NATIONAL QUALITY FORUM
Driving measurable health improvements together

Meeting Summary

Patient Experience and Function Standing Committee – Measure Evaluation Web Meeting

The National Quality Forum (NQF) convened the Patient Experience and Function Standing Committee for two separate web meetings on February 23 and 28, 2023, to evaluate five measures for the fall 2022 cycle.

Welcome, Review of Meeting Objectives, Introductions, and Overview of Evaluation and Voting Process

On February 23, Leah Chambers, NQF director, welcomed the Standing Committee and participants to the web meeting. After the co-chairs provided welcoming remarks, Dr. Matthew Pickering, NQF managing director, informed the Standing Committee that the Centers for Medicare & Medicaid Services' (CMS) contract to serve as the consensus-based entity is set to end on March 26 of this year. CMS recently completed a competitive process to award the next phase of work and announced its award decision: NQF was not awarded the contract, so its work will conclude on March 26, 2023. Dr. Pickering further mentioned that NQF will work with CMS and the successor contractor in the weeks ahead to make a smooth transition, which will include further communication with this Standing Committee and other NQF Committee volunteers. However, Dr. Pickering underscored that this does not change the Standing Committee's focus for the measure evaluation meeting, and NQF looks forward to working with the Committee to review the fall 2022 measures.

NQF staff reviewed the meeting objectives. Following the review, the Standing Committee members each introduced themselves and disclosed any conflicts of interest. There were no conflicts of interest or recusals. Additionally, Erin Buchanan, NQF senior manager, reviewed the Consensus Development Process (CDP) and the measure evaluation criteria.

A quorum of 12 active Standing Committee members was met and maintained during the review of NQF #2958 and NQF #2962. However, quorum was lost during the discussion of NQF #3720 on February 23. Therefore, the Standing Committee discussed all the remaining criteria for NQF #3720 and NQF #3718 and voted after the meeting using an online voting tool. NQF #3734 was discussed on February 28, during which quorum was not achieved, and voting took place after the meeting using an online voting tool. Voting results are provided below.

Measure Evaluation

During the meetings, the Patient Experience and Function Standing Committee evaluated five measures (two maintenance and three new) for endorsement consideration. Prior to the review of the measures, Dr. Pickering noted that for the fall 2022 cycle, measures were reviewed by the Scientific Methods Panel (SMP) if they were deemed as complex (i.e., outcome, cost, composite, and instrument-based measures) and/or if they included testing methods that are not commonly used. For the Patient Experience and Function measures under review, four of the five measures (i.e., NQF #2958, NQF #2962, NQF #3718, and NQF #3720) were evaluated by the SMP.

A measure is recommended for endorsement by the Standing Committee when greater than 60 percent of eligible voting members select a passing vote option (i.e., Pass, High and Moderate, or Yes) on all must-pass criteria and overall suitability for endorsement. A measure is not recommended for endorsement when less than 40 percent of voting members select a passing vote option on any must-pass criterion or overall suitability for endorsement. If a measure does not pass on a must-pass criterion, voting during the measure evaluation meeting will cease. The Standing Committee will not re-vote on the measures during the post-comment meeting unless the Standing Committee decides to reconsider the measure(s) based on submitted comments or a formal reconsideration request from the developer. The Standing Committee has not reached consensus on the measure if between 40 and 60 percent of eligible voting members select a passing vote option on any must-pass criterion or overall suitability for endorsement. The Standing Committee will re-vote on criteria for which it did not reach consensus and potentially on overall suitability for endorsement during the post-comment web meeting.

Voting Legend:

- *Evidence (Outcome Measures) and Use: Pass/No Pass*
- *Accepting the SMP Rating and Overall Suitability for Endorsement: Yes/No*
- *All Other Criterion: H – High; M – Moderate; L – Low; I – Insufficient; NA – Not Applicable*
- *Maintenance Criteria for Which the Standing Committee Decided Additional Discussion/Vote Was Not Needed (Evidence, Reliability, Validity only): Accepted Previous Evaluation*

NQF #2958 Informed, Patient-Centered (IPC) Hip and Knee Replacement Surgery (Massachusetts General Hospital): Maintenance

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; **Measure Type**: Outcome: PRO-PM; **Level of Analysis**: Clinician: Group/Practice; **Setting of Care**: Ambulatory Care, Outpatient Services; **Data Source**: Instrument-Based Data

Measure Steward/Developer Representatives at the Meeting

- Karen Sepucha, Mass General Hospital
- KD Valentine, Mass General Hospital

Standing Committee Votes

- **Evidence**: Total Votes – 13; Pass-10; No Pass-3 (10/13 – 76.9%, Pass)

- **Performance Gap**: Total Votes – 13; H-0; M-9; L-3; I-1 (9/13 – 69.2%, Pass)

- **Reliability**: This measure was evaluated by the SMP and was rated as high for reliability. The Standing Committee accepted the SMP's rating for reliability: Total Votes – 13; Yes-11; No-2; (11/13 – 84.6%, Pass)

- **Validity**: This measure was evaluated by the SMP and was rated as moderate for validity. The Standing Committee accepted the SMP's rating for validity: Total Votes – 13; Yes-10; No-3 (10/13 – 76.9%, Pass)

- **Feasibility**: Total Votes – 13; H-0; M-9; L-3; I-1 (9/13 – 69.2%, Pass)

- **Use**: Total Votes – 13; Pass-10; No Pass-3 (10/13 – 76.9%, Pass)

- **Usability**: Total Votes – 13; H-0; M-9; L-1; I-3 (9/13 – 69.2%, Pass)
- **Standing Committee Recommendation for Endorsement:** Total Votes – 13; Yes-10; No-3 (10/13 – 76.9%, Pass)

The Standing Committee recommended the measure for continued endorsement.

