NATIONAL QUALITY FORUM
Driving measurable health improvements together

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 3720

Measure Title: Patient-Reported Fatigue Following Chemotherapy among Adults with Breast Cancer

Measure Steward: Purchaser Business Group on Health

Brief Description of Measure: 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). The measure is risk-adjusted.

Developer Rationale: Over the past decade, diverse stakeholders in the cancer community have increased calls for the widespread integration of patient reported outcome (PRO) assessment into routine cancer care – as well as the related development of PRO-based performance measures (PRO-PMs) to allow these patient-centered outcomes to be implemented in quality measurement and improvement initiatives. However, PRO assessment in routine care remains underutilized, and very few PRO-PMs have been validated for the cancer population. Moreover, much of the initial exploration and research that has occurred in these areas has focused on cancer patients with advanced disease, despite the fact that the majority of people with cancer are diagnosed with earlier stage, curative disease. A growing body of evidence documents the persistence of symptoms for months and even years after the completion of treatment experienced by people receiving curative cancer treatment (NQF 2017). Among this patient population, it is important to consider the acute symptoms associated with treatment, as well as symptoms of cancer diagnosis and treatment that impact entry into the survivorship phase, hindering patients' abilities to regain functional status following treatment.

This PRO-PM is focused on fatigue in patients with breast cancer. Breast cancer is a common diagnosis treated in both community and hospital-based oncology settings. The PRO-PM fills a gap in the existing measurement set for cancer care, will directly support performance improvement in the delivery of cancer care, and can support accountability and value-based payment. The PROMOnc conceptual development was grounded in the evidence-based premise that medical oncologists who provide the highest quality care (including medical and non-medical support services) to patients receiving curative-intent cytotoxic therapy can reduce longer-term symptom burden and thus improve patient transition into the cancer survivorship period (NCCN 2018; Smith et al. 2019; Bubis et al. 2018).

Research indicates that patient self-reported symptoms are more accurate than clinician assessment of patients' symptoms, where clinicians frequently over-assessed the level of functioning of the patient and under-reported symptoms (Bottomley 2002; Chandwani et al. 2017). Research also reveals that fatigue is a commonly occurring symptom for cancer patients, estimated between 80 – 100% (Cleveland Clinic). Fatigue can be caused by the cancer itself and by

multiple other factors, including treatment-related side effects, such as anemia, pain, stress, and sleep disturbance. Unfortunately, fatigue can persist after the completion of treatment; for example, 20% of patients surveyed at 1-5 and 5-10 years reported fatigue at both time points (Broeckel et al. 1998; Curt et al. 2000; Bower et al. 2006). Management of fatigue is consistently identified as a priority by patients, clinicians, and researchers. Patient reported outcomes are the best source for measurement of fatigue (Borneman et al. 2010; Basch et al. 2015). However, studies show that clinicians do not always assess or treat fatigue, nor do patients always report fatigue to their care teams (Berger et al. 2010; Koornstra et al. 2014; Nyrop et al. 2016).

As a result of oncologists assessing and actively managing symptoms during chemotherapy, patients with breast cancer will experience lower symptom burden, less suffering, and will be better prepared and have lower persistent symptom burden as they enter the survivorship phase. In addition, group-level PRO-PM data are used for quality improvement, leading to practice changes. Payers can promote these practice changes that improve patient outcomes by rewarding high-performing physicians and practices.

References:

- Basch E, Spertus J, Dudley RA, et al. (2015). Methods for developing patient-reported outcome-based performance measures (PRO-PMs). Value in Health 18(4):493 – 504.
- Berger, A. M., Abernethy, A. P., Atkinson, A., Barsevick, A. M., Breitbart, W. S., Cella, D., et al. (2010). Cancer-related fatigue. JNCCN Journal of the National Comprehensive Cancer Network. 8(8):904-931.
- Borneman T et al. (2010). Reducing patient barriers to pain and fatigue management. J Pain Symptom Manage. 39(3):486-501.
- Bottomley A. The cancer patient and quality of life. The Oncologist. 2002. 7(2): 120-125.
- Bower JE, Ganz PA, Desmond KA, Bernaards C, Rowland JH, Meyerowitz BE, Belin TR. (2006). Fatigue in long-term breast carcinoma survivors: a longitudinal investigation. Cancer. 106(4):751-8.
- Broeckel JA, Jacobsen, PB, Horton J, Balducci, L, Lyman GH. (1998). Characteristics and correlates of fatigue after adjuvant chemotherapy for breast cancer. Journal of Clinical Oncology. 16(5):1689-96.
- Bubis, L. D., Davis, L., Mahar, A., Barbera, L., Li, Q., Moody, L., Karanicolas, P., Sutradhar, R., & Coburn, N. G. (2018). Symptom burden in the first year after cancer diagnosis: An analysis of patient-reported outcomes. Journal of Clinical Oncology, 36(11), 1103-1111. https://doi.org/10.1200/jco.2017.76.0876
- Chandwani KD, et al. J Pain Symptom Manage. 2017;53(6):988-998. https://my.clevelandclinic.org/health/diseases/5230-cancer-fatigue
- Cleveland Clinic Health Library. Cancer Fatigue. https://my.clevelandclinic.org/health/diseases/5230-cancer-fatigue
- Curt GA, Breitbart W, Cella D, et al. (2000). Impact of cancer-related fatigue on the lives of patients: new findings from the Fatigue Coalition. Oncologist. 5(5):353–360.
- Koornstra et al. (2014). Management of fatigue in patients with cancer – A practical overview. Cancer Treatment Reviews. 40791–799.
- National Comprehensive Cancer Network (NCCN) clinical practice guidelines in oncology, adult cancer pain Version I. NCCN, 2018 https://www.nccn.org/professionals/physician_gls/pdf/pain.pdf National Comprehensive Cancer Network (NCCN) clinical practice guidelines in oncology, adult cancer fatigue Version I. NCCN, 2018 https://www.nccn.org/professionals/physician_gls/pdf/fatigue.pdf
- National Quality Forum. Cancer 2015-2017 Technical Report. January 13, 2017
- Nyrop K, et al. (2016). Physical activity communication between oncology providers and patients with early-stage breast, colon, or prostate cancer. Cancer. 122(3):470-6.
- Smith, T. G., Troeschel, A. N., Castro, K. M., Arora, N. K., Stein, K., Lipscomb, J., Brawley, O. W., McCabe, R. M., Clauser, S. B., & Ward, E. (2019). Perceptions of patients with breast and colon cancer of the management of cancer-related pain, fatigue, and emotional distress in community oncology. Journal of Clinical Oncology, 37(19), 1666-1676. https://doi.org/10.1200/jco.18.01579

Numerator Statement: The PRO-PM numerator is the group-level PROMIS Fatigue score at the follow-up survey.

Denominator Statement: Adult patients with stages I-III female breast cancer receiving an initial chemotherapy regimen within the measurement window.

Denominator Exclusions:

- Patients on a therapeutic clinical trial
- Patients with recurrence/disease progression
- Patients who leave the practice
- Patients who die

Measure Type: Outcome: PRO-PM

Data Source:

Electronic Health Records

Instrument-Based Data

Paper Medical Records

Level of Analysis:

Clinician: Group/Practice

Preliminary Analysis: New Measure

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a ***health outcome*** measure include providing empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service; if these data not available, data demonstrating wide variation in performance can be used, assuming the data are from a robust number of providers and the results are not subject to systematic bias. For measures derived from a patient report, the evidence also should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.

The developer provides the following description for this measure:

- This is a new outcome PRO-PM measure at the group/clinician that measures fatigue following chemotherapy for adult patients with breast cancer. It is based on the PROMOnc premise that medical oncologists who provide the highest quality care, in particular medical and non-medical support to patients with curative-intent cytotox therapy will be able to reduce symptom burden and, therefore improve patient transition into the cancer survivorship period.
- The developer provides the PROOnc 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 model states that specific evidence-based practices, if delivered by the group practice and clinician will experience lower symptom burden during the survivorship period.

Summary:

- The developer references the 2022 National Comprehensive Cancer Network (NCCN) Cancer-Related Fatigue 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 including the recognizing, evaluating, monitoring, documenting, and treating of fatigue.

Question for the Standing Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*
- *Does the target population value the measured outcome and find it meaningful?*

Guidance From the Evidence Algorithm

Measure is a PRO-PM (box 1)-> Relationship between PRO-PM and at least one healthcare action demonstrated (Box 2)-> Pass

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

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

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

- In 10 clinician groups that participated in the beta field test, there were 744 follow-up surveys and 323 were used for analysis.
- The average adjusted measure score was 48.51, with a range from 42.13 to 53.07 and standard deviation of 3.13.
- The confidence intervals for the highest and lowest groups did not overlap.

Disparities

- During testing, administrative data were collected on race or ethnicity, marital status, and insurance status (Medicaid or dual eligible). Race and ethnicity were also collected via the survey instrument
- The developer states that after adjustment for multiple comparisons, none of these variables were significant in their relationship with the measure but did not provide the data to support this conclusion.
- The developer notes that research studies have found that certain groups of survivors, such as racial/ethnic minorities and those of lower socioeconomic status, report poorer patient-reported outcomes and interventions to address those outcomes.

Questions for the Standing Committee:

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

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

Criteria 2: Scientific Acceptability of Measure Properties

Complex measure evaluated by the Scientific Methods Panel (SMP)? ☒ **Yes** ☐ **No**

Evaluators: Dave Nerenz; Patrick Romano; Jeff Geppert; Zhenqiu Lin; Joe Kunisch; Eric Weinhandl; Daniel Deutscher; John Bott; Ron Walters; Jennifer Perloff; Paul Kurlansky

- The SMP passed the measure on Reliability with a score of: H-0; M-9; L-1; I-0

- The SMP did not reach consensus on Validity with a score of: H-0; M-6; L-3; I-1

2a. Reliability: Specifications and Testing

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

2a2. Reliability testing demonstrates whether the measure data elements are repeatable and producing the same results a high proportion of the time when assessed in the same population in the same time period, and/or whether the measure score is precise enough to distinguish differences in performance across providers.

Specifications:

- The PRO-PM is the risk-adjusted, group-level mean of PROMIS Fatigue scores among adult women with breast cancer entering survivorship after the completion of chemotherapy administered with curative intent.
- Measure specifications are clear and precise.
- Measure specifications for this instrument-based measure also include the specific instrument (e.g., PROM[s]) and standard methods, modes, and languages of administration

Reliability Testing:

- Data were used from 7/1/19 to 4/1/22 at 10 group practices.
- Reliability testing was conducted at the encounter level and accountable-entity level.
    - The developer notes that PROMIS measures, including the Fatigue scale, have undergone rigorous development and validation. Several references are provided in the submission.
        - Reliability testing from the literature demonstrates that for the PROMIS Fatigue, the Cronbach's alpha is 0.86.
    - To test the reliability of the measure score, a signal-to-noise analysis was performed. To evaluate the measure's reliability for group-level reporting, hierarchical linear regression models were used to relate the outcome to providers and covariates. The hierarchy was patients' observations within groups.
        - The estimate of the adjusted ICC was 0.094. 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 estimates that in order to obtain a nominal reliability of 0.7, a minimum sample size of 23 patient respondents would be required. Group specific reliability ranged from 0.38 to 0.88, with a mean of 0.66 (SD=0.21) and a median reliability of 0.68.
        - The proportion of groups in the sample that had sufficient reliability using a reliability threshold of 0.70 was 50 percent.

SMP Summary:

- SMP passed the measure on reliability and, while it was pulled for discussion, chose not to revote on reliability.

Questions for the Standing Committee regarding reliability:

- *Do you have any concerns that the measure cannot be consistently implemented (i.e., are the measure specifications adequate)?*
- *The SMP is satisfied with the reliability testing for the measure. Does the Standing Committee think there is a need to discuss and/or vote on reliability?*

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

2b. Validity: Validity Testing; Exclusions; Risk Adjustment; Meaningful Differences; Comparability; Missing Data

2b2. Validity testing should demonstrate that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

2b2-2b6. Potential threats to validity should be assessed/addressed.

Validity Testing

- Validity testing was conducted at the patient/encounter level:
    - Critical data elements were evaluated by comparing the Patient-Reported Outcomes in Oncology (PROMOnc) and cancer registry datasets.
    - The developer stated that the majority of the clinical and demographic variables could be validated, but several variables were excluded from testing because they were not in the registry used for the validity testing.
        - Five hundred seventy patients were included in this analysis.
        - The percentage agreement by data element ranged from 71.63–100 percent.
        - Reported kappas ranged from 0.64–0.67.
        - The reported sensitivity ranged from 33.33–89.52 percent.
        - The specificity ranged from 60–99.80 percent.
- Validity testing of the measure score was conducted through a systematic assessment of face validity using a panel of 12 oncologist advisors. The following survey question was asked: "Rate your agreement with the following statement: The scores obtained from the measure as specified will provide an accurate reflection of quality and can be used to distinguish good and poor quality."
    - Eight of the 12 advisors participated in the survey.
    - All eight indicated "moderate agreement," "agreement," or "strong agreement" to the above question.
    - Three agreed or strongly agreed that the fatigue measure could differentiate good versus poor quality. Participants who did not rate the measure as 4 or 5 (i.e., agree or strongly agree) felt that fatigue was more susceptible to pandemic-related issues.
    - The four oncologists who declined to participate in the face validity voting expressed concerns regarding the impact of COVID-19 on sample size, and thus, performance scores. They requested additional data prior to voting.
- The developer notes that there is a validated, publicly available quality measure data related to this PRO-PM.
    - The developer states that PROMOnc TEP members hypothesized only moderate correlation between this measure and available measures (H-CAHPS, Outpatient Oncology Press Ganey

(note: different items were used across sites), and QOPI (note: different measures were used across sites).

- The developer analyzed correlations for any measure for which the TEP hypothesized a moderate association and for which we had data for at least 7 test sites.
- The developer reports that Pearson's Correlation Coefficients are in the moderate range (e.g., -0.033 to -0.567), as hypothesized, and in the appropriate direction (e.g., likely to recommend and degree to which care was well coordinated are associated with lower pain).

Exclusions

- There are four exclusions (n=frequency of those exclusions from the measure denominator):
  - Patients on an interventional or therapeutic clinical trial (n=18)
  - Patients who experience relapse or disease progression (n=0)
  - Patients who leave the practice (n=0)
  - Patients who die (n=1)
  - The developer states that it was not able to analyze the impact on measure outcomes of excluding these patients because follow-up survey data was not available for these patients.

