk category/subgroup

Level of Analysis: Facility

Setting of Care: Post Acute/Long Term Care Facility: Long Term Acute Care Hospital

Type of Measure: Outcome

Data Source: Electronic Clinical Data, Electronic Clinical Data: Electronic Health Record, Paper Medical Records

Measure Steward: Uniform Data System for Medical Rehabilitation, a division of UB Foundation Activities, Inc. and its successor in interest, UDSMR, LLC.

STANDING COMMITTEE MEETING [06/06/2016]

1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Y-18; N-1**; 1b. Performance Gap: **H-2**; **M-7**; **L-4**; **I-6 UPDATED Gap: H-0; M-15; L-1; I-0**

Rationale:

- The Committee agreed this measure is very similar to #2776: Functional Change: Change in Motor Score in Long Term Acute Care Facilities, and elected to carry the discussion and votes on the criteria over from that measure. They did not reach consensus on performance gap.
- During the comment period, the developer provided additional testing data from on 39 facilities. They submitted additional information on performance gap demonstrating opportunity for improvement. At the post-comment call, the Committee revoted on gap and the measure passed this subcriterion.

2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-2**; **M-8**; **L-3**; **I-6** 2b. Validity: **H-1**; **M-10**; **L-4**; **I-4**

UPDATED Reliability: H-1; M-13; L-2; I-0 Validity: H-1; M-14; L-1; I-0

Rationale:

- The developer noted they had provided both concurrent and predictive validity testing. They also explained they had attempted to have consistent sample sizes across facility types, which means they could show more variability in IRFs. However, they offered to provide more data on LTACs for the Committee to review.
- The Committee explained that LTACs are a new setting for both the tool and the measures, and that was why they wanted more data for this set of measures. Specifically, they requested information on the facility level distribution of results, and the ICC coefficients at the facility level.
- The Committee agreed this measure is very similar to #2776: Functional Change: Change in Motor Score in Long Term Acute Care Facilities, and elected to carry the votes on the criteria over from that measure. They did not reach consensus on either reliability or validity.

- The developer completed and submitted additional testing during the comment period, including the intra-class correlation scores. The ICC score for this measure was 0.951, p<.001, which they indicated demonstrates very high consistency. Concurrent and predictive validity scores were also provided and were high. The Committee agreed the new testing data was sufficient to ensure the reliability and validity of the measure, and it passed both during the voting on the post-comment call.

3. Feasibility: H-4; M-11; L-3; I-1

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)

Rationale:

- The Committee agreed this measure is very similar to #2776: Functional Change: Change in Motor Score in Long Term Acute Care Facilities, and elected to carry the discussion and votes on the criteria over from that measure.

4. Usability and Use: H-2; M-9; L-3; I-5 UPDATED: H-2; M-14; L-0; I-0

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

Rationale:

- The Committee agreed this measure is very similar to #2776: Functional Change: Change in Motor Score in Long Term Acute Care Facilities, and elected to carry the discussion and votes on the criteria over from that measure. The Committee did not reach consensus on usability.

5. Related and Competing Measures

- No related or competing measures noted.

Standing Committee Recommendation for Endorsement: **Y-9**; **N-10 UPDATED: Y-16; N-0**

Rationale

- After reviewing the additional data submitted during the comment period, the Committee voted to recommend the measure for endorsement.

6. Public and Member Comment

- This measure received one comment supporting endorsement of the measure.

7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X

8. Board of Directors Vote: Y-X; N-X

9. Appeals

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

Submission | Specifications

Description: 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.

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 treatmentare excluded. No other exclusions as long as the respondent has the procedure for the designated condition.