This clinician group-level measure was originally endorsed in 2017. It is not used in any public-reporting programs, and there are no plans to do so. However, it is currently being tested in the following accountability programs: the Blue Cross Blue Shield (BCBS) of Massachusetts Alternative Quality Contract, the Alliance Quality Path Program, and the Shared Decision-Making Program at the Massachusetts General Brigham Health System. In reviewing the evidence for the measure, the Standing Committee acknowledged that this measure is a maintenance patient-reported outcome performance measure (PRO-PM) that is derived from patient responses to the Hip or Knee Decision Quality Instruments. It assesses the proportion of participants who have a passing knowledge score (i.e., 60 percent or higher) and a clear preference for surgery. These participants have met the criteria for an informed, patient-centered (IPC) decision. The target population is adult patients who underwent a primary hip or knee replacement surgery to treat hip or knee osteoarthritis. Since the last review of this measure, a systematic review was conducted, which found that IPC decisions are associated with higher shared decision-making scores. In addition, a cross-sectional survey conducted at four hospitals affiliated with a large health system found that IPC decisions were associated with better physical health and physical function outcomes. A cluster randomized trial of decision support also found that IPC decisions predicted better outcomes following knee replacement surgery.

A Standing Committee member asked the developer to clarify the purpose of the measure because it appears to be more of a psychological measure than a physical improvement measure related to surgery. Another Standing Committee member stated that there is a difference between informed decision making and shared decision making. Specifically, there is no question as to whether this is an important measure in practice, but there is a question as to whether it rises to the importance of a national quality measure. The developer replied that this is a patient-reported outcome (PRO) that assesses the quality of the patient's decision, which is an outcome in and of itself. When patients make more informed decisions, they tend to have better outcomes. The Standing Committee did not have any additional comments and proceeded to pass the measure on evidence.

Moving to performance gap, the Standing Committee acknowledged that the developer reported on data from three new data sets with IPC rates ranging from 70 to 92 percent. With respect to disparities, several differences by age, gender, and race were present across the various samples. The Standing Committee discussed whether a gap existed for other factors, particularly smoking status. The developer explained that those analyses had not yet occurred but would be considered in the future. A Standing Committee member also noted that certain racial groups have a lower rate of undergoing these surgeries. Another Standing Committee member questioned whether the gap had changed over time. The developer replied that they did not have data on changes over time. However, the Standing Committee recognized that a gap still remains and that disparities are present in care across various patient groups. Therefore, the Standing Committee passed the measure on performance gap.

The SMP passed the measure on reliability with a rating of high (H-6; M-2; L-0; I-1). The Standing Committee noted that the reliability testing was conducted at the accountable-entity level. For four testing sites, the reliability was 0.735. Reliability testing was also conducted at the patient/encounter level for the instrument. The Standing Committee recognized that in the 2016 submission, the developer conducted test-retest reliability of the knowledge and preference items from the same individuals four to six weeks apart. The test-retest reliability of the knowledge score was examined in the first sample

with an intraclass correlation coefficient (ICC) of 0.81 (95 percent confidence interval [CI] ranging from 0.71 to 0.87). The test-retest reliability of the item assessing preferred treatment had a Kappa value of 0.801. Moving to a vote, the Standing Committee accepted the SMP's rating of high and passed the measure on reliability.

The SMP passed the measure on validity with a rating of moderate (H-4; M-4; L-1; I-0). The Standing Committee noted that the validity testing was conducted at the accountable-entity level. The Standing Committee also discussed the intent of the measure, which is to evaluate patient knowledge and preferences. One Standing Committee member noted that it is difficult to tease apart knowledge and preferences. In addition, the survey was given after the surgery. Therefore, the survey results came too late and may not be actionable for providers. There was also concern about the face validity of the measure because it had not been brought to a general sample of orthopedic physicians. The developer explained that the measures have been used in a variety of settings and agreed that the ideal timing is preoperative rather than postoperative. However, the challenge is to operationalize the measure in terms of creating a way to reliably test the measure across systems. The goal of combining preoperative and postoperative results was to assess whether there was an informed preference, which is an ultimate measure of decision quality. Lastly, the Standing Committee questioned whether risk adjustment should be used. The developer stated that the measure is not risk-adjusted, as everyone should have the same level of communication and knowledge, whether they have zero or multiple comorbidities. The Standing Committee accepted the developer's responses and voted to accept the SMP's rating for validity.

Regarding feasibility, the patient-reported surveys can be administered online to support electronic capture. If administered via mail or paper, then it will require staff resources. The administration of these questions has been conducted across multiple sites and in multiple modes (predominantly paper and online surveys). A large health system has incorporated the items into their PROs registry for orthopedics, and the data are being collected as part of routine care in that system. The developer stated that patients generally find these surveys acceptable, which was indicated by the good response rates and a low number of missing data. However, the Standing Committee had concerns about implementation costs and response rates. The developer explained that the response rate is approximately 43 percent and implementation costs are dependent upon the setting. A question was raised regarding translating the survey to languages other than English and Spanish. The developer did not have specific plans to do this but was happy to do so based on the demand. The Standing Committee did not have any further questions and passed the measure on feasibility.

Next, the Standing Committee discussed the use criterion. It acknowledged that the measure is being piloted for use in the BCBS of Massachusetts Alternative Quality Contract, the Alliance Quality Path Program, and the Shared Decision-Making Program at the Massachusetts General Brigham Health System. The Alliance Quality Path Program specifies the measurement of decision quality and shared decision making as part of its criteria for recognition. The developer stated that NQF #2958 can be used for this recognition. Additionally, the Shared Decision-Making Program at the Massachusetts General Brigham Health System incorporates items of the IPC measure into the PROs registry. A Standing Committee member questioned whether it is only used in one region. The developer stated that at least one large system has chosen to use the measure in BCBS. The Standing Committee did not raise any additional questions and passed the measure on use.

Regarding usability, the developer did not provide trend data but noted that studies using the IPC measure have found that patients who are provided with decision support interventions have significantly higher rates of knee replacement surgery compared to usual care. The developer also stated that the scores at the Massachusetts General Brigham Health System are shared at the site and at the physician level. The developer is also interested in looking at trends over time but has not done so due to the disruptions related to COVID-19. A Standing Committee member questioned how the

measure has impacted the actual surgeons being held accountable to the measure. The developer explained that this is a group-level measure and groups can drill down to the physician level to obtain feedback. The Standing Committee did not have any further questions and passed the measure on usability. The Standing Committee recommended the measure for overall suitability for endorsement.