Risk Adjustment

- A statistical model is used to risk-adjust this measure using 13 variables.
- To estimate risk-adjusted quality measure scores, hierarchical linear models that relate the patient-measure score to group scores conditioned on risk adjustment covariates were used.
- Model discrimination was tested during the Kendall tau. Comparing scores between null and the multivariate model adjustments for pain interference resulted in a value of 0.87. The Pearson correlation coefficient between the observed and predicted responses was 0.55.

Meaningful Differences

- To examine the ability of the measure to identify high- or low-performing groups, the developer calculated the number and percentage of groups that were significantly above or below the average score using risk adjustment.
- The mean group performance score was 48.51, and the standard deviation was 3.13, with a median score of 48.67 and a range of 42.13–53.07.
- Two of 10 groups had significantly different scores than the overall average, one more favorable and the other less favorable. Among those two groups, the mean absolute difference between the group's scores and the overall average was 4.9 points on a T-score scale (SD=10).
- The developer states that literature in the cancer population has suggested to define meaningful difference as between 3- and 5-point difference on a T-score scale that has a mean of 50 and standard deviation of 10 (Jensen et al., 2017; Yost, 2011).
- The developer reports that among group scores that were significantly above or below the average, the mean absolute difference between the group's scores and the overall average was 4.9 points, very close to half of the standard deviation (5 points).
- The developer concludes that these results indicate that the PRO-PM measure can discriminate between groups' performance.

Missing Data

- Both survey nonresponse and missing data were assessed.
- Across the 10 sites, 896 patients were eligible for follow-up and 19 met the exclusion criteria. The total number of follow-up surveys was 744, making up a survey administration rate of 85 percent. Among those surveys, 323 were completed and nine were ineligible. No statistical significance was identified, except that the respondents and nonrespondents differed on marital status and insurance.
- Missingness ranged from 0.00–0.93 percent for PROMIS item scales.

Comparability

- The measure only uses one set of specifications for this measure.

SMP Summary:

- The preliminary vote for NQF #3720 was consensus not reached (CNR) on validity and remained CNR after the SMP discussion. The CNR decision for validity was due to concerns with face validity testing, the lack of demonstration of meaningful differences, and the missing response rates.
- An SMP member stated that there were concerns with missing data and nonresponse bias, as well as the face validity. The results of the face validity vote did not demonstrate strong agreement among the TEP. Only three of eight TEP members agreed or strongly agreed that the measure could differentiate quality of care. One SMP member commented that given the way the scale was set up for face validity, in which a score of three out of five was "moderately agree," eight out of eight TEP members either strongly agreed, moderately agreed, or agreed that the measure could differentiate good versus poor quality of care. However, another SMP member commented that the four TEP members who did not vote stated that the pandemic confounded the issue of validity for this measure due to the inability of the measure to parse whether the fatigue is due to COVID-19 or cancer.
- An SMP member further questioned via comment the relationship between the COVID-19 pandemic, cancer care, and fatigue. COVID-19, in particular, can cause fatigue, particularly long-COVID, and confound the measure. However, another SMP member stated that this would be difficult to sort out methodologically, even after the pandemic ends and COVID-19 becomes endemic. Ultimately, the SMP re-voted on validity but still did not reach consensus on this criterion.

Questions for the Standing Committee regarding validity:

- *Do you have concerns regarding the validity of the measure (e.g., exclusions, risk adjustment approach, etc.)?*

Preliminary rating for validity: ☐ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient** ☒ **Consensus Not Reached**

Criterion 3. Feasibility

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

- PROMIS measures emanate from a survey that must be collected by staff and entered into the EHR in structured fields.

- The developer noted that during testing, documentation conducted in some provider notes, instead of in structured fields, but noted that this practice is changing. Some EHRs (Epic and Cerner), now include PROMIS surveys. However, this is not an eCQM.
- The developer stated that based on the clinical expertise and feasibility assessment of their technical expert panel, knowledge of the literature in oncology practice trends, the required data were in fact present in the medical record for the majority cases for which they were reported as missing during the testing.
- The developer stated that collecting the baseline survey with the originally defined timeframe from patients taking oral chemotherapy 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. Seven ADCC sites and two MOQC sites responded to the burden questionnaire. The majority of implementation burden was associated with administering the survey rather than collecting the clinical and demographic data elements; patient identification was also a challenge which test sites mitigated by building EHR reports to facilitate patient identification.
- The developer also fielded a survey to patients to assess their understanding of the survey and ease of us. Twelve patients provided feedback. Feedback indicated that 75% of respondents reported that it took them less than 10 minutes to complete the PROMOnc survey; 92% reported that they understood the survey instructions; 83% reported that they didn't have any technical issues completing the survey; and 83% felt that the time that it took to complete the survey was reasonable.

Questions for the Standing Committee:

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

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

Criterion 4: Use and Usability

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

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

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

Current uses of the measure

Publicly reported?	☐	Yes	☒	No		
Current use in an accountability program?	☐	Yes	☒	No	☐	UNCLEAR
Planned use in an accountability program?	☒	Yes	☐	No	☐	NA

Accountability program details

- The developer states that the measure will be submitted to the Measures Under Consideration (MUC) List for potential inclusion the CMS Quality Payment Program.

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

Feedback on the measure provided by those being measured or others

- There were regular meetings between the developer and practice managers and oncologists at test sites. They were actively involved in the development process for the measure.
- Patients and caregivers were engaged throughout the testing process. The developer engaged the Patient and Caregiver Oncology Quality Council from the Michigan Oncology Quality Consortium (MOQC) to provide input into the selection of PROMIS scales for assessing patient-reported outcomes.

***Questions for the Standing Committee**:*

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

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

4b. Usability (4b1. Improvement; 4b2. Benefits of measure)

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

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

Improvement results

- The developer states that the measure just completed testing and has not been used for performance improvement at the time for submission of endorsement

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

Unexpected findings (positive or negative) during implementation

- The developer states that there were no unexpected findings.

Potential harms

- The developer states that there were no potential harms were identified.

***Questions for the Standing Committee**:*

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

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

Criterion 5: Related and Competing Measures

Related Measures

- 0220: Adjuvant hormonal therapy is recommended or administered within 1 year (365 days) of diagnosis for women with AJCC T1cN0M0 or Stage IB – Stage III hormone receptor positive breast cancer
- 0387e: Breast Cancer: Hormonal Therapy for Stage I (T1b)-IIIC Estrogen Receptor/Progesterone Receptor (ER/PR) Positive Breast Cancer

Harmonization

- The developer noted that there are no NQF-endorsed measures with the same focus. NQF measures 0220 and 0387e have overlapping target populations: women receiving curative breast cancer treatment. See 5.05 for more details.

Criteria 1: Importance to Measure and Report

1a. Evidence

1a.01. Provide a logic model.

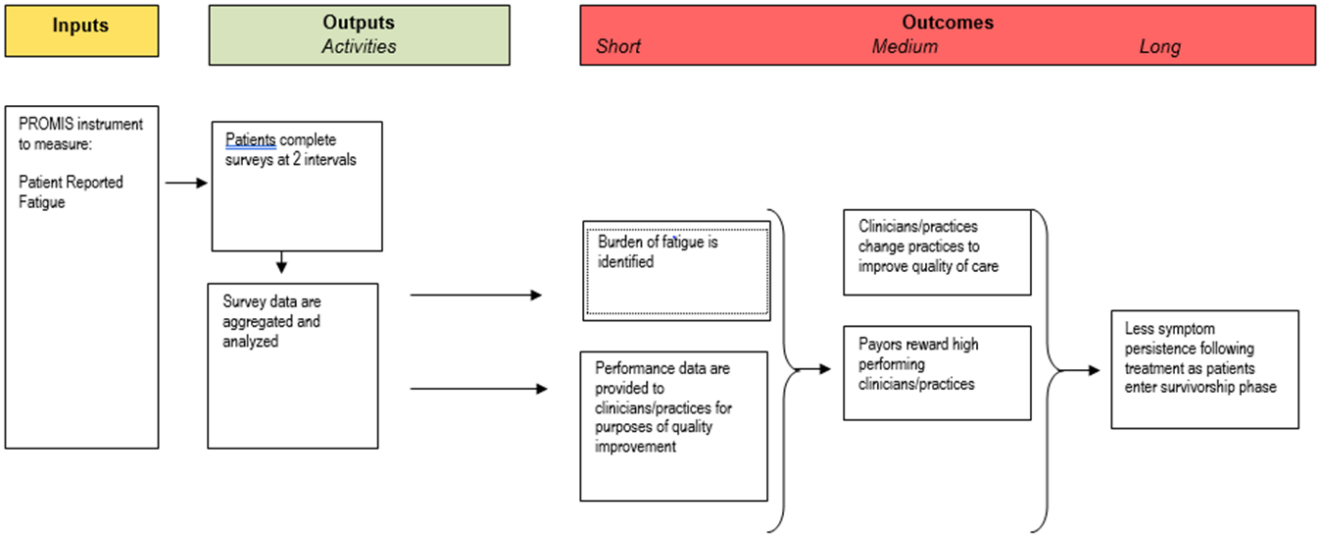
Briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

[Response Begins]

Many patients who undergo chemotherapy with curative intent experience persistent detriments following treatment. Common persistent symptoms include pain, fatigue and detriments to health-related quality of life. Evidence-based practices can manage these symptoms during treatment and position patients better for the survivorship phase. This PRO-PM assesses fatigue following completion of chemotherapy administered for adult patients with breast cancer. Data from this measure provides insight into the effectiveness of medical oncologists in helping patients to minimize the persistent impact of their treatments.

The PROMOnc Logic Model (Figure 1a.1) depicts the anticipated improvements to care provided and received, as well as medium and long term system impacts. As a result of oncologists assessing and actively managing symptoms during chemotherapy, patients will experience lower symptom burden, less suffering, and will be better prepared and have lower persistent symptom interference as they enter the survivorship phase. Group-level PRO-PM data are useful to inform practice improvement. Payers can promote these practice changes that improve patient outcomes by rewarding high-performing physicians and practices.

Figure 1a.1: PROMOnc Logic Model



[Response Ends]

1a.02. Provide evidence that the target population values the measured outcome, process, or structure and finds it meaningful.

Describe how and from whom input was obtained.

[Response Begins]

Patients guided many aspects of the PROMIS survey development process. PROMIS development methods included patient input to inform the development of the questionnaire items, using feedback from patient focus groups about the outcome domains to make sure that the questions reflect how potential respondents experience the symptoms and outcomes. Focus groups included patients with and without chronic illness who had experienced a range of severity or limitation in the domain (or outcome) in question. PROMIS researchers also conducted cognitive interviews to review each question item. In one study, for example, for Fatigue, PROMIS researchers convened 3 workgroups with a total of 17 participants. (DeWalt et al. 2007).

Patients and caregivers were also engaged throughout the PROMOnc testing process. Two representatives from the MOQC Patient and Caregiver Oncology Quality Council participated on the PROMOnc Steering Committee. See Additional (2) for the Steering Committee roster. When the PROMOnc TEP was originally formed, there were two patient representatives, one who was formerly in an advocacy role at Patients Like Me and one who was an administrator at MOQC, nurse practitioner and a patient. During the measure development period, Patients Like Me was acquired by United Health Group (but this representative continued with the TEP) and the other patient excused herself from the TEP when she transitioned to a new job. Moreover, rather than rely on just the personal experience of a small number of patients on the TEP, we engaged the MOQC Patient and Caregiver Oncology Quality Council several times to provide input on key issues such as the outcomes to be measured and the selection of the PROMIS scales for the PROMOnc survey. The Patient and Caregiver Oncology Quality Council is diverse in terms of age, gender, race/ethnicity, cancer type, LGBTQ+, etc. More information about this council can be found here: https://moqc.org/moqc/poqc/And, PROMOnc collaborated with the Seattle Cancer Care Alliance (SCCA) Patient Family Advisory Council (PFAC) to understand acceptability and burden of the PROMIS scales, and in implementation of a patient burden questionnaire during testing.

Reference:

- DeWalt DA, Rothrock N, Yount S, Stone AA. Evaluation of Item Candidates: The PROMIS Qualitative Item Review. Med Care. 2007 May; 45(5 Suppl 1): S12-S21.

[Response Ends]

1a.03. Provide empirical data demonstrating the relationship between the outcome (or PRO) and at least one healthcare structure, process, intervention, or service.

[Response Begins]

PRO-PMs are especially important in cancer care since diagnoses have substantial impact on psychological and physical health (Valderas et al. 2008; Chen et al 2013; Kotronoulas et al. 2014; Basch et al. 2016). Furthermore, multiple treatment modalities are generally used, each of which has potential side effects which may go undetected unless patients provide feedback (Henry et al. 2008; Fromme et al. 2004; Laugsand et al. 2010). Fortunately, a growing body of research suggests that collecting and using patient-reported symptoms during cancer care can improve patient outcomes, including survival (Basch et al. 2016; Seow et al. 2012; Kroenke et al. 2014; Gilbert et al. 2012; Valderas et al. 2008; Chen et al. 2013; Kotronoulas et al. 2014). The collection of PROs can enable providers to assess patients using a standardized symptom assessment process, facilitate appropriate follow-up to ensure patient needs are addressed, and support patient-provider communication and the development of shared care plans, which assess different factors at different points of the treatment journey (e.g., before chemotherapy, during treatment, and into survivorship).

This measure assesses patient-reported fatigue following chemotherapy for adults with breast cancer. Unfortunately, fatigue is a common and is perceived as the most distressing symptom associated with cancer and cancer treatment

(Hinds et al. 2000). Fatigue is reported by about 80 percent of patients receiving chemotherapy (NCCN Fatigue 2022). A metanalysis of breast cancer survivors showed that receipt of chemotherapy was a predictor of severe fatigue (Abrahams et al. 2016). Fatigue can be caused by the cancer itself and by multiple other factors, including treatment-related side effects, such as anemia, pain, stress, and sleep disturbance. Cancer survivors report fatigue as disruptive for months or years after completion of treatment (Broeckel et al. 1998; Curt et al. 2000; Bower et al. 2006). According to the National Comprehensive Cancer Network (NCCN), persistent cancer-related fatigue affects survivors' quality of life and impacts time to return to work after treatment (NCCN Fatigue 2022).