Adjustment/Stratification: No risk adjustment or risk stratification

Level of Analysis: Clinician: Group/Practice

Setting of Care: Ambulatory Care: Clinician Office/Clinic

Type of Measure: PRO

Data Source: Patient Reported Data/Survey

Measure Steward: Massachusetts General Hospital

STANDING COMMITTEE MEETING [06/07/2016]

1. Importance to Measure and Report: <u>The measure meets the Importance criteria</u>

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Y-18; N-1**; 1b. Performance Gap: **H-1**; **M-14**; **L-4**; **I-0**

<u>Rationale</u>:

- This measure assesses the extent to which patients who had elective surgery for hip or knee replacement were well informed and had a clear preference for surgery beforehand. The survey instrument is based on 6 items: 5 knowledge questions and 1 that elicits a patient's preference. These questions focus on the surgical benefits, harms, recovery time, etc. The developer received input from both patients and providers when developing the questions.
- The Committee agreed that asking a patient simple questions such as which treatment do they prefer, do they prefer to have surgery/non-surgical options, etc. should be standard for someone who is actually going to have surgery and if they are not given those options, then they should not be operated on.
- Hip and knee replacements are very common, and the Committee agreed that just because a patient is clinically eligible for one of these procedures, does not mean it is the best choice of treatment. Thus, patients who elect to have one of these procedures should be well informed about the risks and benefits and have a clear preference.
- Additional questions were raised regarding how the questions in the instrument were derived and whether they are meant to be used in conjunction with Healthwise measure #2962: Shared Decision Making. The developer explained that while measuring the quality of the decision and the idea that someone is meaningfully involved in the decision making process is important; this measure is less generic and aims to ensure that a patient is more focused on knowledge.
- During their research, the developer found that there was no correlation between a patient's perception of feeling informed and their ability to answer knowledge-specific questions.
- The measure was tested at 3 different hospitals in the same geographic region in Massachusetts and therefore is not a nationally-representative sample.
- This instrument has a Spanish version available but has not been widely used.

2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-0**; **M-11**; **L-7**; **I-1** **UPDATED Reliability: H-0; M-16; L-0; I-0** 2b. Validity: **H-0**; **M-13**; **L-5**; **I-1**

<u>Rationale</u>:

- One Committee member questioned whether the developer did not or could not compare people with high scores to low scores. The developer responded that in order to test for discriminant validity, they split patients into 2 groups and gave only 1 of the groups decision aids. When comparing the 2 groups, they found significant differences on the knowledge questions, with the decision aids group scoring much higher.
- As with other measures considered during this phase of work, the Committee suggested additional reliability testing, specifically testing at the practice level. It was suggested the developer should perform tests to assess between versus within practice variation.
- The Committee questioned how the developer found the sample of patients and the post-operative

timeline for giving the instrument given to patients. The developer noted that in order to get a reliable sample size, they had to survey patients who had received a hip or knee replacement within the last 2 years. The developer agreed that ideally, patients would be surveyed the week after surgery, but in order to collect enough data to calculate the measure, they recommended allowing a look back period of up to 2 years. The Committee questioned the ability of patients to reliably and validly recollect conversations over that length of time.

- There were additional questions around what is considered to be a passing score when completing this instrument. The developer explained that they had set the criteria and in order to be considered well-informed, a patient must answer 3 or more of the 6 questions correctly.
- Since this measure deals with both hip and knee replacement surgeries, there were concerns about why the correct answer to recovery time was the same for both when those recovering from hip surgery are functional more quickly than those recovering from knee surgery. The developer responded to these concerns by saying that the instrument was not developed to assess actual precision, but more of the general realization that recovery takes a couple of months rather than a few days or a few years. To ensure that these questions and answers remain current, a multi-stakeholder expert panel reviews them every 2 years to ensure that the answers remain accurate and are updated if needed.
- An additional comment was raised around exclusions and looking at non-elective surgeries in addition to primary surgeries. The developer agreed to look into this but also noted that the most evidence supports the importance of shared decision making for elective or preference sensitive surgeries and procedures. It was also noted that non-elective surgeries are not considered exclusions.
- A number of Committee members raised concerns with the instrument being given out up to 2 years after a surgery since so much can change in that time period; they argued that even those with a great memory would have a difficult time remembering such specific details about their surgery. In addition, they noted that patients could have done additional research after the surgery, thus giving them more knowledge than what was provided by their doctor. The developer agreed that it is important to have their knowledge assessed earlier, but explained that they have data on a study they did among breast cancer patients where they surveyed patients right after their surgery and then a year later. While they had predicted the numbers would drop, after data analysis they did not find a big difference in knowledge scores.
- Due to the testing concerns, the Committee did not reach consensus on reliability at the in-person meeting. The measure passed validity.
- During the comment period, the developer submitted additional information for review, including: a clarification of the exclusions; the addition of clear time periods for the survey to be conducted after the surgery; and more information on the methods and results of the reliability testing. They also conducted additional reliability testing by calculating the IPC within each sample. After reviewing this new data, the Committee voted that the measure met the reliability criterion.