No related or competing measures were identified for this measure, and no NQF member or public comments were received prior to the measure evaluation meeting.

NQF #2962 Shared Decision-Making Process (Massachusetts General Hospital): Maintenance

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; **Measure Type**: Outcome: PRO-PM; **Level of Analysis**: Clinician: Group/Practice; **Setting of Care**: Ambulatory Care, Inpatient/Hospital, Outpatient Services; **Data Source**: Instrument-Based Data

Measure Steward/Developer Representatives at the Meeting

- Karen Sepucha, Mass General Hospital
- KD Valentine, Mass General Hospital

Standing Committee Votes

- **Evidence**: Total Votes – 12; Pass-11; No Pass-1 (11/12 – 91.7%, Pass)
- **Performance Gap**: Total Votes – 12; H-3; M-7; L-1; I-1 (10/12 – 83.3%, Pass)
- **Reliability**: This measure was evaluated by the SMP and was rated as moderate for reliability. The Standing Committee accepted the SMP's rating for reliability: (Total Votes – 12; Yes-10; No-2 (10/12 – 83.3%, Pass)
- **Validity**: This measure was evaluated by the SMP and was rated as moderate for validity. The Standing Committee accepted the SMP's rating for validity: (Total Votes – 12; Yes-8; No-4 (8/12 – 66.7%, Pass)
- **Feasibility**: Total Votes – 13; H-0; M-10; L-1; I-2 (10/13 – 76.9%, Pass)
- **Use**: Total Votes – 14; Pass-12; No Pass-2 (12/14 – 85.7%, Pass)
- **Usability**: Total Votes – 13; H-0; M-9; L-1; I-3 (9/13 – 69.2%, Pass)
- **Standing Committee Recommendation for Endorsement**: Total Votes – 13; Yes-10; No-3 (10/13 – 76.9%, Pass)

The Standing Committee recommended the measure for continued endorsement.

This clinician group-level measure was originally endorsed in 2016. It is not used in any public-reporting programs, and there are no plans to do so. However, it is currently used in the following accountability programs: the BCBS of Massachusetts Alternative Quality Contract, the Alliance Quality Path Program, and the Shared Decision-Making Program at the Massachusetts General Brigham Health System.

In reviewing the evidence, the Standing Committee noted that the measure is a PRO-PM that assesses shared decision making. The Standing Committee noted that since the last endorsement of the measure, a 2021 meta-analysis of the shared decision making (SDM) process scale was conducted for surgical

decisions in which researchers found that SDM process scores were associated with higher decision quality, less decisional conflict, and lower decision regret. The developer also provided a study on hip and knee replacement and spine surgery decisions in which researchers found that SDM process scores were related to less regret and higher patient satisfaction. A Standing Committee member asked whether the measure was a composite. The developer clarified that the measure is not a composite. The Standing Committee had no additional questions and passed the measure on evidence.

The Standing Committee pointed out the large differences in performance gap. A Standing Committee member asked the developer whether they reviewed income disparities. Another Standing Committee member noted that many of the procedures were moved to different settings of care, such as ambulatory surgical centers. The developer noted that they were unable to find detailed information on race and ethnicity. The developer further stated that income data were not available, and because this measure was tested in Massachusetts, almost all (98 percent) of the data were insured, making the data less meaningful to test by insurance status. The Standing Committee had no other concerns and passed the measure on performance gap.

The SMP passed the measure on reliability with a rating of high (H-0; M-8; L-0; I-2). The Standing Committee noted that reliability testing was conducted at the accountable-entity level. The average signal-to-noise estimate for reliability was 0.69 (95 percent CI = [0.685, 0.69]). The developers also reported an ICC of 0.96. For reliability testing conducted at the patient/encounter level in the 2016 submission, the developer noted that Cronbach alpha may not be an appropriate measure of reliability due to the nature of the measure. However, the developer calculated the alphas for some decisions, noting that they were in the 0.5–0.7 range. The short-term, test-retest data on some variations of the measure obtained ICC values ranging from 0.7 to 0.8. Several Standing Committee members had concerns with the data provided regarding reliability. In addition, the Standing Committee was concerned that the accountable entity-level testing did not demonstrate adequate reliability for all surgical procedures. The developer clarified that there were more data in some areas than others and provided the videotape recording to strengthen the reliability of the measure. The Standing Committee had no further questions, accepted the SMP's rating of moderate, and passed the measure on reliability.

The SMP passed the measure on validity with a rating of moderate (H-3; M-4; L-1; I-2). The developer provided evidence from three published studies, which showed correlations, in the predicted direction, with other decision-making outcomes (e.g., higher confidence; higher satisfaction; less regret; and higher rates of informed, patient-centered surgery). Validity testing was also conducted at the accountable-entity level, citing two studies at the practice-site level. The developers tested whether clinical practices that implemented shared decision making had higher SDM scores than sites practicing usual care. For osteoarthritis of the knee and hip, patients in practices where decision aids were used reported significantly better decision processes (2.9 versus 2.5, $p < 0.001$ and 2.9 versus 2.1, $p < 0.001$, respectively). The difference in the SDM process scores for spine practices that did and did not use decision support (3.0 versus 2.75, $p=0.12$) was in the expected direction but was not large enough to reach statistical significance. Regarding breast cancer practices, those that had formal decision support had significantly better scores than cancer practices without any decision support interventions (2.7 versus 2.3, $p < 0.05$). The Standing Committee raised some concerns regarding the lack of risk adjustment. However, one Standing Committee member agreed with the choice to not risk-adjust the measure, stating that shared decision making is a fundamental process and that any differences should not be risk-adjusted. Additionally, the developer further clarified that risk adjustment is not needed because disparities were not being observed. They explained that this is a fundamental process that should be performed but should not be influenced by patient comorbidities. The Standing Committee accepted the developer's explanations and the SMP's rating of moderate and passed the measure on validity.