Management of fatigue is consistently identified as a priority by patients, clinicians, and researchers. Yet, fatigue following a cancer diagnosis is underreported, underdiagnosed, and undertreated (NCCN Fatigue 2022). Patient reported outcomes are the best source for measurement of fatigue (NCCN Fatigue 2022; Borneman et al. 2010; Basch et al. 2015). Studies show that clinicians do not always assess or treat fatigue, nor do patients always report fatigue to their care teams (Berger et al. 2010; Koornstra et al. 2014; Nyrop et al 2016).

The National Comprehensive Cancer Network (NCCN) issues clinical practice guidelines for oncology. The NCCN Cancer-Related Fatigue Guideline (2022) states that "Fatigue should be recognized, evaluated, monitored, documented, and treated promptly for all age groups, at all stages of disease prior to, during, and following treatment." The guideline contains comprehensive algorithms and guidelines for the treatment of fatigue in cancer patients, during and following treatment. (https://www.nccn.org/professionals/physician_gls/pdf/fatigue.pdf). The NCCN Survivorship Guideline (2022) further notes that "Fatigue is a subjective experience that should be systematically assessed using patient self-reports and other sources of data for cancer survivors in the months and years after diagnosis". Both guidelines state that "Patients and family/caregiver(s) should be informed that management of fatigue is an integral part of total health care and that fatigue can persist following treatment" (NCCN Fatigue 2022; NCCN Survivorship 2022).

The PROMOnc PRO-PM for fatigue will provide oncologists with data that can drive improvements in the management of fatigue during chemotherapy and after the completion of chemotherapy. In one example of an improvement study, Bennett et al. (2016) found that clinician support to patients had an effect on reducing fatigue intensity, fatigue's interference with daily life, and general fatigue.

References:

- Abrahams HJG, Gielissen MFM, Schmits IC, Verhagen CAHHVM, Rovers MM, Knoop H. Risk factors, prevalence, and course of severe fatigue after breast cancer treatment: a meta-analysis involving 12,327 breast cancer survivors. Ann Oncol. 2016 Jun;27(6):965-974. doi: 10.1093/annonc/mdw099. Epub 2016 Mar 2. PMID: 26940687.
- Basch, E., Deal, A. M., Kris, M. G., Scher, H. I., Hudis, C. A., Sabbatini, P., Rogak, L., Bennett, A. V., Dueck, A. C., Atkinson, T. M., Chou, J. F., Dulko, D., Sit, L., Barz, A., Novotny, P., Fruscione, M., Sloan, J. A., & Schrag, D. (2016). Symptom monitoring with patient-reported outcomes during routine cancer treatment: A randomized controlled trial. Journal of Clinical Oncology, 34(6), 557-565. https://doi.org/10.1200/jco.2015.63.0830
- Basch E, Spertus J, Dudley RA, et al. (2015). Methods for developing patient-reported outcome-based performance measures (PRO-PMs). Value in Health 18(4):493 – 504.
- Bennett, S., Pigott, A., et al. Educational interventions for the management of cancer-related fatigue in adults. Cochrane Database of Systematic Reviews 2016, Issue 11. Art. No.: CD008144. DOI: 10.1002/14651858.CD008144.pub2.
- Berger, A. M., Abernethy, A. P., Atkinson, A., Barsevick, A. M., Breitbart, W. S., Cella, D., et al. (2010). Cancer-related fatigue. JNCCN Journal of the National Comprehensive Cancer Network. 8(8):904-931. Borneman T et al. (2010). Reducing patient barriers to pain and fatigue management. J Pain Symptom Manage. 39(3):486-501.
- Borneman T et al. (2010). Reducing patient barriers to pain and fatigue management. J Pain Symptom Manage. 39(3):486-501.
- Bower JE, Ganz PA, Desmond KA, Bernaards C, Rowland JH, Meyerowitz BE, Belin TR. (2006). Fatigue in long-term breast carcinoma survivors: a longitudinal investigation. Cancer. 106(4):751-8.
- Broeckel JA, Jacobsen, PB, Horton J, Balducci, L, Lyman GH. (1998). Characteristics and correlates of fatigue after adjuvant chemotherapy for breast cancer. Journal of Clinical Oncology. 16(5):1689-96.
- Chen, J., Ou, L., & Hollis, S. J. (2013). A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. BMC Health Services Research, 13(1). https://doi.org/10.1186/1472-6963-13-211

- Curt GA, Breitbart W, Cella D, et al. (2000). Impact of cancer-related fatigue on the lives of patients: new findings from the Fatigue Coalition. Oncologist. 5(5):353–360.
- Fromme, E. K., Eilers, K. M., Mori, M., Hsieh, Y., & Beer, T. M. (2004). How accurate is clinician reporting of chemotherapy adverse effects? A comparison with patient-reported symptoms from the quality-of-Life questionnaire C30. Journal of Clinical Oncology, 22(17), 3485-3490. https://doi.org/10.1200/jco.2004.03.025
- Gilbert, J. E., Howell, D., King, S., Sawka, C., Hughes, E., Angus, H., & Dudgeon, D. (2012). Quality improvement in cancer symptom assessment and control: The provincial palliative care integration project (PPCIP). Journal of Pain and Symptom Management, 43(4), 663-678. https://doi.org/10.1016/j.jpainsymman.2011.04.028
- Henry, D. H., Viswanathan, H. N., Elkin, E. P., Traina, S., Wade, S., & Cella, D. (2008). Symptoms and treatment burden associated with cancer treatment: Results from a cross-sectional national survey in the U.S. Supportive Care in Cancer, 16(7), 791-801. https://doi.org/10.1007/s00520-007-0380-2
- Hinds PS, Quargnenti A, Bush AJ, Pratt C, Fairclough D, Rissmiller G, Betcher D, Gilchrist GS. An evaluation of the impact of a self-care coping intervention on psychological and clinical outcomes in adolescents with newly diagnosed cancer. Eur J Oncol Nurs. 2000 Mar;4(1):6-17; discussion 18-9. doi: 10.1054/ejon.1999.0051. PMID: 12849624.
- Koornstra et al. (2014). Management of fatigue in patients with cancer – A practical overview. Cancer Treatment Reviews. 40791–799.
- Kotronoulas, G., Kearney, N., Maguire, R., Harrow, A., Di Domenico, D., Croy, S., & MacGillivray, S. (2014). What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials. Journal of Clinical Oncology, 32(14), 1480-1501. https://doi.org/10.1200/jco.2013.53.5948
- Kroenke, K., Krebs, E. E., Wu, J., Yu, Z., Chumbler, N. R., & Bair, M. J. (2014). Telecare collaborative management of chronic pain in primary care. JAMA, 312(3), 240. https://doi.org/10.1001/jama.2014.7689
- Laugsand, E. A., Sprangers, M. A., Bjordal, K., Skorpen, F., Kaasa, S., & Klepstad, P. (2010). Health care providers underestimate symptom intensities of cancer patients: A multicenter European study. Health and Quality of Life Outcomes, 8(1), 104. https://doi.org/10.1186/1477-7525-8-104
- National Comprehensive Cancer Network (NCCN) Clinical Practice Guidelines in Oncology, Cancer-Related Fatigue, Version 2.2022. NCCN, 2022 https://www.nccn.org/guidelines/category_3
- National Comprehensive Cancer Network (NCCN) Clinical Practice Guidelines in Oncology, Survivorship, Version 1.2022. NCCN, 2022 https://www.nccn.org/guidelines/category_3
- Nyrop K, et al. (2016). Physical activity communication between oncology providers and patients with early-stage breast, colon, or prostate cancer. Cancer. 122(3):470-6.
- Seow, H., Sussman, J., Martelli-Reid, L., Pond, G., & Bainbridge, D. (2012). Do high symptom scores trigger clinical actions? An audit after implementing electronic symptom screening. Journal of Oncology Practice, 8(6), e142-e148. https://doi.org/10.1200/jop.2011.000525
- Valderas, J. M., Kotzeva, A., Espallargues, M., Guyatt, G., Ferrans, C. E., Halyard, M. Y., Revicki, D. A., Symonds, T., Parada, A., & Alonso, J. (2008). The impact of measuring patient-reported outcomes in clinical practice: A systematic review of the literature. Quality of Life Research, 17(2), 179-193. https://doi.org/10.1007/s11136-007-9295-0

[Response Ends]

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

1b.01. Briefly explain the rationale for this measure.

Explain how the measure will improve the quality of care, and list the benefits or improvements in quality envisioned by use of this measure.

[Response Begins]

Over the past decade, diverse stakeholders in the cancer community have increased calls for the widespread integration of patient reported outcome (PRO) assessment into routine cancer care – as well as the related development of PRO-based performance measures (PRO-PMs) to allow these patient-centered outcomes to be implemented in quality measurement and improvement initiatives. However, PRO assessment in routine care remains underutilized, and very

few PRO-PMs have been validated for the cancer population. Moreover, much of the initial exploration and research that has occurred in these areas has focused on cancer patients with advanced disease, despite the fact that the majority of people with cancer are diagnosed with earlier stage, curative disease. A growing body of evidence documents the persistence of symptoms for months and even years after the completion of treatment experienced by people receiving curative cancer treatment (NQF 2017). Among this patient population, it is important to consider the acute symptoms associated with treatment, as well as symptoms of cancer diagnosis and treatment that impact entry into the survivorship phase, hindering patients' abilities to regain functional status following treatment.

This PRO-PM is focused on fatigue in patients with breast cancer. Breast cancer is a common diagnosis treated in both community and hospital-based oncology settings. The PRO-PM fills a gap in the existing measurement set for cancer care, will directly support performance improvement in the delivery of cancer care, and can support accountability and value-based payment. The PROMOnc conceptual development was grounded in the evidence-based premise that medical oncologists who provide the highest quality care (including medical and non-medical support services) to patients receiving curative-intent cytotoxic therapy can reduce longer-term symptom burden and thus improve patient transition into the cancer survivorship period (NCCN 2018; Smith et al. 2019; Bubis et al. 2018).

Research indicates that patient self-reported symptoms are more accurate than clinician assessment of patients' symptoms, where clinicians frequently over-assessed the level of functioning of the patient and under-reported symptoms (Bottomley 2002; Chandwani et al. 2017). Research also reveals that fatigue is a commonly occurring symptom for cancer patients, estimated between 80 – 100% (Cleveland Clinic). Fatigue can be caused by the cancer itself and by multiple other factors, including treatment-related side effects, such as anemia, pain, stress, and sleep disturbance. Unfortunately, fatigue can persist after the completion of treatment; for example, 20% of patients surveyed at 1-5 and 5-10 years reported fatigue at both time points (Broeckel et al. 1998; Curt et al. 2000; Bower et al. 2006). Management of fatigue is consistently identified as a priority by patients, clinicians, and researchers. Patient reported outcomes are the best source for measurement of fatigue (Borneman et al. 2010; Basch et al. 2015). However, studies show that clinicians do not always assess or treat fatigue, nor do patients always report fatigue to their care teams (Berger et al. 2010; Koornstra et al. 2014; Nyrop et al. 2016).

As a result of oncologists assessing and actively managing symptoms during chemotherapy, patients with breast cancer will experience lower symptom burden, less suffering, and will be better prepared and have lower persistent symptom burden as they enter the survivorship phase. In addition, group-level PRO-PM data are used for quality improvement, leading to practice changes. Payers can promote these practice changes that improve patient outcomes by rewarding high-performing physicians and practices.

References:

- Basch E, Spertus J, Dudley RA, et al. (2015). Methods for developing patient-reported outcome-based performance measures (PRO-PMs). Value in Health 18(4):493 – 504.
- Berger, A. M., Abernethy, A. P., Atkinson, A., Barsevick, A. M., Breitbart, W. S., Cella, D., et al. (2010). Cancer-related fatigue. JNCCN Journal of the National Comprehensive Cancer Network. 8(8):904-931.
- Borneman T et al. (2010). Reducing patient barriers to pain and fatigue management. J Pain Symptom Manage. 39(3):486-501.
- Bottomley A. The cancer patient and quality of life. The Oncologist. 2002. 7(2): 120-125.
- Bower JE, Ganz PA, Desmond KA, Bernaards C, Rowland JH, Meyerowitz BE, Belin TR. (2006). Fatigue in long-term breast carcinoma survivors: a longitudinal investigation. Cancer. 106(4):751-8.
- Broeckel JA, Jacobsen, PB, Horton J, Balducci, L, Lyman GH. (1998). Characteristics and correlates of fatigue after adjuvant chemotherapy for breast cancer. Journal of Clinical Oncology. 16(5):1689-96.
- Bubis, L. D., Davis, L., Mahar, A., Barbera, L., Li, Q., Moody, L., Karanicolas, P., Sutradhar, R., & Coburn, N. G. (2018). Symptom burden in the first year after cancer diagnosis: An analysis of patient-reported outcomes. Journal of Clinical Oncology, 36(11), 1103-1111. https://doi.org/10.1200/jco.2017.76.0876
- Chandwani KD, et al. J Pain Symptom Manage. 2017;53(6):988-998. https://my.clevelandclinic.org/health/diseases/5230-cancer-fatigue
- Cleveland Clinic Health Library. Cancer Fatigue. https://my.clevelandclinic.org/health/diseases/5230-cancer-fatigue
- Curt GA, Breitbart W, Cella D, et al. (2000). Impact of cancer-related fatigue on the lives of patients: new findings from the Fatigue Coalition. Oncologist. 5(5):353–360.

- Koornstra et al. (2014). Management of fatigue in patients with cancer – A practical overview. Cancer Treatment Reviews. 40791–799.
- National Comprehensive Cancer Network (NCCN) clinical practice guidelines in oncology, adult cancer pain Version I. NCCN, 2018 https://www.nccn.org/professionals/physician_gls/pdf/pain.pdf National Comprehensive Cancer Network (NCCN) clinical practice guidelines in oncology, adult cancer fatigue Version I. NCCN, 2018 https://www.nccn.org/professionals/physician_gls/pdf/fatigue.pdf
- National Quality Forum. Cancer 2015-2017 Technical Report. January 13, 2017
- Nyrop K, et al. (2016). Physical activity communication between oncology providers and patients with early-stage breast, colon, or prostate cancer. Cancer. 122(3):470-6.
- Smith, T. G., Troeschel, A. N., Castro, K. M., Arora, N. K., Stein, K., Lipscomb, J., Brawley, O. W., McCabe, R. M., Clauser, S. B., & Ward, E. (2019). Perceptions of patients with breast and colon cancer of the management of cancer-related pain, fatigue, and emotional distress in community oncology. Journal of Clinical Oncology, 37(19), 1666-1676. https://doi.org/10.1200/jco.18.01579

[Response Ends]

1b.02. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.