3. Feasibility: H-0; M-15; L-3; I-1

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)

Rationale:

- Some Committee members wanted information about the burden of collecting the data and how much time is required in collecting the responses. The developer explained the patient burden is very limited as it only takes a few minutes to complete the questions. In terms of burden on the provider, the developer thought it depended on the practice as some likely already have resources in place to assess patient-reported outcomes.

4. Usability and Use: H-0; M-9; L-6; I-3 UPDATED: H-4; M-11; L-1; I-0

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

Rationale:

- The measure is currently used in a quality recognition program but is not publically reported or used in an accountability program. The developer stated they would like to see this incorporated into programs that

are assessing the quality of the surgical process of care, including whether the right patient was in the operating room, whether patients were well informed, and whether they had a clear preference for surgical treatments prior to surgery.
- A Committee member was concerned that if endorsed, this measure could be used for both evaluating quality improvement and for holding providers accountable, but the Committee member did not think the measure was ready to be used for payment programs.
- The measure is new, but is based on a patient reported survey that has been used by thousands of patients and has been well tested. The developer provided some additional information on use and usability, and at the post-comment call, the Committee agreed the measure met this criterion.

5. Related and Competing Measures

- No related or competing measures noted.

Standing Committee Recommendation for Endorsement: **Y-10**; **N-8 UPDATED: Y-16; N-0**

Rationale

- After reviewing the clarifications and new information, the Committee voted to recommend the measure during the post-comment call.

6. Public and Member Comment

- No comments were received on this measure.

7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X

8. Board of Directors Vote: Y-X; N-X

9. Appeals

2778 Functional Change: Change in Mobility Score for Long Term Acute Care Facilities

Submission

Description: 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.

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.

Denominator Statement: Facility adjusted adjusted expected change in rasch derived values, adjusted at the Case Mix Group level.

Exclusions: Excluded in the measure are patients who died in the LTAC or patients less than 18 years old.

Adjustment/Stratification: Stratification by risk category/subgroup

Level of Analysis: Facility

Setting of Care: Post Acute/Long Term Care Facility: Long Term Acute Care Hospital

Type of Measure: Outcome

Data Source: Electronic Clinical Data, Electronic Clinical Data: Electronic Health Record, Paper Medical Records

Measure Steward: Uniform Data System for Medical Rehabilitation

STANDING COMMITTEE MEETING [06/06/2016]

1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Y-18; N-1**; 1b. Performance Gap: **H-2**; **M-7**; **L-4**; **I-6 UPDATED Gap: H-0; M-15; L-1; I-0**

Rationale:

- The Committee agreed this measure is very similar to #2776: Functional Change: Change in Motor Score in Long Term Acute Care Facilities, and elected to carry the discussion and votes on evidence and performance gap over from that measure. The Committee did not reach consensus on performance gap.
- During the comment period, the developer provided additional testing data from on 39 facilities. They submitted additional information on performance gap demonstrating opportunity for improvement. At the post-comment call, the Committee revoted on gap and the measure passed this subcriterion.