For feasibility, the developer reported that data for this measure are generated online or via paper. Patient surveys administered via paper are required to be entered into an online database by staff for reporting. The only difficulty regarding the data collection was obtaining sufficient response rates due to the administrative burden on clinic staff. The Standing Committee raised a concern about whether people who do not speak English or Spanish or those with developmental disabilities are able to respond to the survey. The Standing Committee also had concerns about the cost of the infrastructure to administer surveys by practice and whether proxy surveys could be used in patients with cognitive impairment. The developer replied that data were collected from older adults who showed mild or moderate cognitive impairments to see how the surveys performed and hope to have more data to review in the future; however, nothing had become definitive yet. The developer clarified that other entities (e.g., insurance companies) could administer the survey, which would reduce the financial and measurement burden on practices. Based on the discussion, the Standing Committee passed the measure on feasibility.

Regarding use, the Alliance Quality Path Program, similar to NQF #2958, specifies the measurement of decision quality and shared decision making as part of its criteria for recognition. The developer noted that NQF #2962 can be used for this recognition. The Shared Decision-Making Program at the Massachusetts General Brigham Health System incorporates items from the SDM measure into the Patient-Reported Outcomes registry. The Standing Committee did not raise any questions and passed the measure on use.

Regarding usability, the developer did not provide trend data; however, they noted that higher scores were associated with less decisional conflict and less decision regret reported by patients; they also noted that the scores improved after the introduction of formal decision support programs. The Standing Committee had questions and concerns with regard to usability, specifically regarding when to ask the survey questions and what practices can do with low scores. The Standing Committee also raised questions about the temporality of receiving the scores. The developer stated that there was an average score for a practice and survey questions would ideally be asked right after the surgery to reduce recall bias. The Standing Committee did not raise any additional concerns and passed the measure on usability.

There was no additional discussion on the measure; therefore, the Standing Committee voted to recommend the measure for continued endorsement. No NQF member or public comments were received prior to the measure evaluation meeting.

Two related measures were discussed. The first was NQF #0005 *CAHPS Clinician & Group Surveys (CG-CAHPS) Version 3.0 – Adult, Child*. The developer stated that this measure has an optional supplement of the shared decision-making items that are adaptations of the items used in NQF #2962. The developer explained that the problem with integrating NQF #2962 into NQF #0005 pertained to both the sample sizes required and sampling strategies. NQF #2962 refers to more specific procedures. In addition, the developer noted that NQF #2962 focuses more on detail and has a larger sample size to assess a focused group of procedures. The second related measure was NQF #3227 *CollaboRATE Shared Decision-Making Score.* This measure assesses the perceptions of shared decision making but does not target specific decisions or concrete behaviors during the clinical encounter. The developer noted that shared decision-making quality can vary by type. Therefore, it was important to collect data on specific procedures and medications. A Standing Committee member stated that the key data element is assessing the time spent with the doctor. However, none of the measures capture this. The Standing Committee member also noted that this key data element could be integrated into future iterations of the measure. Another Standing Committee member again mentioned the heterogeneity of NQF #2962 and how it might work operationally when there are so many different types of procedures assessed. Ultimately, the Standing

Committee did not raise any major concerns with the developer's rationale regarding the differences with the related measures.

NQF #3720 Patient-Reported Fatigue Following Chemotherapy Among Adults With Breast Cancer (Purchaser Business Group on Health): New

Description: The PRO-PM assesses fatigue among adult women with breast cancer entering survivorship after completion of chemotherapy administered with curative intent. Fatigue is assessed using the PROMIS Fatigue 4a scale administered at baseline (prior to chemotherapy) and at follow-up (about three months following completion of chemotherapy); **Measure Type**: Outcome: PRO-PM; **Level of Analysis**: Clinician: Group/Practice; **Setting of Care**: Ambulatory Care, Outpatient Services; **Data Source**: Instrument-Based Data, Paper Medical Records, Electronic Health Records

Measure Steward/Developer Representatives at the Meeting

- Rachel Brodie, Purchaser Business Group on Health
- Feifei Ye, Purchaser Business Group on Health
- Kristen McNiff Landrum - Purchaser Business Group on Health
- Jennifer Grigg, University of Michigan

Standing Committee Votes

- **Evidence**: Total Votes – 12; Pass-9; No Pass-3 (9/12 – 75.0%, Pass)
- **Performance Gap**: Total Votes – 12; H-0; M-10; L-0; I-2 (10/12 – 83.3%, Pass)
- **Reliability**: This measure was evaluated by the SMP and was rated as moderate for reliability. The Standing Committee accepted the SMP's rating for reliability: (Total Votes – 12; Yes-10; No-2 (10/12 – 83.3%, Pass)
- **Validity**: This measure was evaluated by the SMP and was rated as "consensus not reached" for validity. The Standing Committee votes are listed below:
    - Total Votes – 12; H-0; M-8; L-1; I-3 (8/12 – 66.7%, Pass)
- **Feasibility**: Total Votes – 12; H-0; M-8; L-3; I-1 (8/12 – 66.7%, Pass)
- **Use**: Total Votes – 12; Pass-8; No Pass-4 (8/12 – 66.7%, Pass)
- **Usability**: Total Votes – 12; H-1; M-5; L-3; I-3 (6/12 – 50.0%, Consensus Not Reached)
- **Standing Committee Recommendation for Endorsement:** Total Votes – 12; Pass-8; No Pass-4 (8/12 – 66.7%, Pass)

The Standing Committee recommended the measure for endorsement.

This clinician group-level measure was newly submitted for endorsement. Because this is a new measure, it is not currently in any federal program. The developer stated that the measure will be submitted to CMS for consideration for the Measures Under Consideration (MUC) list.