Include mean, std dev, min, max, interquartile range, and scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.

[Response Begins]

A total of 21 groups participated in the beta field test and 10 groups were included in the final sample. We fielded a total of 744 follow-up surveys, and 323 completed surveys were used for analysis.

Based on the testing sample (N=10 groups), the average adjusted measure score is 48.51. Adjusted group scores range from 42.13 to 53.07 with a standard deviation of 3.13. Confidence intervals for the highest and lowest group scores do not overlap: Lowest group CI: (39.36, 44.90); Highest group CI: (47.63, 58.51). One group has significantly higher score than the average, while one other group has significantly lower score. The observed variability across groups supports the potential of the measure to distinguish among groups with high, medium, and low performance.

Table 1b.1 shows the mean, standard deviation, minimum, maximum, and interquartile range of the group adjusted scores. Table 1b.2 shows the deciles of the observed group adjusted scores (N=10).

Table 1b.1: Distribution of Group-Level Scores

Measure	Mean	Standard Deviation	Median	Minimum	Maximum	1st Quartile	3rd Quartile	Inter-Quartile Range
Fatigue	48.51	3.13	48.67	42.13	53.07	47.46	50.44	2.98

Table 1b.2: Deciles of the Observed Group Adjusted Scores (N=10)

Measure	10th Percentile	20th Percentile	30th Percentile	40th Percentile	50th Percentile	60th Percentile	70th Percentile	80th Percentile	90th Percentile
Fatigue	43.82	46.48	47.74	48.10	48.67	49.21	49.85	51.18	52.50

Table 1b.2 shows the deciles of the observed group adjusted scores (N=10).

[Response Ends]

1b.03. If no or limited performance data on the measure as specified is reported above, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement. Include citations.

[Response Begins]

See 1b.02.

[Response Ends]

1b.04. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.

Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included. Include mean, std dev, min, max, interquartile range, and scores by decile. For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.

[Response Begins]

To understand if and to what extent disparities in measure reporting and patient experience exist, we evaluated the relationship of various social risk factors to the measure score and the groups.

For all eligible patients during testing, administrative data were collected on race or ethnicity, marital status, and insurance status (Medicaid or dual eligible). Race and ethnicity were also collected via the survey instrument. Among survey respondents included in the measure, 7.7 percent are Hispanic, 10.5 percent are non-Hispanic black, 7.7 percent are non-Hispanic Asian, and 66.9 percent are non-Hispanic white; 3.4 percent have Medicaid or are dual eligible; 72.1 percent are married. After adjustment for multiple comparisons, none of these variables were significant in their relationship with the measure.

[Response Ends]

1b.05. If no or limited data on disparities from the measure as specified is reported above, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in above.

[Response Begins]

Myriad research studies reveal that certain groups of survivors, such as racial/ethnic minorities and those of lower socioeconomic status, report poorer patient-reported outcomes and interventions to address those outcomes. African American women diagnosed with breast cancer are more likely to experience higher pain levels compared to other groups (Green et al. 2011). Black patients report significantly higher pain intensity, more pain-related distress, and more pain-related interference with function than white patients (Vallerand et al. 2005). Cancer survivors who are nonwhite, less educated, older, and/or have comorbidities are less likely to receive adequate cancer pain management (Stein et al. 2016). Black women are more likely to experience cancer-related fatigue than women of other racial and ethnic groups (Swen et al. 2017). Moreover, research indicates income disparities in the quality of life of cancer survivors (Short et al. 2006), along with racial and ethnic disparities, with Hispanics and blacks reporting a higher burden of poor QOL compared with white patients (Hildebrandt 2017; Short et al. 2006). A 2021 study of women with early stage breast cancer in Tennessee combined EHR and patient-reported data, and found that pain perception was significantly associated with

poverty and blight level of the neighborhood, after adjustment for demographic characteristics, cancer stage, and chemotherapy (Choi et al. 2022). Madison et al. (2021) assessed patient-reported cancer-related distress, perceived stress, anxiety and depressive symptoms, fatigue, and pain. They found significantly more cancer-related distress, perceived stress, emotional fatigue, and vigor among Black compared to White survivors, with symptoms improving by 6 months post-treatment among White women but persisting among Black women. Unfortunately, disparities are reflected throughout many breast cancer outcomes, including survival. Women who are Black and of lower socioeconomic status, for instance, have higher breast cancer mortality rates (Kantor et al. 2022).

References:

- Choi HY, Graetz I, Shaban-Nejad A, Schwartzberg L, Vidal G, Davis RL, Shin EK (2022). Social Disparities of Pain and Pain Intensity Among Women Diagnosed With Early Stage Breast Cancer. Frontiers in oncology, 12, 759272. https://doi.org/10.3389/fonc.2022.759272
- Green C, Hart-Johnson T, Loeffler D (2011) Cancer-related chronic pain: examining quality of life in diverse cancer survivors. Cancer 117:1994–2003
- Hildebrandt M. Minority Colorectal Cancer Patients Report Higher Burden of Poor Quality-of-Life than Whites. MD Anderson Newsroom. March 27, 2017. https://www.mdanderson.org/newsroom/minority-colorectal-cancer-patients-report-higher-burden-of-poor-quality-of-life-than-whites.h00-159143667.html
- Kantor O, Wang ML, Bertrand K, et al. Racial and socioeconomic disparities in breast cancer outcomes within the AJCC pathologic prognostic staging system. Ann Surg Oncol. 2022;29(1):686–96. https://doi.org/10.1245/s10434-021-10527-8.
- Madison, A. A., Peng, J., Shrout, M. R., Renna, M. E., Alfano, C. M., Povoski, S. P., Lipari, A. M., Agnese, D. M., Carson, W. E., Malarkey, W. B., & Kiecolt-Glaser, J. K. (2021). Distress Trajectories in Black and White Breast Cancer Survivors: From Diagnosis to Survivorship. Psychoneuroendocrinology, 131, 105288. https://doi.org/10.1016/j.psyneuen.2021.105288
- Swen M, Mann A, Paxton RJ, Dean LT. Do Cancer-Related Fatigue and Physical Activity Vary by Age for Black Women With a History of Breast Cancer?. Prev Chronic Dis. 2017;14:E122. Published 2017 Nov 30. doi:10.5888/pcd14.170128
- Short PF, Mallonee EL. Income disparities in the quality of life of cancer survivors. Med Care. 2006;44(1):16-23. doi:10.1097/01.mlr.0000188986.84819.3a
- Stein KD, Alcaraz KI, Kamson C, Fallon EA, Smith TG. Sociodemographic inequalities in barriers to cancer pain management: a report from the American Cancer Society's Study of Cancer Survivors-II (SCS-II). Psychooncology. 2016 Aug 12. doi: 10.1002/pon.4218
- Vallerand AH, Hasenau S, Templin T, Collins-Bohler D. Disparities Between Black and White Patients with Cancer Pain: The Effect of Perception of Control over Pain, Pain Medicine, Volume 6, Issue 3, May 2005, Pages 242–250, https://doi.org/10.1111/j.1526-4637.2005.05038.x

[Response Ends]

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

sp.01. Provide the measure title.

Measure titles should be concise yet convey who and what is being measured (see What Good Looks Like).

[Response Begins]

Patient-Reported Fatigue Following Chemotherapy among Adults with Breast Cancer

[Response Ends]

sp.02. Provide a brief description of the measure.

Including type of score, measure focus, target population, timeframe, (e.g., Percentage of adult patients aged 18-75 years receiving one or more HbA1c tests per year).

[Response Begins]

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). The measure is risk-adjusted.

[Response Ends]

sp.03. Provide a rationale for why this measure must be reported with other measures to appropriately interpret results.

[Response Begins]

The Patient-Reported Symptoms Following Chemotherapy grouped measures assess pain interference, fatigue and overall physical health. As PRO-PMs, these measures were developed as grouped to facilitate implementation; reduce burden for providers and patients; and contribute to interpretation/clinical meaningfulness. A single survey integrates the PROMIS scales that assess pain interference, fatigue and overall quality of life. The PROMIS scales generate specific scores for pain interference, fatigue and overall physical health. The three measures have a common denominator, denominator exclusions, and risk adjustment model, which maximizes use of the clinical and demographic data and thus reduces reporting burden.

[Response Ends]

sp.04. Check all the clinical condition/topic areas that apply to your measure, below.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Surgery: General*

[Response Begins]

Cancer: Breast

[Response Ends]

sp.05. Check all the non-condition specific measure domain areas that apply to your measure, below.

[Response Begins]

Health and Functional Status

Other (specify)

[Other (specify) Please Explain]

Fatigue

Person-and Family-Centered Care: Person-and Family-Centered Care

[Response Ends]

sp.06. Select one or more target population categories.

Select only those target populations which can be stratified in the reporting of the measure's result.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Populations at Risk: Populations at Risk*

[Response Begins]

Adults (Age >= 18)

Women

[Response Ends]

sp.07. Select the levels of analysis that apply to your measure.

Check ONLY the levels of analysis for which the measure is SPECIFIED and TESTED.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Clinician: Clinician*
- *Population: Population*

[Response Begins]

Clinician: Group/Practice

[Response Ends]

sp.08. Indicate the care settings that apply to your measure.

Check ONLY the settings for which the measure is SPECIFIED and TESTED.

[Response Begins]

Ambulatory Care

Outpatient Services

[Response Ends]

sp.09. Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials.

Do not enter a URL linking to a home page or to general information. If no URL is available, indicate “none available".

[Response Begins]

None available

[Response Ends]

sp.12. Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). Excel formats (.xlsx or .csv) are preferred.

Attach an excel or csv file; if this poses an issue, contact staff. Provide descriptors for any codes. Use one file with multiple worksheets, if needed.

[Response Begins]

Available in attached Excel or csv file

[Response Ends]

Attachment: 3720_3741_3720_PROMOnc Data Dictionary_BreastCancer_NQF_Revised-508.xlsx

For the question below: state the outcome being measured. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.13. State the numerator.

Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome).

DO NOT include the rationale for the measure.

[Response Begins]

The PRO-PM numerator is the group-level PROMIS Fatigue score at the follow-up survey.

[Response Ends]

For the question below: describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.14. Provide details needed to calculate the numerator.

All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets.

Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]

The PRO-PM is the risk adjusted group-level mean of PROMIS Fatigue scores among adult women with breast cancer entering survivorship after completion of chemotherapy administered with curative intent. The numerator is calculated as follows:

1. Patient-level PROMIS Fatigue scores captured during the measurement window (baseline and follow-up period) are calculated in accordance with the PROMIS scoring manual (https://www.healthmeasures.net/score-and-interpret/interpret-scores/promis) using the following steps:
    a. Calculate PROMIS raw score for each survey respondent by summing responses to Fatigue Short Form 4a question items

b. Convert raw score for each survey respondent to a T-score using conversion table (see Data Dictionary for conversion table)
c. Calculate the mean of the patient-level T-scores

2. A risk-adjusted mean score at the follow-up survey is calculated for each reporting group.

[Response Ends]

For the question below: state the target population for the outcome. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.15. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]

Adult patients with stages I-III female breast cancer receiving an initial chemotherapy regimen within the measurement window.

[Response Ends]

For the question below: describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.16. Provide details needed to calculate the denominator.

All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets.

Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]

The denominator population includes the following patients:

- >= age 18 on the date of diagnosis, AND
- Stages I-III female breast cancer (ICD-10 C50XX; see Data Dictionary) AND
- Receiving an initial chemotherapy regimen with a defined duration at the test site AND
- Patients with baseline and follow-up PROMIS surveys

Only patients with complete baseline and follow-up surveys are included in the denominator. Reporting sites are required to meet a minimum threshold for survey completion among patients who are eligible for the denominator (minus exclusions); see section sp.29.

Surveys must be administered within the defined measurement window to be included. Upon implementation within a defined measurement window (e.g., 18 months), new patient accrual (defined by administration of the baseline survey) for reporting within that measurement window should end 6 months prior to the end of the measurement window. Note that an 18-month measurement window will include all eligible patients starting chemotherapy over the course of one year. This allows patients in the responding oncology groups to complete the planned chemotherapy regimen and meet time to follow up survey requirements (about 3 months after completion of chemotherapy). For example, for a

measurement window of January 1, 2023 – June 30, 2024, reporting will include patients who completed the baseline survey/started chemotherapy between January 1, 2023 and December 31, 2023.

- Chemotherapy is defined as one or more cytotoxic drugs used in the treatment of cancer. (See Data Dictionary for a list of chemotherapy drugs).
    - All routes of chemotherapy administration are eligible, including oral chemotherapy. Maintenance chemotherapy (i.e., a chemotherapy regimen intended for ongoing treatment and therefore without a defined number of cycles/end date) is not eligible.
    - Immunotherapies, biologics, targeted therapies, HER-2 directed therapies, and/or endocrine therapies are not considered chemotherapy. Patients receiving these therapies should be included only if they are also receiving a chemotherapy drug.
- Chemotherapy must be initiated at the reporting site.
    - Patients who previously received chemotherapy for the breast cancer diagnosis are not eligible.
    - Patients who started the current chemotherapy regimen at another practice/institution, and then continue treatment at the reporting site, are not eligible.
- Chemotherapy may be administered to a patient with any other treatment modality (e.g., surgery, radiation). Chemotherapy may be administered with any treatment sequence. For instance, chemotherapy may be administered prior to surgery (pre-operative or neoadjuvant chemotherapy) or following definitive surgery (adjuvant chemotherapy).

[Response Ends]

sp.17. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]

- Patients on a therapeutic clinical trial
- Patients with recurrence/disease progression
- Patients who leave the practice
- Patients who die

[Response Ends]

sp.18. Provide details needed to calculate the denominator exclusions.