2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-2**; **M-8**; **L-3**; **I-6** 2b. Validity: **H-1**; **M-10**; **L-4**; **I-4**

UPDATED Reliability: H-1; M-13; L-2; I-0 Validity: H-1; M-14; L-1; I-0

Rationale:

- The Committee agreed this measure is very similar to #2776: Functional Change: Change in Motor Score in Long Term Acute Care Facilities, and elected to carry the discussion and votes on the reliability and validity over from that measure. The Committee did not reach consensus on either reliability or validity.
- The developer completed and submitted additional testing during the comment period, including the intra-class correlation scores. The ICC score for this measure was 0.938, p<.001, which they indicated demonstrates very high consistency. Concurrent and predictive validity scores were also provided and were high. The Committee agreed the new testing data was sufficient to ensure the reliability and validity of the measure, and it passed both during the voting on the post-comment call.

3. Feasibility: H-4; M-11; L-3; I-1

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)

Rationale:

- The Committee agreed this measure is very similar to #2776: Functional Change: Change in Motor Score in Long Term Acute Care Facilities, and elected to carry the discussion and votes on feasibility over from that measure.

4. Usability and Use: H-2; M-9; L-3; I-5 UPDATED: H-2; M-14; L-0; I-0

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

Rationale:

- The Committee agreed this measure is very similar to #2776: Functional Change: Change in Motor Score in Long Term Acute Care Facilities, and elected to carry the discussion and votes on usability over from that measure. At the in-person meeting, they did not reach consensus on usability.
- The developer provided a verbal response to the concerns regarding number of LTACs using this measure. They explained that the 6 facilities used in the initial measure submission represented a sample of facilities, which they were able to augment with data from an additional 39 facilities to update their analyses. They did acknowledge the FIM is not as widely used in LTACs as compared to other settings, but it's use promotes alignment and some aspects of comparability that re needed in the market. After consideration of the additional information, the Committee agreed the measure met the usability criterion.

5. Related and Competing Measures

- No related or competing measures noted.

Standing Committee Recommendation for Endorsement: **Y-7**; **N-11 UPDATED: Y-16; N-0**

Rationale

- After reviewing the additional information and data submitted by the developer, the Committee voted to recommend the measure for endorsement.

6. Public and Member Comment • This measure received one comment supporting endorsement of the measure.
7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X
8. Board of Directors Vote: Y-X; N-X
9. Appeals

2967 CAHPS® Home and Community Based Services (HCBS) Survey Measures
Submission
Description: CAHPS® Home and Community Based Services (HCBS) Experience of Care (EoC) 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. 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

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 responseon 3 survey items

4. Choosing the services that matter to you - average proportion of respondents that gave the most positive responseon 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 responseon 3 survey items

7. Planning your time and activities - average proportion of respondents that gave the most positive responseon 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. 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.

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.

Adjustment/Stratification: Statistical risk model

Level of Analysis: HCBS Program

Setting of Care: Other: Home and Community-Based Services Program

Type of Measure: PRO

Data Source: Patient Reported Data/Survey

Measure Steward: Centers for Medicare and Medicaid Services

STANDING COMMITTEE MEETING [06/06/2016]

1. Importance to Measure and Report: this measure met the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Y-17; N-1**;

1b. Performance Gap: **Split by domain**

Scale: **H-1**; **M-2**; **L-13**; **I-2 – Did not meet the Importance Criteria**

Global Ratings: H-0; **M-10**; **L-7**; **I-1 – Did not reach consensus on the Importance Criteria**

Recommendations: H-0; **M-12**; **L-5**; **I-1 – Met the Importance Criteria**

Unmet Needs: H-9; **M-7**; **L-2**; **I-0 – Met the Importance Criteria**

Physical Safety: H-0; **M-4**; **L-7**; **I-7 – Did not meet the Importance Criteria**

UPDATED (all domains together): H-5; M-11; L-0; I-0

Rationale:

- This is a package of 19 different measures, split into 5 domains: scale, global ratings, recommendations, unmet needs, and physical safety. The measures assess experience of care for long term home and community based service programs.
- The measures are scored at the state program level (Medicaid programs including both fee-for-service and Managed Long Term Services and Supports programs), and the developer noted there are 3-11 programs per state. The programs serve groups including frail elderly; people with physical, intellectual and developmental disabilities; and people with brain injuries. The data for the measures is collected via a 95 question survey (the developer noted there are many skip patterns so not all items are asked).
- Some of the Committee had serious concerns with the level of accountability for this measure. Since

there are multiple agencies providing many staff members, the Committee was concerned it would be difficult to make the measures actionable for improvement. Committee members with experience in this area noted that while the services are provided via "a hodgepodge of a lot of different programs" what matters to consumers is that their needs are met, not who is meeting them. Therefore, an overall assessment of whether care is being provided and the quality at the aggregate level is also important, not just the quality of any individual provider.

- It was also noted that these services are vital for many people to be able to live in the community with minimal support, and are particularly important to allow young people to live on their own, away from their parents. However, people who rely on these services may not be able to follow up on care issues independently, so being asked is important.
- After an overview discussion, Committee members turned to the specific measures within the submission. They requested clarification that the endorsement would be of the measure, not the experience of care survey, and of how many measures are potentially being endorsed. In addition, they wanted more information on whether all or some components would be used. It was clarified that states could select to only report on some of the measures. Committee members noted this could affect the reliability.
- Committee members asked the developer to explain why there are both a global ratings set and a recommendations set, given that they are assessing something very similar (patient satisfaction) using a different approach. The developer indicated that consistent with CAHPS surveys, the general overall ratings and recommendations are considered behavioral intentions and the global ratings are used a validation items for those subscales. Thus you want to keep that subscale structure because it tells a program where to focus improvement.
- It was noted that some HCBS programs also provide employment services. In response, the developer noted there was a supplement regarding employment, but because so few of the people in the testing population answered in a way that would trigger the appropriate series of questions, it was not adequately tested and therefore not included for potential endorsement.
- Committee members noted that the quality of these services is tremendously important to the disability community, and that the measures could be very useful for states as they assess whether their programs are meeting goals and are effective.
- Committee members discussed the possibility of deferring the measures, noting that while they agreed they address an important area they are still very new and that questions remain about the limited amount of testing conducted thus far. NQF staff provided information on deferral and the processes around it.
- The Committee decided to vote on evidence all together, and then split the measure set into 5 measure batches and vote on each of the domains separately for performance gap and the remaining criteria. They agreed that they were not comfortable voting on it as a single measure, but also did not think 19 separate votes was appropriate. The domains are: scale measures, global measures, recommend measures, unmet needs measures, and physical safety.
- Committee members noted that some of the questions on the scale measure are similar to other surveys that patients may be receiving, and wanted to know if this would be duplicative. The developer explained that an HCBS program would likely field this survey at most once a year, and that while individuals may receive services and surveys from other providers, these will be administered either face to face or over the phone, making it a different kind of survey. They also noted that it would be conducted on a sample, not a full population, and states would likely be careful about burden for their participants.
- Each of the items in the measures are on a "never, sometimes, usually, always" scale which is then transformed to a 0 to 100 scale to make it easier to understand. After some discussion of this scale, Committee members reviewed the data provided and were concerned about the lack of room for improvement on some of the measures. They requested information on whether the sampling or something about how the survey was administered may have led to much higher scores than might be expected based on the literature. The developer agreed the scores were high. They noted it was a random sample but only respondents who passed the cognitive screening were included, and that the

modes of survey administration were appropriate for the population.