For evidence, the Standing Committee noted that this measure is a new PRO-PM at the group/clinician level that measures fatigue following chemotherapy for adult patients with breast cancer. It is based on the Patient-Reported Outcomes in Oncology (PROMOnc) premise that medical oncologists who provide

the highest quality of care, specifically medical and nonmedical support to patients with curative-intent cytotoxic therapy, will be able to reduce symptom burden and therefore improve the patient's transition into the cancer survivorship period. The Standing Committee noted that that there are randomized trials showing a clear link between process and the PRO-PM. During the pre-evaluation review, a Standing Committee member commented that the evidence was low in the National Comprehensive Cancer Network (NCCN) guidelines regarding the impact of interventions on fatigue. A Standing Committee member asked about how fatigue should be tracked in different stages of cancer treatment. Another question was asked about whether males should be included because men can be treated for breast cancer in rare cases. The developer replied that the practices that perform well are the ones that address fatigue while treatment is ongoing. Regarding males with breast cancer, the developer explained that only one male was in the sample, and therefore, the numbers were not high enough to include males in the measure. Another question was asked about the treatability of fatigue. The developer explained that there are meaningful interventions to evaluate the intensity of fatigue, nutritional imbalances, and anemia and to perform interventions such as diet, exercise, and medication review. During this discussion, the Standing Committee lost quorum for this measure. Therefore, voting for the evidence criterion and the remaining criteria was conducted offline after the meeting. During offline voting, the Standing Committee passed the measure on evidence.

Regarding performance gap, the Standing Committee noted that the developer found an average adjusted measure score of 48.51, with a range from 42.13 to 53.07 and standard deviation (SD) of 3.13. The Standing Committee also noted that in the measure submission, the developer stated that the measure was adjusted for race and ethnicity, marital status, and insurance status; however, none of these variables were significant in the relationship with the measure. The developer reported the presence of outliers and a meaningful distribution in the scores. During offline voting, the Standing Committee passed the measure on performance gap.

The SMP passed the measure on reliability with a rating of moderate (H-0; M-9; L-1; I-0). The Standing Committee noted that the mean estimate of the reliability for the measure score was 0.77, with a range from 0.38 to 0.88 across clinician group practices. The Standing Committee had no questions or concerns and accepted the SMP's rating on reliability during offline voting.

The Standing Committee noted that the SMP did not reach consensus on validity (H-0; M-6; L-3; I-1) due to concerns with the face validity testing, the lack of meaningful differences in performance, and concerns with missing response rates. During the discussion, the Standing Committee asked for clarification on how much an improvement in fatigue is really about the quality of care in this population. A comment was also made about understanding what sorts of interventions had been done and what had been effective to reduce fatigue. The developer explained that practices that performed well on the measure assessed and intervened on fatigue early. The developer also noted that they risk-adjusted for baseline fatigue, which would be a proxy for prior interventions. Stage IV cancer patients were excluded because the goal was to address patients with curative cancer, considering there are measures for Stage IV patients that are separate. With respect to the face validity testing, the Standing Committee discussed the SMP's concern regarding four of the 12 oncologists, noting that the fatigue may not be due to cancer but to fatigue related to the COVID-19 pandemic. The developer stated that COVID-19 may have been a confounding factor. However, the guidelines recommend assessing stressors like the COVID-19 pandemic. The developer also stated that meaningful differences were present between the test sites, which were tested in part during the COVID-19 pandemic. The Standing Committee did not have any further questions passed the measure on validity during offline voting.

Regarding feasibility, the Standing Committee noted that measures from the Patient-Reported Outcomes Measurement Information System (PROMIS) originate from a survey that must be collected

by staff, then entered into the electronic health record (EHR) in structured fields. Some EHRs (e.g., Epic and Cerner) now include PROMIS surveys. The developer stated that collecting the baseline survey with the originally defined time frame from patients taking oral chemotherapy was challenging. According to the feedback, 75 percent of respondents reported that it took them less than 10 minutes to complete the PROMOnc survey; 92 percent reported that they understood the survey instructions; 83 percent reported that they did not have any technical issues completing the survey; and 83 percent felt that the time that it took to complete the survey was reasonable. One of the concerns mentioned was having staff enter the data. The developer stated that many of the test sites programmed the survey into the EHR or on an iPad. Some of the sites programmed it into the Research Electronic Data Capture (RedCAP). The Michigan practices also had a centralized vendor. Health plans could also potentially collect the data. Regarding clinical trials, patients were excluded because including them could potentially alter the data for the measure. During offline voting, the Standing Committee passed the measure on feasibility.

Regarding use, the Standing Committee stated that the measure will be submitted to CMS for consideration for the MUC list for potential use in a CMS Quality Payment Program. A Standing Committee member further stated that the measure information could also be given to the patient. The developer stated that the measure could be used for quality improvement. Regarding how to provide the data to the patient, the developer stated that PRO-PM data are being provided to patients. During offline voting, the Standing Committee passed the measure on use.

During the discussion on usability, the Standing Committee noted that there were no unexpected findings. However, concerns were raised about potential harm or unintended consequences if the measure was used for accountability, and many Standing Committee members expressed that the measure would be more appropriate for use as a quality improvement measure. The general concern was that it may incentivize a pharmacological solution. A Standing Committee member expressed that they would not like to see this measure in a value-based payment program. The developer explained that the measure was developed iteratively with the technical expert panel (TEP) to facilitate its usability. There was no additional discussion on usability. During offline voting, the Standing Committee did not reach consensus on usability, which is not a must-pass criterion. This was primarily due to the concern regarding the potential unintended consequences of the measure's use in public programs versus quality improvement.

The Standing Committee had no further points to discuss regarding this measure and concluded the discussion. During offline voting, the Standing Committee recommended the measure for overall suitability for endorsement. No NQF member or public comments were received prior to the measure evaluation meeting. The Standing Committee was unable to discuss related and competing measures during the meeting; therefore, this discussion will occur during the post-comment meeting.