All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]

Denominator exclusions for this PRO-PM are:

- Patients on an interventional or therapeutic clinical trial (excluded at identification)
- Patients who experience relapse or disease progression (excluded during follow-up survey administration period)
- Patients who leave the practice (excluded during the follow-up survey administration period)
- Patients who die (excluded during the follow-up survey administration period)

An interventional or therapeutic trial is one in which patients are prospectively assigned to an intervention, the study evaluates the effect of that intervention, and the effect being evaluated is a biomedical or behavioral outcome. By this

definition, studies that involve secondary research with biological specimens or health information are not interventional or therapeutic clinical trials.

[Response Ends]

sp.19. Provide all information required to stratify the measure results, if necessary.

Include the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate. Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format in the Data Dictionary field.

[Response Begins]

The Fatigue measure scores are used for reporting at the group-level (i.e., not stratified by region or other characteristics).

[Response Ends]

sp.20. Is this measure adjusted for socioeconomic status (SES)?

[Response Begins]

No

[Response Ends]

sp.21. Select the risk adjustment type.

Select type. Provide specifications for risk stratification and/or risk models in the Scientific Acceptability section.

[Response Begins]

Statistical risk model

[Response Ends]

sp.22. Select the most relevant type of score.

Attachment: If available, please provide a sample report.

[Response Begins]

Continuous variable, e.g. average

[Response Ends]

sp.23. Select the appropriate interpretation of the measure score.

Classifies interpretation of score according to whether better quality or resource use is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score

[Response Begins]

Better quality = Lower score

[Response Ends]

sp.24. Diagram or describe the calculation of the measure score as an ordered sequence of steps.

Identify the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period of data, aggregating data; risk adjustment; etc.

[Response Begins]

Survey Completion Steps:

- Patient completes PROMIS baseline survey at initiation of chemotherapy (fatigue scores at baseline)
- Patient completed PROMIS follow-up survey at about 3 months following completion of chemotherapy (fatigue scores at follow up)

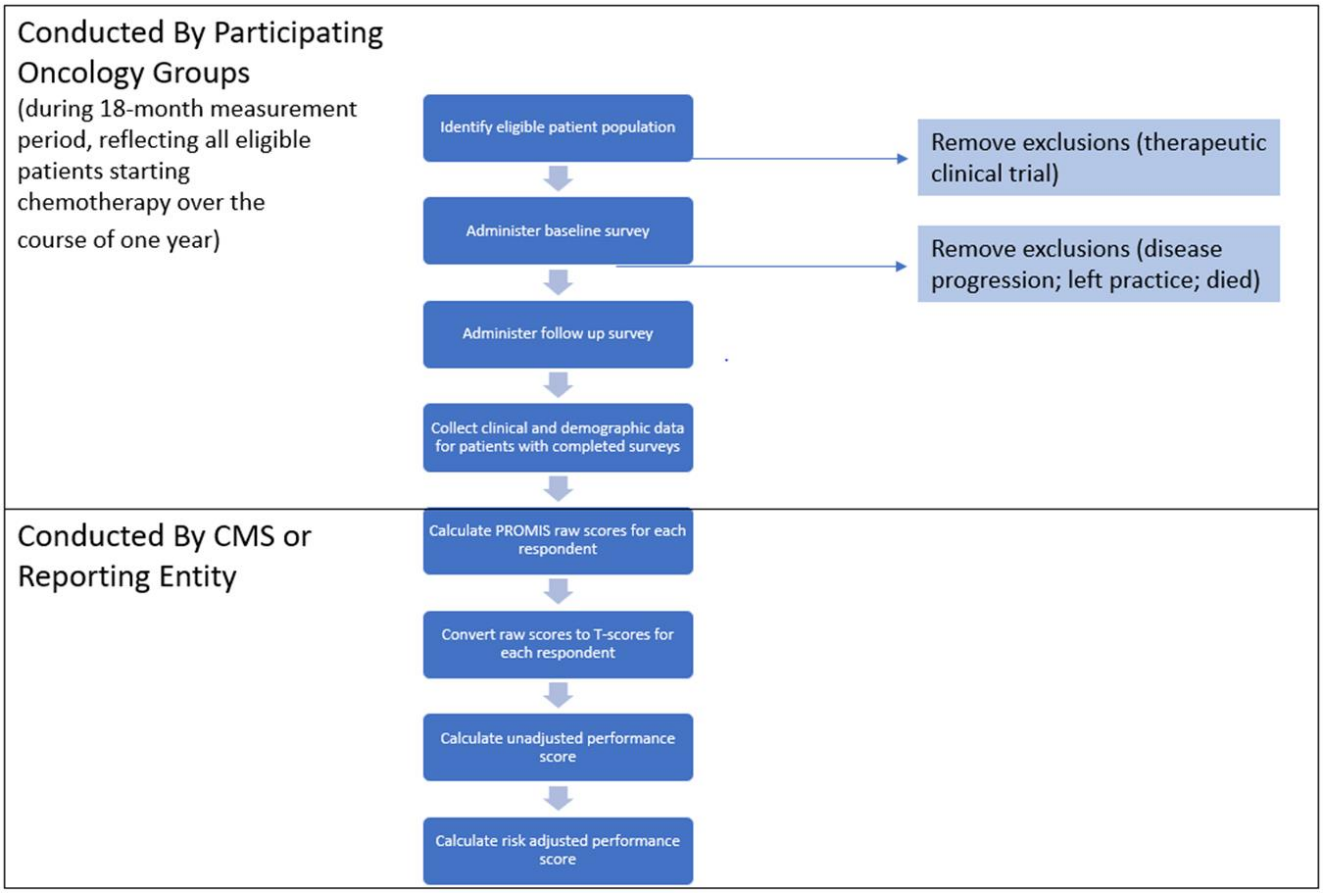
Calculation Logic:

Patient-level PROMIS Fatigue scores are calculated for baseline and follow-up in accordance with the PROMIS scoring manual, and a mean is then calculated for the follow-up survey for each reporting group. Detailed instructions are the following:

1. Calculate PROMIS raw score for each survey respondent by summing responses to Fatigue Short Form 4a question items
2. Convert raw score for each survey respondent to a T-score using conversion table (see Data Dictionary for conversion table)
3. Calculate the mean of the patient-level T-scores

The PRO-PM score is a risk-adjusted average score for each group. The resulting performance measure score will be on a T-score scale. The group is the unit of analysis. Baseline measure scores are included as part of the risk adjustment method.

The PROMOnc measure calculation flow is below. Additional detail (e.g., eligible patient definition) is in the denominator/numerator/exclusion details, and the data dictionary.



[Response Ends]

sp.25. Attach a copy of the instrument (e.g. survey, tool, questionnaire, scale) used as a data source for your measure, if available.

[Response Begins]

Copy of instrument is attached.

[Response Ends]

Attachment: 3720_3741_3720_PROMOnc PROM Instrument-508.pdf

sp.26. Indicate the responder for your instrument.

[Response Begins]

Patient

[Response Ends]

sp.27. If measure testing is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.

Examples of samples used for testing:

- *Testing may be conducted on a sample of the accountable entities (e.g., hospital, physician). The analytic unit specified for the particular measure (e.g., physician, hospital, home health agency) determines the sampling strategy for scientific acceptability testing.*
- *The sample should represent the variety of entities whose performance will be measured. The 2010 Measure Testing Task Force recognized that the samples used for reliability and validity testing often have limited generalizability because measured entities volunteer to participate. Ideally, however, all types of entities whose performance will be measured should be included in reliability and validity testing.*
- *The sample should include adequate numbers of units of measurement and adequate numbers of patients to answer the specific reliability or validity question with the chosen statistical method.*
- *When possible, units of measurement and patients within units should be randomly selected.*

[Response Begins]

No sampling was used.

[Response Ends]

sp.28. Identify whether and how proxy responses are allowed.

[Response Begins]

Responses by family or other caregivers is allowed, which is consistent with PROMIS implementation guidance.

[Response Ends]

sp.29. Survey/Patient-reported data.

Provide instructions for data collection and guidance on minimum response rate. Specify calculation of response rates to be reported with performance measure results.

[Response Begins]

PROMOnc measures use PROMIS (http://www.healthmeasures.net/explore-measurement-systems/promis), specifically the PROMIS Global Health v1.2 (for overall physical health), Pain Interference Short Form 4a (for pain interference) and Fatigue Short Form 4a (for fatigue). With the exception of the pain intensity question, which is a 1-10 scale, the other questions have consistent response options. The PROMOnc survey question items are provided as an attachment. This measure uses the fatigue score from the PROMIS Fatigue Short Form 4a scale. PROMIS PROM scores can be calculated from the PROMIS scoring manual, or obtained via HealthMeasures Scoring Services, powered by Assessment Center.

PROMIS defines validated administration methods for surveys. PROMIS is available in multiple validated translations, which can be selected based on the reporting group's patient characteristics.

Timing of Survey Implementation

Additional details of PROMIS implementation are outlined in detail in the PROMOnc Implementation Guide.

IV Chemotherapy

- Baseline: Survey administered on the first day of chemotherapy administration
  - Allowable window: first day of chemotherapy administration - 2 weeks (14 days) before
- Post-chemotherapy/Follow-Up: Survey administered 3 months after the last chemotherapy administration
  - Allowable window: 3 months after last chemotherapy + 2 months after (90-150 days after last day of chemotherapy)

Oral Chemotherapy

- Baseline: Survey administered on the 1) the oral chemotherapy start date documented in the medical record, or, if that date is missing 2) the date the oral chemotherapy prescription is written
  - Allowable window: oral chemotherapy start date/ prescription date, - 2 weeks (14 days) before and + 1 week (7 days) after
- Post-chemotherapy/Follow-Up: Survey administered 3 months after the oral chemotherapy completion date
  - 3 months after last chemotherapy + 2 months after (90-150 days after last day of chemotherapy)

Sites should attempt to administer the PROMIS instrument to all patients in the target population during the defined measurement window. Consistent with current data completeness criteria for the quality performance category (CMS, 2021), any measured group should obtain survey responses for at least 70% of the target population. In addition, minimum sample size requirements should be met to promote measure reliability – see section 2a.11.

Reference:

- Centers for Medicare & Medicaid Services, "Medicare Program; CY 2022 Payment Policies Under the Physician Fee Schedule and Other Changes to Part B Payment Policies; Medicare Shared Savings Program Requirements; Provider Enrollment Regulation Updates; and Provider and Supplier Prepayment and Post-Payment Medical Review Requirements," *Federal Register*, Vol. 86, No. 221, Washington, D.C.: U.S. Government Printing Office, November 19, 2021c. As of July 26, 2022: https://www.federalregister.gov/documents/2021/11/19/2021-23972/medicare-program-cy-2022-payment-policies-under-the-physician-fee-schedule-and-other-changes-to-part.

[Response Ends]

sp.30. Select only the data sources for which the measure is specified.

[Response Begins]

Electronic Health Records

Instrument-Based Data

Paper Medical Records

[Response Ends]

sp.31. Identify the specific data source or data collection instrument.

For example, provide the name of the database, clinical registry, collection instrument, etc., and describe how data are collected.

[Response Begins]

This measure is calculated based on data from PROMIS: Fatigue Short Form 4a. The measure also requires clinical and demographic risk adjustment variables which are derived from oncology medical records.

[Response Ends]

sp.32. Provide the data collection instrument.

[Response Begins]

Available in attached appendix in Question 1 of the Additional Section

[Response Ends]

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate fields in the Scientific Acceptability sections of the Measure Submission Form.

- Measures must be tested for all the data sources and levels of analyses that are specified. If there is more than one set of data specifications or more than one level of analysis, contact NQF staff about how to present all the testing information in one form.
- All required sections must be completed.
- For composites with outcome and resource use measures, Questions 2b.23-2b.37 (Risk Adjustment) also must be completed.
- If specified for multiple data sources/sets of specifications (e.g., claims and EHRs), Questions 2b.11-2b.13 also must be completed.
- An appendix for supplemental materials may be submitted (see Question 1 in the Additional section), but there is no guarantee it will be reviewed.
- Contact NQF staff with any questions. Check for resources at the Submitting Standards webpage.
- For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for the 2021 Measure Evaluation Criteria and Guidance.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a. Reliability testing demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.

2b1. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure;

AND

If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

2b3. For outcome measures and other measures when indicated (e.g., resource use):

- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; 14,15 and has demonstrated adequate discrimination and calibration
- rationale/data support no risk adjustment/ stratification.

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful 16 differences in performance;

OR

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

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

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

2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:

2c1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and

2c2. the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.

(if not conducted or results not adequate, justification must be submitted and accepted)

Definitions

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

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

Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

Risk factors that influence outcomes should not be specified as exclusions.

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

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:

Current Submission:

Updated testing information here.

Previous (Year) Submission:

Testing from the previous submission here.

2a.01. Select only the data sources for which the measure is tested.

[Response Begins]

Electronic Health Records

Instrument-Based Data

Paper Medical Records

[Response Ends]

2a.02. If an existing dataset was used, identify the specific dataset.

The dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry).

[Response Begins]

N/A

[Response Ends]

2a.03. Provide the dates of the data used in testing.

Use the following format: "MM-DD-YYYY - MM-DD-YYYY"

[Response Begins]

07/01/2019 - 04/01/2022

[Response Ends]

2a.04. Select the levels of analysis for which the measure is tested.

Testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Clinician: Clinician*
- *Population: Population*

[Response Begins]

Clinician: Group/Practice

[Response Ends]

2a.05. List the measured entities included in the testing and analysis (by level of analysis and data source).

Identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample.

[Response Begins]

Group practices: Testing was planned on a sample of 21 oncology groups. Due to the impact of the COVID 19 public health emergency during the testing period, however, only 10 sites submitted sufficient data for inclusion in testing analyses. The group practices included in testing are described below, and reflect geographic, size, and practice type variation.