- Committee members discussed the potential for both "ceiling effect" and "floor effect" problems with the scores, given that some have very small standard deviations and some are very large, and also noted that since this is voluntary, they may only be getting high performers to participate.
- However, the Committee noted that what both HCBS providers and patients really care about is whether people are doing well, and the details are less important to measure, except to the level that the details are needed to discover whether the reason people are not doing well is due to their needs not being met. They also noted this is a patient-reported outcome measure, and the data are reported by the people receiving the services. Committee members sought and were reassured that part of the consent process of the survey made it clear that this is a care optimizing tool and that patients were not at risk of losing care based on their answers.
- While the gap on most of the measures was small, it was very high on the unmet needs category; however, the Committee was concerned that not all of this was under control of the program as the decision of what services to provide may be under the control of a state budget office.
- In a single vote, all of the measure domains passed evidence.
- The recommendation and unmet needs measures passed performance gap. The global measure did not achieve consensus on gap. The scale and physical safety measures did not pass performance gap and did not move forward in the discussion.
- During the comment period, the developer submitted additional information. The developer clarified that the small performance gaps for the personal safety-related measures are because they are "never events". The unmet need measures are also expected to have a low prevalence and therefore a small gap. With this information, the Committee agreed the measure met the performance gap subcriterion.

2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **Split by domain**

Scale: **H-X**; **M-X**; **L-X**; **I-X**

Global Ratings: H-0; **M-7**; **L-8**; **I-3**

Recommendations: H-0; **M-4**; **L-12**; **I-2**

Unmet Needs: H-1; **M-2**; **L-12**; **I-3**

Physical Safety: H-X; **M-X**; **L-X**; **I-X**

UPDATED: H-2; M-13; L-1; I-0

Validity: **H-1**; **M-14**; **L-1**; **I-0**

Rationale:

- Committee members were concerned about the exclusion of people with cognitive limitations from the measure, as this group represents a substantial part of the population receiving these services, and reiterated the need for proxy reporters. However, they noted there are typically a lot of disagreements between proxy reporters and people reporting on their own behalf. They suggested that the proxy and self-reported scores be reported separately since they may not be comparable.
- Given that some states have one program and other states have multiple programs, Committee members were concerned about being able to distinguish state variation from program variation as well as the within versus the between program variation within and across states. The developer explained that these will be administered by the states, so they might be administered differently within each state. The measures are not intended to be used to compare states to each other at this point, only to compare performance within a state. It will also be a voluntary measure.
- Committee members noted that people who cannot pass a cognitive screening would be excluded, which would include a lot of frail elderly who are receiving in-home services, and wondered whether the developer would consider including caregivers or family members. The developer explained that they had to exclude these patients for testing as they hoped for a CAHPS trademark, and CAHPS surveys do not allow proxies. However, as the testing progressed, they realized that they were receiving proxy responses

so the testing pool was expanded to allow them after a period of time. In the Testing Experience and Functional Tools (TEFT) demonstration for round 2 of data collection, TEFT state grantees are including proxies since it became clear they were necessary. However, the measure testing submitted to NQF did not include this data because at the time it had not been consistently administered to proxies.