NQF #3718 Patient-Reported Pain Interference Following Chemotherapy Among Adults With Breast Cancer (Purchaser Business Group on Health): New

Description: The PRO-PM assesses pain interference among adult women with breast cancer entering survivorship after completion of chemotherapy administered with curative intent. Pain interference is assessed using the PROMIS Pain Interference 4a scale administered at baseline (prior to chemotherapy) and at follow-up (about three months following completion of chemotherapy); **Measure Type**: Outcome: PRO-PM; **Level of Analysis**: Clinician: Group/Practice; **Setting of Care**: Ambulatory Care, Outpatient Services; **Data Source**: Instrument-Based Data, Paper Medical Records, Electronic Health Records

Measure Steward/Developer Representatives at the Meeting

- Rachel Brodie, Purchaser Business Group on Health
- Feifei Ye, Purchaser Business Group on Health
- Kristen McNiff Landrum - Purchaser Business Group on Health
- Jennifer Grigg – University of Michigan

Standing Committee Votes

- **Evidence**: Total Votes – 12; Pass-10; No Pass-2 (10/12 – 83.3%, Pass)
- **Performance Gap**: Total Votes – 12; H-1; M-7; L-1; I-3 (8/12 – 66.7%, Pass)
- **Reliability**: This measure was evaluated by the SMP. The Standing Committee accepted the SMP's moderate rating for reliability: Total Votes – 12; Pass-10; No Pass-2 (10/12 – 83.3%, Pass)
- **Validity**: This measure was evaluated by the SMP. The Standing Committee accepted the SMP's moderate rating for validity: Total Votes – 12; Pass-10; No Pass-2 (10/12 – 83.3%, Pass)
- **Feasibility**: Total Votes – 12; H-0; M-10; L-1; I-1 (10/12 – 83.3%, Pass)
- **Use**: Total Votes – 12; Pass-10; No Pass-2 (10/12– 83.3%, Pass)
- **Usability**: Total Votes – 12; H-1; M-9; L-0; I-2 (10/12 – 83.3%, Pass)
- **Standing Committee Recommendation for Endorsement:** Total Votes – 12; Pass-10; No Pass-2 (10/12– 83.3%, Pass)

The Standing Committee recommended the measure for endorsement.

This clinician group-level measure was newly submitted for endorsement. It is not currently in any quality or accountability programs. The developer stated that the measure will be submitted to CMS for consideration for the MUC list for potential use in a Quality Payment Program.

The measure assesses patient-reported pain interference following chemotherapy for adult patients with breast cancer. Similar to NQF #3720, it is based on the PROMOnc premise that medical oncologists who provide medical and nonmedical support to patients with curative-intent cytotoxic therapy will be able to reduce symptom burden and therefore improve the patient's transition into the cancer survivorship period. In their submission, the developer provided a logic model that depicts that patients who are undergoing chemotherapy with curative intent experience persistent symptoms following treatment, such as pain, fatigue, and other issues impacting health-related quality of life. The developer attested that if evidence-based practices are delivered by the group practice and clinician, patients will experience lower symptom burden during the survivorship period. The developer referenced the 2022 NCCN Adult Cancer Pain Guideline and 2022 NCCN Survivorship Guideline recommendations to demonstrate relationships between the PRO-PM and healthcare actions that can be utilized to achieve the desired outcome. The Standing Committee noted a clear process-outcome link in terms of ways that clinicians can impact this measure. A Standing Committee member expressed that they had a favorable view of the measure, particularly with respect to addressing pain through both pharmacologic and non-pharmacologic methods. However, another Standing Committee member mentioned that the evidence was graded as a low-quality level of evidence for the interventions. The Standing Committee did not have any additional discussion on this concern, nor did it have any additional discussion on evidence. During offline voting, the Standing Committee passed the measure on evidence.

Regarding performance gap, ten clinician groups participated in the field testing of this measure. The developer stated that the number of surveys used in the analysis decreased from 744 to 323 because 323 participants completed both pre- and post-surveys. Therefore, a total of 323 follow-up surveys were used for analysis. The developer expected the completion rate to be higher, given it is an accountability measure, particularly as these surveys become more integrated into clinical care. The average adjusted measure score was 50.51. The range was 43.92 to 54.11 with an SD of 2.83, which was considered clinically significant by the developer. In terms of disparities, no socioeconomic status (SES), race, or insurance variables were deemed significant in their relationship with the measure following adjustment for multiple comparisons. The developer noted that research studies have found that certain groups of survivors, such as racial and ethnic minorities and those of lower socioeconomic status, report poorer PROs and interventions to address those outcomes. A Standing Committee member asked about the ethnicity distribution, particularly ensuring equal proportions of race and ethnicity in the sample, given the potential differences in pain perception. Another Standing Committee member stated that this is not the case regarding perception of pain; rather, it is more about differences in treatment by racial disparities. Another question was asked about whether test sites were representative of different types of practices. The developer stated that a wide variety of practices exists across several states. The Standing Committee did have any further discussion and passed the measure on performance gap during offline voting.

The SMP passed the measure on reliability (H-0; M-9; L-1; I-0). Reliability testing was conducted at the patient/encounter and accountable-entity levels. The developer noted that PROMIS measures, including the pain interference scale, have undergone rigorous development and validation. Reliability testing from the literature demonstrated that Cronbach's alpha is 0.99 for PROMIS Pain Interference. To test the reliability at measure-score level (i.e., accountable-entity level), a signal-to-noise analysis was performed. The estimate of the adjusted ICC was 0.097. The estimate of the reliability at the average sample size for a group (32 patients per group) was 0.77. Using the Spearman-Brown prophecy formula, the developer estimated that to obtain a nominal reliability of 0.7, a minimum sample size of 22 patient respondents would be required. Group-specific reliability ranged from 0.39 to 0.88, with a mean of 0.66 (SD=0.20) and a median reliability of 0.68. The proportion of groups in the sample with sufficient reliability using a reliability threshold of 0.70 was 50 percent. The Standing Committee did not have any concerns about the reliability testing and accepted the SMP's rating for reliability during offline voting.

The SMP passed the measure on validity (H-2; M-5; L-1; I-2). The Standing Committee acknowledged that the developer conducted validity testing at the patient/encounter level. All critical data elements were evaluated using the PROMOnc data registry, and 570 patients were included in the analysis. Percent agreement by data element ranged from 71.6 to 100 percent. Reported Kappa values ranged from 0.64 to 0.67. Reported sensitivity estimates ranged from 33.3 to 89.5 percent, and specificity ranged from 60.0 to 99.5 percent. The developer also assessed face validity using a panel of 12 oncologists. Eight of the 12 participated in the survey; however, four out of the 12 stated that they could not participate due to disruptions with the COVID-19 pandemic. Seven of the 12 agreed or strongly agreed that the pain interference measure could differentiate between good and poor quality of care. The Standing Committee noted that the exclusions for the measure were clear and that the measure was risk-adjusted. Regarding meaningful differences, two of the clinician groups were either above or below the average. A Standing Committee member commented on understanding the upstream management of pain, particularly whether patients are opioid naïve or whether they had chronic pain prior to treatment. The developer stated that they did not want to incentivize the overuse of opioids because they can be unsafe and cause adverse effects. The Standing Committee did not have any additional discussion on validity and accepted the SMP's rating on validity during offline voting.