- City of Hope Comprehensive Cancer Center is an academic practice site in Duarte, CA. City of Hope has about 300,000 oncology outpatient visits per year.
- Henry Ford Macomb is a community hospital site of an academic health system in Clinton Township, MI. Henry Ford Macomb sees about 1600 new oncology patients per year.
- The James Cancer Hospital is an academic practice site in Columbus, OH. The James has about 750,000 oncology outpatient visits per year.
- Karmanos Cancer Institute at McLaren-Macomb is a community hospital site of an academic health system in Mount Clemens, MI. The practice sees about 1600 new oncology patients per year.
- MD Anderson Cancer Center is an academic practice site in Houston, TX. MD Anderson has approximately 1.5 million oncology outpatient visits per year.
- Memorial Sloan Kettering Cancer Center is an academic practice site in New York, NY. Memorial Sloan Kettering has about 800,000 oncology outpatient visits per year.
- Munson Cancer Center is a community hospital in Traverse City, MI. Munson sees about 4000 new oncology patients per year.
- Roswell Park Cancer Institute is an academic practice site in Buffalo, NY. Roswell Park has about 270,000 oncology outpatient visits per year.
- The Seattle Cancer Care Alliance/Fred Hutchinson Cancer Center is an academic practice in Seattle, WA. SCCA has approximately 90,000 oncology outpatient visits per year.
- USC Norris Comprehensive Cancer Center is an academic practice site in Los Angeles, CA. USC Norris has about 140,000 oncology outpatient visits per year.

[Response Ends]

2a.06. Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis), separated by level of analysis and data source; if a sample was used, describe how patients were selected for inclusion in the sample.

If there is a minimum case count used for testing, that minimum must be reflected in the specifications.

[Response Begins]

Table 2a.1 and 2a.2 show descriptive characteristics of the 323 patients who completed the baseline and follow-up surveys.

Table 2a.1 Demographic Characteristics of Patients Completing Baseline and Follow-up Surveys

Characteristic	**Category**	**N**	**%**
Marital Status	Unmarried	79	24.46
Marital Status	Married	233	72.14
Marital Status	Undisclosed	11	3.41
Race	Hispanic	25	7.74
Race	Non-Hispanic White	216	66.87
Race	Non-Hispanic Black	34	10.53
Race	Non-Hispanic Asian	25	7.74
Race	Other	23	7.12
Insurance	Missing	7	2.17
Insurance	Private Insurance	166	51.39
Insurance	Medicare	35	10.84
Insurance	Medicaid or Dual Eligible	11	3.41
Insurance	Self-Pay or Uninsured	29	8.98
Insurance	Combination Private and Medicare	20	6.19
Insurance	Other	55	17.03
Smoking Status	Undocumented	8	2.48
Smoking Status	Never Smoker	218	67.49
Smoking Status	Former Smoker	74	22.91
Smoking Status	Current Smoker	23	7.12
*	*	**N**	**M(SD)**
Age at Diagnosis		323	54.60(11.67)
Body Mass Index (BMI)		318	29.98(7.16)

*Cell intentionally left blank

Table 2a.2 Clinical Characteristics of Patients Completing Baseline and Follow-up Surveys

Characteristic	**Category**	**N**	**%**
AJCC Breast Cancer Pathologic/Clinical Stage	Missing	14	4.33

Characteristic	Category	N	%
AJCC Breast Cancer Pathologic/Clinical Stage	Stage I	5	1.55
AJCC Breast Cancer Pathologic/Clinical Stage	Stage IA	77	23.84
AJCC Breast Cancer Pathologic/Clinical Stage	Stage IA2	0	0.00
AJCC Breast Cancer Pathologic/Clinical Stage	Stage IB	50	15.48
AJCC Breast Cancer Pathologic/Clinical Stage	Stage II	2	0.62
AJCC Breast Cancer Pathologic/Clinical Stage	Stage IIA	72	22.29
AJCC Breast Cancer Pathologic/Clinical Stage	Stage IIB	45	13.93
AJCC Breast Cancer Pathologic/Clinical Stage	Stage III	5	1.55
AJCC Breast Cancer Pathologic/Clinical Stage	Stage IIIA	19	5.88
AJCC Breast Cancer Pathologic/Clinical Stage	Stage IIIB	22	6.81
AJCC Breast Cancer Pathologic/Clinical Stage	Stage IIIC	12	3.72
Estrogen Receptor Status	Missing	9	2.79
Estrogen Receptor Status	Positive	216	66.87
Estrogen Receptor Status	Negative	98	30.34
Progesterone Receptor Status	Missing	23	7.12
Progesterone Receptor Status	Positive	180	55.73
Progesterone Receptor Status	Negative	120	37.15
HER2 Receptor Status	Missing	27	8.36
HER2 Receptor Status	Positive	82	25.39
HER2 Receptor Status	Negative	211	65.33
HER2 Receptor Status	Equivocal	3	0.93
Performance Status at Baseline	Missing	46	14.24
Performance Status at Baseline	Normal activity level	241	74.61
Performance Status at Baseline	Symptomatic and ambulatory; cares for self	33	10.22
Performance Status at Baseline	Ambulatory > 50% of time; occasional assistance	3	0.93

Characteristic	Category	N	%
Performance Status at Baseline	Ambulatory </= 50% of time; nursing care needed	0	0.00
Chemotherapy Regimen	Missing	0	0.00
Chemotherapy Regimen	Dose-Dense AC (doxorubicin/cyclophosphamide) followed by paclitaxel	119	36.84
Chemotherapy Regimen	TC (docetaxel and cyclophosphamide)	56	17.34
Chemotherapy Regimen	Dose-Dense AC	13	4.02
Chemotherapy Regimen	AC Every 3 Weeks	2	0.62
Chemotherapy Regimen	CMF (cyclophosphamide/methotrexate/fluorouracil)	18	5.57
Chemotherapy Regimen	AC followed by paclitaxel	7	2.17
Chemotherapy Regimen	AC followed by docetaxel	1	0.31
Chemotherapy Regimen	EC (epiribicin/cyclophosphamide)	0	0.00
Chemotherapy Regimen	TAC (docetaxel/doxorubicin/cyclophosphamide)	0	0.00
Chemotherapy Regimen	AC followed by T (paclitaxel) + trastuzumab	2	0.62
Chemotherapy Regimen	AC followed by T + trastuzumab + pertuzumab	3	0.93
Chemotherapy Regimen	Paclitaxel + trastuzumab	25	7.74
Chemotherapy Regimen	TCH (docetaxel/carboplatin/trastuzumab)	4	1.24
Chemotherapy Regimen	TCH + pertuzumab	40	12.38
Chemotherapy Regimen	Docetaxel + cyclophosphamide + trastuzumab	1	0.31
Chemotherapy Regimen	AC followed by docetaxel + trastuzumab	0	0.00
Chemotherapy Regimen	AC followed by docetaxel + trastuzumab + pertuzumab	1	0.31
Chemotherapy Regimen	Other	31	9.60
Neoadjuvant or Adjuvant Chemotherapy	Missing	0	0.00
Neoadjuvant or Adjuvant Chemotherapy	Neoadjuvant	166	51.39
Neoadjuvant or Adjuvant Chemotherapy	Adjuvant	157	48.61
Aromatase Inhibitor	Missing	39	12.07
Aromatase Inhibitor	No	175	54.18
Aromatase Inhibitor	Yes, administered	83	25.70
Aromatase Inhibitor	Yes, planned	26	8.05
SERM (e.g., Tamoxifen, Raloxifene, Toremifene)	Missing	52	16.10

Characteristic	Category	N	%
SERM (e.g., Tamoxifen, Raloxifene, Toremifene)	No	249	77.09
SERM (e.g., Tamoxifen, Raloxifene, Toremifene)	Yes, administered	19	5.88
SERM (e.g., Tamoxifen, Raloxifene, Toremifene)	Yes, planned	3	0.93
LHRH Agonists	Missing	49	15.17
LHRH Agonists	No	250	77.40
LHRH Agonists	Yes, administered	21	6.50
LHRH Agonists	Yes, planned	3	0.93
Trastuzumab (Herceptin)	Missing	69	21.36
Trastuzumab (Herceptin)	No	236	73.07
Trastuzumab (Herceptin)	Yes, administered	14	4.33
Trastuzumab (Herceptin)	Yes, planned	4	1.24
Pertuzumab (Perjeta)	Missing	52	16.10
Pertuzumab (Perjeta)	No	228	70.59
Pertuzumab (Perjeta)	Yes, administered	42	13.00
Pertuzumab (Perjeta)	Yes, planned	1	0.31
Neratinib (Nerlynx)	Missing	70	21.67
Neratinib (Nerlynx)	No	253	78.33
Neratinib (Nerlynx)	Yes, administered	0	0.00
Neratinib (Nerlynx)	Yes, planned	0	0.00
Other Cancer Directed Therapy	Missing	41	12.69
Other Cancer Directed Therapy	No	207	64.09
Other Cancer Directed Therapy	Yes, administered	72	22.29
Other Cancer Directed Therapy	Yes, planned	3	0.93
Breast Cancer Surgery Received	Missing	5	1.55
Breast Cancer Surgery Received	No	24	7.43
Breast Cancer Surgery Received	Yes	294	91.02
Radiation Therapy Received	Missing	35	10.84
Radiation Therapy Received	No, Radiation	106	32.82
Radiation Therapy Received	Yes, Radiation	182	56.35

Multiple comorbidities were collected based on a modified Elixhauser comorbidity tool. Among those comorbidities, diabetes, hypertension, and depression had responses sufficient for analyses. 4.95% (N=16) of the patients above had a

reported diabetic comorbidity, 13.62% (N=44) had a hypertension comorbidity and 2.48% (N=8) had a depression comorbidity.

[Response Ends]

2a.07. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing.

[Response Begins]

The analysis of the validity of data elements used the collected dataset (see the Data Dictionary) and Cancer Registry data. All other analyses were conducted with the same collected dataset.

[Response Ends]

2a.08. List the social risk factors that were available and analyzed.

For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

[Response Begins]

Social risk factors in the data include insurance Dual eligibility for Medicare and Medicaid, for which the distribution is presented in Table 2a.3.

Table 2a.3. Social Risk Characteristics of Respondents

Characteristic	Category	N	%
Insurance	Missing	7	2.17
Insurance	Private Insurance	166	51.39
Insurance	Medicare	35	10.84
Insurance	Medicaid or Dual Eligible	11	3.41
Insurance	Self-Pay or Uninsured	29	8.98
Insurance	Combination Private and Medicare	20	6.19
Insurance	Other	55	17.03

[Response Ends]

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.09 check patient or encounter-level data; in 2a.010 enter “see validity testing section of data elements”; and enter “N/A” for 2a.11 and 2a.12.

2a.09. Select the level of reliability testing conducted.

Choose one or both levels.

[Response Begins]

Patient or Encounter-Level (e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements)

Accountable Entity Level (e.g., signal-to-noise analysis)

[Response Ends]

2a.10. For each level of reliability testing checked above, describe the method of reliability testing and what it tests.

Describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used.

[Response Begins]

PROMIS Internal Consistency

PROMIS measures, including the fatigue scale, have undergone rigorous development and validation for use in both a general population and in individuals with chronic conditions. The original psychometric testing of PROMIS fatigue scales included a broad range of diseases including cancer and reported internal consistency reliability using Cronbach's alpha coefficients (Cook et al. 2016; Stone et al. 2016).

References:

- Cook KF, Jensen SE, Schalet BD, Beaumont JL, Amtmann D, Czajkowski S, Dewalt DA, Fries JF, Pilkonis PA, Reeve BB, Stone AA, Weinfurt KP, Cella D. PROMIS Measures of Pain, Fatigue, Negative Affect, Physical Function, and Social Function Demonstrated Clinical Validity Across a Range of Chronic Conditions. J Clin Epidemiol. 2016 May;73:89-102. doi: 10.1016/j.jclinepi.2015.08.038. Epub 2016 Mar 4. PMID: 26952842
- Stone AA, Broderick JE, Junghaenel DU, Schneider S, Schwartz JE. PROMIS fatigue, pain intensity, pain interference, pain behavior, physical function, depression, anxiety, and anger scales demonstrate ecological validity. J Clin Epidemiol. 2016 Jun;74:194-206. doi: 10.1016/j.jclinepi.2015.08.029. Epub 2015 Nov 25.

PRO-PM Reliability

To test the reliability of the performance measure, we used a traditional "signal-to-noise" analysis that decomposes variability in the measure score into a) between-subject variability and b) within-subject variability. If there is a large amount of between-subject variability (i.e., "signal") compared to within-subject variability (i.e., "noise"), then there is more evidence that it is possible to discriminate performance among groups.

To evaluate quality measure reliability for group-level reporting, we used hierarchical linear regression models to relate our outcome measures to our providers and their covariates, where the hierarchy of data is patient observations within groups. The variance of the model can be decomposed using the adjusted intraclass correlation coefficient (ICC), which provides a summary of the reliability of the measure as tested, with higher values implying more variability between groups. Additionally, we incorporate risk adjustment variables into our models to provide fair comparisons among groups and to provide a best effort to ensure that the observed differences among groups are truly from differences in performance and not due to baseline differences in risk variables that represent the groups. The reliability from the measure test is then projected out based on observed variances and sample sizes from each group, using the Spearman-Brown prophecy formula. This allows us to estimate the required within-group sample size to achieve a desired reliability for the measure. Reliability values of approximately 0.7 were a target of an acceptable level of reliability and helped determine required sample sizes (Nunnally & Bernstein, 1994), and are recommended in the NQF-commissioned paper on PRO-PMs (NQF, 2013).

References:

- Nunnally JC & Bernstein IH. Psychometric Theory. New York: McGraw Hill; 1994.
- National Quality Forum (NQF). Patient Reported Outcomes (PROs) in Performance Measurement. January 10, 2013.

[Response Ends]

2a.11. For each level of reliability testing checked above, what were the statistical results from reliability testing?

For example, provide the percent agreement and kappa for the critical data elements, or distribution of reliability statistics from a signal-to-noise analysis. For score-level reliability testing, when using a signal-to-noise analysis, more than just one overall statistic should be reported (i.e., to demonstrate variation in reliability across providers). If a particular method yields only one statistic, this should be explained. In addition, reporting of results stratified by sample size is preferred (pg. 18, NQF Measure Evaluation Criteria).

[Response Begins]

Results of PROMIS Internal Consistency Reliability Testing

Reliability testing results of the PROMIS instrument are reported in the literature. For PROMIS Fatigue, the Cronbach's alpha coefficient, which measures internal consistency reliability, is .86.

See 2a.10 for PROMIS references.