- In response to questions, the developer confirmed that three rounds of cognitive testing had been performed in both English and Spanish. The Committee requested more information about the results of this testing and the developer agreed to provide it at a later date.
- Committee members wanted to know if the measures performed differently based on whether the survey was admitted by phone or in-person. The developer said the differences were significant on some but not all of the measures and said they recommending adjusting for survey mode to account for this.
- The Committee noted the measures were tested in 26 different programs and the total responses were 2,300; they were concerned this sample was too small. The developer explained that going forward they recommend a larger sample size (400) in order to get a reliability score of 0.7. In addition, they noted in 2012, 25% of programs have less than 400 enrollees, 30% have between 400-3,000, and 41% have 3,000-50,000 enrollees. They noted that after the 2014 HCBS rule, waiver programs are expected to consolidate and grow over time. However, other Committee members were concerned a larger sample might affect the validity as some programs will be assessed with half their population and others with a very small portion. They also noted potentially underrepresented samples such as traumatic brain injury patients.
- Committee members wanted to see additional testing, such as Spearman-Brown prophecy formula, to discover whether a larger sample or more items are needed to better distinguish between facility variation. They also requested ICC coefficients to better assess within versus between program comparisons.
- None of the measures passed the reliability criteria at the in-person meeting, but the Committee offered some additional feedback to the developers to assist them in continuing to refine the measures.
- During the comment period, the developer completed additional testing and submitted more information. They reanalyzed the data with a larger sample that includes proxy respondents which improves both the gap and the reliability scores.
- The measure was changed to use of top-box scoring instead of mean scores, which is more consistent with CAHPS and which also improved the reliability. They also provided data from the Spearman-Brown prophecy formula, the inter-class correlations, and the factor analysis, as well as more information on the cognitive testing of the survey the measure uses.
- The developer also clarified several outstanding questions, including the unit of analysis (an HCBS program) and accountable entity (the operating entity), and summarized the minor changes made to comply with CAHPS standardized requirements as the survey the measure is based on is now a CAHPS survey.
- The Committee requested clarification on whether the CAHPS designation included a review of the psychometric properties of the instrument and of the measures derived from the instrument. It was clarified that the CAHPS Consortium only evaluates at the instrument level, but would have found the reliability and validity of the tool acceptable to carry the CAHPS trademark.
- There continued to be some concerns about the reliability of some of the individual measures and clarification was sought regarding a threshold NQF would consider minimally acceptable. It was explained that NQF does not set threshold standards, but relies on the expertise of the respective Standing Committees to determine what satisfies the respective criteria. During the post-comment call, the Committee discussed the new information submitted as well as the comments received. Voting was conducted on a post-call voting survey, and the Committee voted to recommend the measure.

3. Feasibility: H-4; M-10; L-2; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)

Rationale:

- During the feedback portion of the discussion, the Committee requested more information on the feasibility of getting the optimal sample size of 400.

- The Committee also requested information on how long the survey takes to complete and the burden on individual patients/caregivers.
- During the comment period the developer updated the Feasibility information, noting that the inclusion of proxy respondents significantly improves the response rate; they also noted improvements made to survey administration that improve rates.
- The developer estimates 30 minutes is needed to complete the survey, as compared to 20 minutes for Nursing Home CAHPS. They noted that the average response includes 51 out of 96 items (due to skip patterns) which would indicate an expected response time of around 13 minutes.

4. Usability and Use: H-5; M-11; L-0; I-0

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

Rationale:

- During the Importance section, the Committee did discuss the intended use of this set of measures and wanted to know if this would be publically reported. Given that for some patients, the only way to receive improved care would be to move to a different state with a better program, Committee members questioned how public reporting could be useful. The developer reiterated that the measures are still voluntary and that states could decide how to use it or report on it. Round 1 data were not reported publically, but was given to the states in individual reports, and the states wanted to keep the results internal.

5. Related and Competing Measures

- No related or competing measures noted.

Standing Committee Recommendation for Endorsement: **Y-15**; **N-1**

- The Committee agreed that it is really important that we capture the experience of patients who are provided with care by Home and Community Based services and so we need to make sure that the measures we use reflect that care appropriately. They noted that the reliability, feasibility, and usability should be monitored as the measure is implemented to ensure the measure remains psychometrically sound.

6. Public and Member Comment

- During the public and NQF Member comment period, the developer submitted additional information responding to the Committee's concerns.
- This measure received 11 comments, which primarily focused on requesting reconsideration of the measure. However, the comments reflected possibility of lack of understanding of the NQF policy of not endorsing the actual survey and the process for consideration of measures. The comments support the importance of experience of care measures for the HCBS community, which the Committee had also previously supported.
- During the post-comment call, the Committee discussed the new information submitted as well as the comments received. During voting conducted on a post-call voting survey, the Committee voted to recommend all 19 measures included in the submission for endorsement. At this second vote, the Committee voted on the measure in one batch rather than splitting it out into domains.

7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X

8. Board of Directors Vote: Y-X; N-X

9. Appeals