In terms of feasibility, the PROMIS measures come from a survey that must be collected by staff and entered into the EHR in structured fields. Some EHRs (e.g., Epic and Cerner) now include PROMIS surveys. Based on the clinical expertise and the feasibility assessment from the developer's TEP and from knowledge of the literature in oncology practice trends, the required data were present in the medical record for the majority of the cases. However, collecting the baseline survey with the originally defined time frame was challenging. During the testing period, the developer fielded a questionnaire to assess the burden and feasibility related to data abstraction and implementation and patient-related activities, to which nine sites responded. Most of the implementation burden was associated with administering the survey rather than collecting the clinical and demographic data elements. Patient identification was also a challenge. The developer also surveyed patients to assess their understanding of the survey and ease of use. Twelve patients provided feedback. Seventy-five percent reported that it took them less than 10 minutes to complete the PROMOnc survey; 92 percent reported that they understood the survey instructions; 83 percent reported that they did not have any technical issues completing the survey; and 83 percent felt that the time it took to complete the survey was reasonable. A Standing Committee member stated that because EHRs can be nonstandardized, they would need to be integrated into diverse systems (e.g., Cerner systems tend to differ, while Epic seem to be more standard). The developer explained that it would be feasible in terms of implementing the measure in rural areas due to less resources for measure collection. During offline voting, the Standing Committee passed the measure on feasibility.

Regarding use, the developer stated that as a new measure, the measure will be submitted to CMS for consideration for the MUC list for potential use in a CMS Quality Payment Program. The Standing Committee did not have any concerns or questions and passed the measure on use during offline voting.

In terms of usability, there was nothing to report for improvement because the measure is not currently in use. The Standing Committee did not have any major concerns and passed the measure on usability during offline voting. Notably, this measure, which is similar to NQF #3720, passed on usability, while the Standing Committee did not reach consensus on usability for NQF #3720. This was because some Standing Committee members expressed concern with the measure's use in public programs where the Standing Committee agreed that pain was more addressable than fatigue. Therefore, the measure may be more appropriate in public programs.

During offline voting, the Standing Committee recommended the measure for overall suitability for endorsement. No NQF member or public comments were received prior to the measure evaluation meeting. The Standing Committee was unable to discuss related and competing measures during the meeting; therefore, this discussion will occur during the post-comment meeting.

NQF #3734 Alignment of Person-Centered Service Plan (PCSP) With Functional Assessment Standardized Items (FASI) Needs (CMS/Lewin Group): New

Description: The percentage of home and community-based services (HCBS) recipients aged 18 years or older whose PCSP documentation addresses needs in the areas of self-care, mobility, and instrumental activities of daily living (IADL) as determined by the most recent FASI assessment; **Measure Type**: Process; **Level of Analysis**: Other, population: regional and state; **Setting of Care**: Ambulatory Care, Outpatient Services, Post-Acute Care; **Data Source**: Assessment Data, Instrument-Based Data, Electronic Health Records, Paper Medical Records

Measure Steward/Developer Representatives at the Meeting

- Colleen McKiernan, Lewin Group
- Kenneth Harwood, Lewin Group

Standing Committee Votes

- **Evidence**: Total Votes – 13; H-0; M-10; L-3; I-0 (10/13– 76.9%, Pass)
- **Performance Gap**: Total Votes – 13; H-1; M-9; L-2; I-1 (10/13– 76.9%, Pass)
- **Reliability**: Total Votes – 13; H-0; M-7; L-4; I-2 (7/13– 53.8%, Consensus Not Reached)
- **Validity**: Total Votes – 13; H-0; M-8; L-3; I-2 (8/13– 61.5%, Pass)
- **Feasibility**: Total Votes – 13; H-0; M-7; L-6; I-0 (7/13– 53.8%, Consensus Not Reached)
- **Use**: Total Votes – 13; Pass-10; No Pass-3 (10/13– 76.9%, Pass)
- **Usability**: Total Votes – 13; H-0; M-9; L-3; I-1 (9/13– 69.2%, Pass)
- **Standing Committee Recommendation for Endorsement**: Vote Not Taken

The Standing Committee did not vote on the recommendation for endorsement because it did not reach consensus on reliability—a must-pass criterion. The Standing Committee will re-vote on the measure during the post-comment web meeting.

This is a new measure at the regional-state level. It is used in the Veterans Health Administration (VHA) for its Program of Comprehensive Assistance for Family Caregivers as part of its Veteran Functional Assessment Instrument (VFAI). There are plans to use this measure in CMS' HCBS quality measure set for voluntary adoption by states' HCBS programs.

Regarding evidence, the Standing Committee noted that the developer provided a logic model that shows an individual's person-centered service plan (PCSP). The logic model addressed self-care, mobility, and IADL needs leading to long-term and short-term outcomes. The Standing Committee noted that the developer did not provide a systematic review but did conduct a literature review. A Standing Committee member commented on the importance of these measures to people in Medicaid and in home-based services. One Standing Committee member stated a plan of care does improve outcomes in this population. Another Standing Committee member described the evidence as not particularly strong for people with intellectual disabilities. The developer replied that they had cited publications showing that building services plans are important for people intellectual disabilities. Additionally, the developer clarified that group homes are included in the measure. The Standing Committee did not raise any additional discussion and passed the measure on evidence during offline voting.

Moving to performance gap, the Standing Committee noted that the developer examined data from June and July 2018 from nine organizations across four different states and found performance measure scores ranging from 42.5 percent for individuals with an intellectual or developmental disability to 85.5 percent for individuals with an acquired brain injury. The overall mean of the performance scores was 66.3 percent. Regarding disparities, the Standing Committee noted that the developer found significant differences by race but no significant differences by ethnicity. African Americans had the highest performance measure scores, while those of White race had the lowest performance measure scores. A Standing Committee member asked whether the data were divided by home or group home, to which the developer replied that the data were not analyzed in that way. The Standing Committee also asked why the performance gap may have been higher in the population with intellectual disabilities; however, the developer did not have an explanation for this occurrence. The developer clarified that they also had groups in the sample with an acquired brain injury and older adults with Alzheimer's

disease and the rates were not lower in those groups. There was no further discussion on performance gap. During offline voting, the Standing Committee passed the measure on this criterion.