Results of Group-Level Reliability Testing

The estimate of the adjusted ICC is 0.094 and the estimate of the reliability at the average sample size for a group (32 patients per group) is 0.77. We then extend our reliability results to future samples using the Spearman-Brown prophecy formula, which estimates the average number of patient respondents within groups to achieve a desired reliability for a given ICC. We estimate that in order to obtain a nominal reliability of 0.7, a minimum sample size of 23 patient respondents would be required. Group specific reliability ranges from 0.38 to 0.88, with a mean of 0.66 (SD=0.21), and a median reliability of 0.68. We assessed the proportion of groups in our sample that have sufficient reliability, using a reliability threshold of 0.70; 50% of groups have reliability that is .70 or greater.

[Response Ends]

2a.12. Interpret the results, in terms of how they demonstrate reliability.

(In other words, what do the results mean and what are the norms for the test conducted?)

[Response Begins]

Interpretation of PROMIS Internal Consistency Reliability Testing

The reliability testing results of the PROMIS instrument reported in the literature demonstrate alpha values of 0.70 or greater, which is an acceptable minimum for group-level assessment.

See 2a.10 for PROMIS references.

Interpretation of Group-Level Reliability Testing

The measure exhibits acceptable group-level reliability of 0.70 or greater at the average number of completed surveys per group.

[Response Ends]

2b. Validity

2b.01. Select the level of validity testing that was conducted.

[Response Begins]

Patient or Encounter-Level (data element validity must address ALL critical data elements)

Systematic assessment of face validity of performance measure score as an indicator of quality or resource use (i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance)

[Response Ends]

2b.02. For each level of testing checked above, describe the method of validity testing and what it tests.

Describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used.

[Response Begins]

Patient-Level Data Element Validity

For patient-level data element validity, cancer registries provide gold standard data for interdisciplinary cancer care (https://www.cdc.gov/cancer/npcr/value/registries.htm). However, not all PROMOnc data elements are available in the cancer registry and no other feasible gold standard validation sources could be identified. In November 2020, PROMOnc test sites were asked to provide cancer registry data for PROMOnc eligible patients for critical data elements used to identify denominator, denominator exclusion, and risk adjustment variables. Seven test sites with cancer registries submitted data for 570 PROMOnc eligible patients.

The majority of the PROMOnc clinical and demographic variables were validated; however, certain variables could not be validated. Elements included and excluded from testing are described below.

- Elements determining patient eligibility/denominator: all included in validity testing.
- Elements in denominator exclusions: death and cancer recurrence included in validity testing. Clinical trial enrollment is not captured in the cancer registry. Patients leaving the practice administering chemotherapy is not captured in the cancer registry.
- Elements determining numerator: PRO scores only; no clinical or demographic data.
- Elements included in risk adjustment model: Elements to calculate derived variables associated with time since diagnosis; receipt of radiation and timing; receipt of surgery, type of surgery, and surgical timing; and receipt of an aromatase inhibitor were evaluated as described above. BMI, comorbidities, smoking status and performance status were not evaluated as they are not routinely captured by the cancer registry.

Among the PROMOnc clinical and demographic variables that were validated, we computed percentage of exact agreement for all data elements, Kappa coefficient for cancer stages (I, II, III, and IV) that are on an ordinal scale, and sensitivity and specificity for data elements that are dichotomous.

Face Validity

Face validity of the quality measure scores was determined through a systematic and transparent process by convening experts who explicitly addressed whether scores resulting from the measure, as specified, can be used to distinguish good from poor quality. In May 2022, following completion of testing, a panel of 12 oncologist advisors were asked to review the final measure specifications and testing results and rate face validity of the measure score. Advisors were asked to respond to the question “Rate your agreement with the following statement: The scores obtained from the measure as specified will provide an accurate reflection of quality and can be used to distinguish good and poor quality.” with response categories: 1=Disagree, 2, 3=Moderate Agreement, 4, 5=Agree. (Scale adapted from NQF’s ‘What Good Looks Like’ example for Validity Testing (Systematic Assessment of Face Validity) in which the rating scale had five levels with the following narrative anchors (with no anchors for 2 and 4): 1=Disagree, 3=Moderate Agreement and 5=Agree.)

[Response Ends]

2b.03. Provide the statistical results from validity testing.

Examples may include correlations or t-test results.

[Response Begins]

Patient-Level Data Element Validity

Table 2b.1 summarizes data element validity between submitted PROMOnc data and cancer registry data for breast cancer patients.

Table 2b.1: Data Element Validity Among Patients with Data in PROMOnc and Cancer Registry Datasets

Purpose of the PROMOnc Data Element	Data Element	Number of Patients	Agreement Index	Sensitivity	Specificity
Identify Patients in Denominator	Date of birth	570	Percentage of exact agreement: 100%	*	*
Identify Patients in Denominator	Gender	570	Percentage of exact agreement: 100%	N/A; All patients are female in both datasets	N/A; All patients are female in both datasets
Identify Patients in Denominator	Breast cancer diagnosis	570	Percentage of exact agreement: 100%	N/A; All patients are patients with breast cancer in both datasets	N/A; All patients are patients with breast cancer in both datasets
Identify Patients in Denominator	Breast cancer pathologic stage	243	Percentage of exact agreement: 80.25% Kappa: 0.64	*	*
Identify Patients in Denominator	Breast cancer clinical stage	141	Percentage of exact agreement: 71.63% Kappa: 0.67	*	*
Denominator Exclusions	Death	541	99.45%	N/A; All patients in the registry dataset are alive	99.45%
Denominator Exclusions	Recurrence	503	99.01%	33.33%	99.80%
Risk Adjustment Variable: Patient Age	Date of birth	570	100%	*	*

Purpose of the PROMOnc Data Element	Data Element	Number of Patients	Agreement Index	Sensitivity	Specificity
Data Related to Derived Risk Adjustment Variable: Number of Days Between Diagnosis Date and The Date of Follow-Up Survey	Diagnosis date (within 14 days**)	569	79.61%	*	*
Data Related to Derived Risk Adjustment Variable: Radiation Within Two Weeks Before the Date of Follow-Up Survey	Radiation administered	319	84.64%	89.52%	75.23%
Data Related to Derived Risk Adjustment Variable: Radiation Within Two Weeks Before the Date of Follow-Up Survey	Start date of radiation	189	96.83%	*	*
Data Related to Derived Risk Adjustment Variable: Radiation Within Two Weeks Before the Date of Follow-Up Survey	Ending date of radiation	180	93.89%	*	*
Data Related to Derived Risk Adjustment Variable: Surgery Severity Level	Surgery received	530	82.83%	89.16%	60.00%
Data Related to Derived Risk Adjustment Variable: Surgery Severity Level	Surgery type	410	91.50%	*	*
Data Related to Derived Risk Adjustment Variable; Number of Days Between the Latest Surgery and the Date of Follow-Up Survey	Surgery date (within 24 hours**)	393	91.90%	*	*
Risk Adjustment Variable: Aromatase Inhibitor	AI administered ***	154	75.97%	N/A	N/A

*Cell intentionally left blank

**Date precision for date of diagnosis allows for slight differences in diagnosis date definition for PROMOnc vs the cancer registry (due to feasibility challenges with the latter); date precision for dates of treatment allows for reasonable variation (e.g., radiation treatment planning vs first administration; date of surgery vs date of discharge).

***Sensitivity and specificity cannot be evaluated for AI administered as the registry data includes other hormonal therapies in a single variable. Percent agreement indicates presence of hormonal therapy in the cancer registry data, which could include therapies other than aromatase inhibitors.

Face Validity

Advisors were asked to respond to the question "Rate your agreement with the following statement: The scores obtained from the measure as specified will provide an accurate reflection of quality and can be used to distinguish good and poor quality." with response categories: 1=Disagree, 2, 3=Moderate Agreement, 4, 5=Agree. Eight responded to the face validity survey, with eight indicating that they "moderately agree" to "agree" (e.g., rated a 3, 4 or 5) that the measure can differentiate good from poor quality care among accountable entities. If we remove the 5 ratings of Moderate Agreement (e.g., rated a 3), 3 of 8 agreed that the Fatigue measure could differentiate good versus poor quality; the participants that did not rate the measure as a 4 or 5 felt that fatigue was more susceptible to pandemic related issues. Four oncologists declined to participate in face validity voting for the measures; these oncologists expressed concerns regarding the impact of COVID-19 on sample size and potentially performance scores. They requested additional testing data and thus more patients included in the testing analysis prior to voting.

[Response Ends]

2b.04. Provide your interpretation of the results in terms of demonstrating validity. (i.e., what do the results mean and what are the norms for the test conducted?)

[Response Begins]

Patient-Level Data Element Validity

Data element validity is supported with percentages of exact agreement range from 71.63% to 100%. Items for which sensitivity and specificity could be analyzed demonstrated acceptable specificity. Sensitivity was low for recurrence variable but reflected data from only 6 patients who were identified as having recurrence in the registry data; a low rate of recurrence is expected in this population (note that no recurrence exclusions were captured in the final testing cohort, see 2b.16).

Face Validity

These face validity ratings provided by 8 expert advisors in oncology and quality measurement reflect support for face validity of the proposed quality measure.

[Response Ends]

2b.05. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified.

Describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided in Importance to Measure and Report: Gap in Care/Disparities.

[Response Begins]

To examine the ability of the measure to identify high or low performing groups, we calculated the number and percentage of groups that were significantly above or below the average score. All scores were risk adjusted. A two-sided alpha=0.05 level test was used to test for significance.

[Response Ends]

2b.06. Describe the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities.

Examples may include number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined.

[Response Begins]

Results indicate a mean group performance score of 48.51 and the standard deviation is 3.13, median score is 48.67, with a range of 42.13 to 53.07. Two out of 10 groups have significantly different scores than the overall average, one more favorable and the other less favorable. Among group scores that were significantly above or below the average, the mean absolute difference between the group's scores and the overall average for was 4.9 points on a T-score scale (SD=10).

[Response Ends]

2b.07. Provide your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities.

In other words, what do the results mean in terms of statistical and meaningful differences?

[Response Begins]

Testing in the cancer population indicates meaningful variation with at least a 3-point difference on a T-score scale that has a mean of 50 and standard deviation of 10 (Jensen et al., 2017). Among group scores that were significantly above or below the average, the mean absolute difference between the group's scores and the overall average was 4.9 points. Results indicate that the PRO-PM measure can discriminate between groups' performance.

Reference:

- Jensen RE, Moinpour CM, Potosky AL, Lobo T, Hahn EA, Hays RD, Cella D, Smith AW, Wu XC, Keegan TH, Paddock LE, Stroup AM, Eton DT. Responsiveness of 8 Patient-Reported Outcomes Measurement Information System (PROMIS) Measures in Large, Community-Based Cancer Study Cohort. Cancer. 2017 Jan 1;123(2):327-335. doi: 10.1002/cncr.30354. Epub 2016 Oct 3. PMID: 27696377

[Response Ends]

2b.08. Describe the method of testing conducted to identify the extent and distribution of missing data (or non-response) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders). Include how the specified handling of missing data minimizes bias.

Describe the steps—do not just name a method; what statistical analysis was used.

[Response Begins]

Survey Non-Response

The Patient-Reported Symptoms Following Chemotherapy grouped measures assess pain interference, fatigue and overall physical health. PRO data were collected with a single survey containing 18 items that integrates the PROMIS scales that assess pain interference, fatigue and overall quality of life.

The survey administration rate was calculated as follows:

Administration Rate = (Total Number of Follow-up Surveys Fielded) / (Total Number of Patients in the Target Population – Total Number of Patients Meeting the Denominator Exclusion Criteria)

The survey response rate was calculated as follows:

Response Rate = (Total Number of Completed Surveys) / (Total Number of Follow-up Surveys Fielded – Total Number of Ineligible Surveys)

The Total Number of Completed Surveys is the total number of surveys for which the respondent answers at least 50 percent (9 items in the follow-up survey), which is a threshold commonly used in patient-reported survey measures, of the questions. Total Number of Ineligible Surveys is the total number of surveys for which it is determined that the

patient met the denominator exclusion criteria outlined above in Section Sp.17 (e.g., on a therapeutic clinical trial, left the practice, disease progressed, or deceased) plus those that have a language barrier or who had mental/physical incapacity. The following are not removed from the denominator of the response rate calculation: break-off surveys, refusals, non-response.

We assessed the association between survey nonresponse and several patient characteristics, including demographic characteristics (ethnicity and race, age, marital status, insurance), baseline clinical factors (smoking, BMI, performance status, pathology and clinical stage, receptor status, comorbidities), cancer treatment (surgery severity level, with or without radiation, chemo regimen), and baseline measure scores (pain interference, fatigue, physical and mental health at baseline before the start of chemotherapy).

Item Non-Response

In 2b.09, we present nonresponse to evaluative items among respondents. Specifically, we report the total proportion of missing data for each evaluative item on the follow-up survey.

[Response Ends]

2b.09. Provide the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data.

For example, provide results of sensitivity analysis of the effect of various rules for missing data/non-response. If no empirical sensitivity analysis was conducted, identify the approaches for handling missing data that were considered and benefits and drawbacks of each).

[Response Begins]

Survey Non-Response

Across ten sites, 896 patients in the target population were eligible for the follow-up survey, and 19 patients met the denominator exclusion criteria outlined above in Section Sp.17 (18 patients on a therapeutic clinical trial and 1 died). The total number of follow up surveys fielded was 744. The survey administration rate is calculated as 744/(896-19) = 84.8%.

Among the 744 follow-up surveys fielded, there were 323 completed surveys, and 9 ineligible surveys. The response rate is calculated as 323/(744-9)=43.95%.

We compared patients in the target population, excluding patients meeting the denominator exclusion criteria, who completed the follow-up survey (n=323) and those who did not (n=554) on patient characteristics stated in Section 2b.08. No statistical significance was identified except that the respondents and nonrespondents differed on marital status and insurance.

The portion of patients who are married or with a partner was higher among respondents, compared to nonrespondents (72.14% vs 63.45%); the results from a chi-squared test indicates that this difference is statistically significant (p = 0.03). Respondents were more likely to have a combination of private and Medicare insurance (6.33% for respondents vs 1.47% for non-respondents), but less likely to have Medicaid (3.48% vs 11.95%), and such differences are significant at p<.001.