Regarding reliability, the Standing Committee noted that the SMP did not review this measure and that reliability testing was conducted at the patient/encounter level. In a field test conducted from March to September 2017, reviewers interviewed and observed individuals enrolled in a Medicaid HCBS program in four geographically diverse states. The developer spoke with primary caregivers, or family members, and reviewed case notes from both field tests. The Standing Committee noted that the data elements are clearly defined but that high quality services are subjective and may lead to variation in the reliability of the results. The Standing Committee also asked for clarification on the elements used for the survey score. The developer clarified that the survey looks at three areas: alignment of self-care, mobility, and activities of daily living (ADLs). The Standing Committee asked about the Kappa value of 0.2, which was low, for people with intellectual disabilities. The developer confirmed that the low Kappa value was accurate and that they did not have an explanation as to why it was low. Regarding the three areas listed above, the developer noted that if all the areas of intervention were missing, the survey results would have been removed from the analysis. To be included in the measure, at least one of the areas needed to be captured in the survey. The developer noted that only a small percentage was removed for this reason. During offline voting, the Standing Committee did not reach consensus on the reliability criterion.

As noted above, the SMP did not review this measure. In terms of validity testing, the Standing Committee noted that face validity testing for the measure was conducted by surveying a TEP as well as the reviewers who participated in the reliability testing. The reviewers and TEP members consistently gave the elements a 90 percent “agree” or “strongly agree” rating as to whether the data elements could differentiate quality. The Standing Committee questioned the developer regarding how difficult it is for people with intellectual disabilities to adequately answer the questions. The developer replied that there is a low risk of collecting survey information; in the field test, there were only 36 missing data abstractions forms, which was 7.8 percent of the sample. A Standing Committee member asked whether any patients or family members served on the TEP. The developer confirmed that four participants consisting of patients and family members served on the TEP. Another Standing Committee member asked NQF staff how to vote on validity if there were an issue with a subset of the validity (e.g., the intellectual disabilities population). NQF staff explained that since the number of missing data was minimal, it did not cause any major concern and is reflected in the preliminary rating for validity. Another question was asked about whether the functional assessment standardized items (FASI) measure cognitive features (e.g., depression) rather than functional features. The developer replied that they discussed this subject in depth when they were developing the FASI and decided to focus fully on functional measures. A Standing Committee member mentioned that this is an important area that should be addressed in the future. During offline voting, the Standing Committee passed the measure on validity.

In terms of feasibility, data elements were abstracted from a record by someone other than the person obtaining the original information. A comment was made about whether measures of social determinants of health are captured in this measure. There were also significant concerns with the amount of time and effort required to implement the survey, as well as the large variety of organizations that needed to gather the data. The developer stated that the FASI is available with Logical Observation Identifier Names and Codes (LOINC) codes; therefore, it can be captured using EHR data. The developer also explained that by using standardized templates, they could improve the feasibility of the data collection for sites. This is particularly true for patients with intellectual disabilities, where service planning is even more important. During offline voting, the Standing Committee was unable to reach consensus on feasibility.

Next, the Standing Committee discussed the use criterion. A Standing Committee member noted the measure is in use or is being considered for use in three states. It was being used by VHA, particularly for its Program of Comprehensive Assistance for Family Caregivers as part of its VFAI. A Standing Committee member asked whether the measure can be used for value-based purchasing. The developer replied that some of the states use fee-for-service delivery. However, it could also be used to gather data for value-based purchasing, assuming enough data are gathered in enough states to make comparisons. A Standing Committee member suggested avoiding using the measure for accountability until data collection could be streamlined. The developer replied that this measure captures the voice of the individual in the PCSP. The developer also stated that the measure could also be stratified by different types of individuals within a state. Based on this information, the Standing Committee member thought that because the measure was at the state level, it would be difficult to hold entities who are providing the service accountable for the measure. During offline voting, the Standing Committee passed the measure on the use criterion.

Regarding usability, the Standing Committee noted that because this is a new measure, its performance has not been measured over time. The developer stated in their submission that they were surprised by the extent of the TEP and reviewer agreement on the importance of the measure for aligning functional needs with service planning. A Standing Committee member provided a supportive comment, stating that the benefits of this measure outweigh any potential harm. Another Standing Committee member questioned whether the measure would be usable to truly measure quality at the state level because it may be difficult to understand how performance is improving and how the end user may use the information. A question was asked about whether the measure is being used at the state level for quality improvement. The developer noted that the measure is expected to be used in states for the HCBS quality improvement projects and public reporting. The developer further stated that they are hoping to identify best practices and potentially stratify by groups (e.g., brain injury) at some point. During offline voting, the Standing Committee passed the measure on usability.

Because the Standing Committee did not reach consensus on reliability, a must-pass criterion, it did not vote on overall suitability for endorsement. The Standing Committee will re-vote on reliability, feasibility, and overall suitability for endorsement during the post-comment meeting.

The Standing Committee did not have a formal discussion on related or competing measures for this measure because it did not vote the recommendation for endorsement because it did not reach consensus on reliability–a must-pass criterion. However, a member of the Standing Committee did mention that insurance companies are measuring similar data through the Health Risk Assessment. Therefore, the private sector is already starting to measure similar information. No NQF member or public comments were received prior to the measure evaluation meeting.

Public Comment

Dr. Pickering opened the lines for NQF members and public comments. No public or NQF member comments were provided during either measure evaluation meeting.

Next Steps

Dr. Pickering provided an overview of the next steps. NQF will begin drafting a meeting summary of the Standing Committee's deliberations. Dr. Pickering reiterated the earlier statement about the transition of the endorsement and maintenance work to the new successor. Dr. Pickering thanked the Standing Committee for its time, engagement, and participation in this work and adjourned the call.