Item Non-Response

Table 2b.2. Item Missingness, PROMIS Fatigue Scale

Item of Fatigue	% Missing
During the past 7 days… I feel fatigued.	0.31
During the past 7 days… I have trouble starting things because I am tired.	0.93
In the past 7 days… How run-down did you feel on average?	0.62

Item of Fatigue	% Missing
In the past 7 days… How fatigued were you on average?	0.00

[Response Ends]

2b.10. Provide your interpretation of the results, in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders), and how the specified handling of missing data minimizes bias.

In other words, what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis was conducted, justify the selected approach for missing data.

[Response Begins]

The survey administration rate exceeds the data completeness criteria for the quality performance category of 70%. The survey response rate is higher than response rate in similar studies. Although our analyses indicate that response propensity varies by certain patient characteristics, previous work in patient experience of care surveys has demonstrated that nonresponse weighting to account for potential bias is not needed after case-mix adjustment (see, for example, Elliott, Edwards et al. 2005 and Elliott, Zaslavsky et al. 2009).

Across evaluative items, less than 2 percent of respondents missed at least one item. This finding suggests that it is unlikely that item results are biased due to systematic skipping of items by respondents.

References:

- Elliott MN, Edwards C, Angeles J, Hays RD (2005). "Patterns of unit and item non-response in the CAHPS® Hospital Survey." *Hlth Serv Res* 40(6): 2096-2119.
- Elliott MN, Zaslavsky AM, Goldstein E, Lehrman W, Hambarsoomian K, Beckett MK, Giordano L (2009). "Effects of survey mode, patient mix, and nonresponse on CAHPS Hospital Survey scores." *Hlth Serv Res* 44(2): 501-508.

[Response Ends]

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

2b.11. Indicate whether there is more than one set of specifications for this measure.

[Response Begins]

No, there is only one set of specifications for this measure

[Response Ends]

2b.12. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications.

Describe the steps—do not just name a method. Indicate what statistical analysis was used.

[Response Begins]

[Response Ends]

2b.13. Provide the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications.

Examples may include correlation, and/or rank order.

[Response Begins]

[Response Ends]

2b.14. Provide your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications.

In other words, what do the results mean and what are the norms for the test conducted.

[Response Begins]

[Response Ends]

2b.15. Indicate whether the measure uses exclusions.

[Response Begins]

Yes, the measure uses exclusions.

[Response Ends]

2b.16. Describe the method of testing exclusions and what was tested.

Describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used?

[Response Begins]

We finalized four exclusions from the measure denominator. Those exclusions and their frequencies obtained from testing are:

- Patients on an interventional or therapeutic clinical trial (n=18)
- Patients who experience relapse or disease progression (n=0)
- Patients who leave the practice (n=0)
- Patients who die (n=1)

We could not analyze the impact on measure outcomes of excluding these patients because follow-up survey data was not available for these patients.

Inclusion in the PROMOnc denominator requires patient completion of PROMIS baseline and follow-up surveys. As described in detail in 2b.09, we compared patients who completed the follow-up survey (n=323) and those who did not (n=554) on patient characteristics stated in Section 2b.08. No statistical significance was identified except that the respondents and nonrespondents differed on marital status and insurance.

[Response Ends]

2b.17. Provide the statistical results from testing exclusions.

Include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores.

[Response Begins]

N/A

[Response Ends]

2b.18. Provide your interpretation of the results, in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results.

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

[Response Begins]

N/A

[Response Ends]

2b.19. Check all methods used to address risk factors.

[Response Begins]

Statistical risk model with risk factors (specify number of risk factors)

[Statistical risk model with risk factors (specify number of risk factors) Please Explain]

All three measures (including Pain Interference, Fatigue, Physical Health) are risk adjusted for 13 risk adjustment variables, which are listed in 2b.20.

[Response Ends]

2b.20. If using statistical risk models, provide detailed risk model specifications, including the risk model method, risk factors, risk factor data sources, coefficients, equations, codes with descriptors, and definitions.

[Response Begins]

To ensure that comparisons between groups reflect differences in performance rather than differences in patient characteristics, follow-up survey responses are adjusted for “case mix” (i.e., variations of such characteristics across groups). To estimate risk-adjusted performance measure scores, we use hierarchical linear models that relate the patient-level PROMIS measure scores to group scores (conditioned on risk adjustment covariates, i.e., case mix); the hierarchy of data is patient observations within the designated accountable group. To calculate performance measure scores at the group level, it is necessary to perform hierarchical regressions with outcomes and a group-level random effect that will best estimate an adjusted score. The model was fit using the PROC MIXED procedure in SAS 9.4.

Risk Adjustment Variables

All three measures (including Pain Interference, Fatigue, Physical Health) are risk adjusted for:

- Patient age
- BMI at baseline
- Race and ethnicity (Hispanic, non-Hispanic white, non-Hispanic white, non-Hispanic Asian, other)
- Smoking status at baseline (current smoker, former smoker, non-smoker)
- Comorbidity of depression (Yes/No)
- Comorbidity of diabetic (Yes/No)
- Performance status at baseline (0=Normal activity level, 1=symptomatic and ambulatory; cares for self, 2=ambulatory > 50% of time; occasional assistance)
- Number of days between diagnosis date and the date of follow-up survey completion
- Radiation within two weeks before the date of follow-up survey completion (Yes/No)
- Number of days between the latest surgery and the date of follow-up survey completion
- Surgery severity Level (1= Lumpectomy (BCS) ± SLND, 2=Mastectomy with implant reconstruction ± SLND/ALND or Lumpectomy (BCS) with ALND, 3= Breast surgery + SLND/ALND + autologous reconstruction)
- Aromatase inhibitor (Yes/No)
- Baseline score of the outcome measure (before the start of chemotherapy)

Calculating Measure Scores

To estimate risk-adjusted quality measure scores, we utilize hierarchical linear models that relate the patient-level measure score to group scores (conditioned on risk adjustment covariates). The hierarchy of data is patient observations within the designated accountable group.

Measure scores are calculated with the model assessed at all baseline covariate values (i.e., assuming patients all are white, non-smoker, not diabetic, not having depression, at normal activity level, no radiation within two weeks of the follow-up survey completion date, having level 1 surgery, no aromatase inhibitor, and with continuous covariates, including age, BMI, number of days between the latest surgery and the date of follow-up survey completion, number of days between diagnosis date and the date of follow-up survey completion, and baseline measure score, all at sample average).

Coefficients obtained in hierarchical linear regression models estimate the tendency of patients to respond more positively or negatively. Group performance measure scores are adjusted to the overall mean of case-mix variables across respondents from all reporting groups. Thus, whether the scores of a given group are adjusted upward or downward for a given measure depends not only on these case-mix adjustments, but also on the case mix of that group relative to the overall average of these case-mix characteristics. Specifically, the total case mix-adjustment for a given group is the sum of a series of products, where each product multiplies the adjustments by the difference between the group's mean on the corresponding case-mix variable and the overall mean on that case-mix variable.

[Response Ends]

2b.21. If an outcome or resource use measure is not risk-adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (i.e., case mix) is not needed to achieve fair comparisons across measured entities.

[Response Begins]

[Response Ends]

2b.22. Select all applicable resources and methods used to develop the conceptual model of how social risk impacts this outcome.

[Response Begins]

Published literature

Internal data analysis

[Response Ends]

2b.23. Describe the conceptual and statistical methods and criteria used to test and select patient-level risk factors (e.g., clinical factors, social risk factors) used in the statistical risk model or for stratification by risk.

Please be sure to address the following: potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10 or other statistical tests; correlation of x or higher. Patient factors should be present at the start of care, if applicable. Also discuss any "ordering" of risk factor inclusion; note whether social risk factors are added after all clinical factors. Discuss any considerations regarding data sources (e.g., availability, specificity).

[Response Begins]

The Patient-Reported Symptoms Following Chemotherapy grouped measures assess pain interference, fatigue and overall physical health. PRO data were collected with a single survey that integrates the PROMIS scales that assess pain interference, fatigue and overall quality of life. As PRO-PMs, these measures were developed as grouped to facilitate implementation; reduce burden for providers and patients; and contribute to interpretation/clinical meaningfulness. The development of the risk adjustment model was performed to all three measures simultaneously.

Based on review of the literature and expert guidance, 38 risk adjustment variables were considered for inclusion in testing. TEP members participated in a structured Delphi process to rank feasibility and importance of gathering each of the candidate variables. This process resulted in 26 variables that were collected during PROMOnc testing, including:

- Patient demographics
- Social risk factors or proxies (e.g., race/ethnicity; dual eligibility)
- Clinical variables related to cancer and cancer treatments
- Other clinical variables (e.g., comorbidities)
- Survey scores at baseline

Each of these 26 risk adjustment variables collected was reviewed with the TEP after testing for missingness and threats to reliability. Five variables were removed at this review, and multiple discreet variables were converted into categories (using an iterative, evidence-based expert review process). Next, we examined the predictive ability of each potential risk adjustment variable by conducting bivariate analyses between each of the potential variables and each PROMOnc measure using regression analysis. We reviewed these data with the TEP, with a goal of including in further modeling only those that reach a significance level or were considered to meet the following criterion: have very high face validity/clinical meaningfulness plus little to no reporting burden. Variables with a significant association with any one of the three measures ($p < .10$) were included for review, with an a priori plan to create one common risk adjustment model. All final variables were tested for collinearity; none was found. These variables were used to create the final model, as described in the section below.

We also tested the survey mode as a potential risk adjustor to determine whether survey mode adjustments were needed to fairly compare survey scores across groups using different modes of administration. Groups use one of the following modes of survey administration: tablet in office, paper in office, electronic at home, or other. Due to the impact of the COVID pandemic, surveys from 78.55% of responding patients were administered via electronic at home and 17.03% via phone, while 3.79% via tablet in office and 0.63% via paper in office. We conducted linear regression analysis predicting each of the outcomes from survey mode. We found no significant effects of survey mode on responses to any of the PROMOnc outcomes. With no significant association between mode of survey administration and outcome measures, we do not need to adjust for mode of survey administration in scoring.

[Response Ends]

2b.24. Detail the statistical results of the analyses used to test and select risk factors for inclusion in or exclusion from the risk model/stratification.

[Response Begins]

Risk Adjustment Results

We present the risk-adjustment model coefficient estimates in Table 2b.3.

Table 2b.3. Regression Coefficients in Risk Adjustment Models - Fatigue

Risk Adjustor	Regression Coefficient	Standard Error	p-value
Baseline PROMIS Score	0.43	6.88	0.00
Surgery Level 1	0.43	0.22	0.83
Surgery Level 2	-1.22	-0.62	0.53
Surgery Level 3	8.77	1.34	0.18
Hispanic	1.40	0.68	0.50
Non-Hispanic Black	0.43	0.25	0.80
Non-Hispanic Asian	3.62	1.77	0.08
Other Race	1.43	0.68	0.50
Former Smoker	0.79	0.64	0.52
Current Smoker	0.57	0.30	0.77
Depression	0.89	0.28	0.78
Diabetic	-1.51	-0.63	0.53
Performance Status	1.23	0.83	0.41
Age	-1.03	-2.02	0.05
BMI	1.16	2.22	0.03
Aromatase Inhibitor	-2.92	-2.43	0.02
Days Between Diagnosis and Follow-Up Survey	-0.51	-1.00	0.32
Days Between Latest Surgery and Follow-Up Survey	1.35	2.27	0.02
Radiation Within Two Weeks of Follow-Up Survey	0.26	0.19	0.85

[Response Ends]

2b.25. Describe the analyses and interpretation resulting in the decision to select or not select social risk factors.

Examples may include prevalence of the factor across measured entities, availability of the data source, empirical association with the outcome, contribution of unique variation in the outcome, or assessment of between-unit effects and within-unit effects. Also describe the impact of adjusting for risk (or making no adjustment) on providers at high or low extremes of risk.

[Response Begins]

We considered insurance status of Medicaid or Medicare/Medicaid dual eligibility in the analyses to select risk factors. Among 323 survey respondents, 11 patients (3.41%) are eligible for Medicaid or dual eligible. Its association with the three PRO measures is not significant (r's < .003 with p-values >.33). Thus, we decided not to adjust for this factor.

[Response Ends]

2b.26. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (describe the steps—do not just name a method; what statistical analysis was used). Provide the statistical results from testing the approach to control for differences in patient characteristics (i.e., case mix) below. If stratified ONLY, enter "N/A" for questions about the statistical risk model discrimination and calibration statistics.

Validation testing should be conducted in a data set that is separate from the one used to develop the model.

[Response Begins]

To investigate the overall effect of the risk adjustment model, we compared group-level estimates without adjustment versus group-level estimates after adjusting using the multivariate model. We calculated Kendall's tau, a measure of rank correlation, which expresses the proportion of group pairs whose relative rankings were reversed by adjustment, scaled from 1 for no changes to −1 for a complete reversal of rankings. A tau value near 0 would indicate very little correlation between the unadjusted and adjusted scores and a tau value near 1 would indicate almost perfect correlation between the scores. A tau estimate equal to 1 would indicate that risk adjustment has no effect on the group-level scores, which would be concerning since adjustment is expected to have some effect. A tau estimate very close to -1 would indicate almost perfect negative correlations, meaning that risk adjustment almost completely re-ranked all groups, which would also be concerning since risk adjustment would not be expected to have such a dramatic effect.

Kendall's tau comparing scores between null and multivariate model adjustments for fatigue is .87.

[Response Ends]

2b.27. Provide risk model discrimination statistics.

For example, provide c-statistics or R-squared values.

[Response Begins]

To assess the contribution of the risk adjustment model to the measures, we calculated the proportional reduction of the unexplained variance, a multilevel version of R-squared values. We followed the approach in Snijders & Bosker (2012) and presented results in Table 2b.4.

Table 2b.4. Proportional Reduction of the Unexplained Variance of Risk Adjustment Model

Measure	Total Variance	Residual Variance After Including Baseline Measure	Residual Variance After Including Baseline Measure and Other Risk Adjustors	Proportion Reduction in Unexplained Variance Due to Baseline Measure	Proportion Reduction in Unexplained Variance Due to Other Risk Adjustors	Proportion Reduction of Unexplained Variance Due to All Risk Adjustors
Pain Interference	77.12	68.19	63.93	0.12	0.06	0.17
Fatigue	96.72	83.74	79.77	0.13	0.04	0.18
Physical Health	58.53	49.89	48.70	0.15	0.02	0.17

Reference:

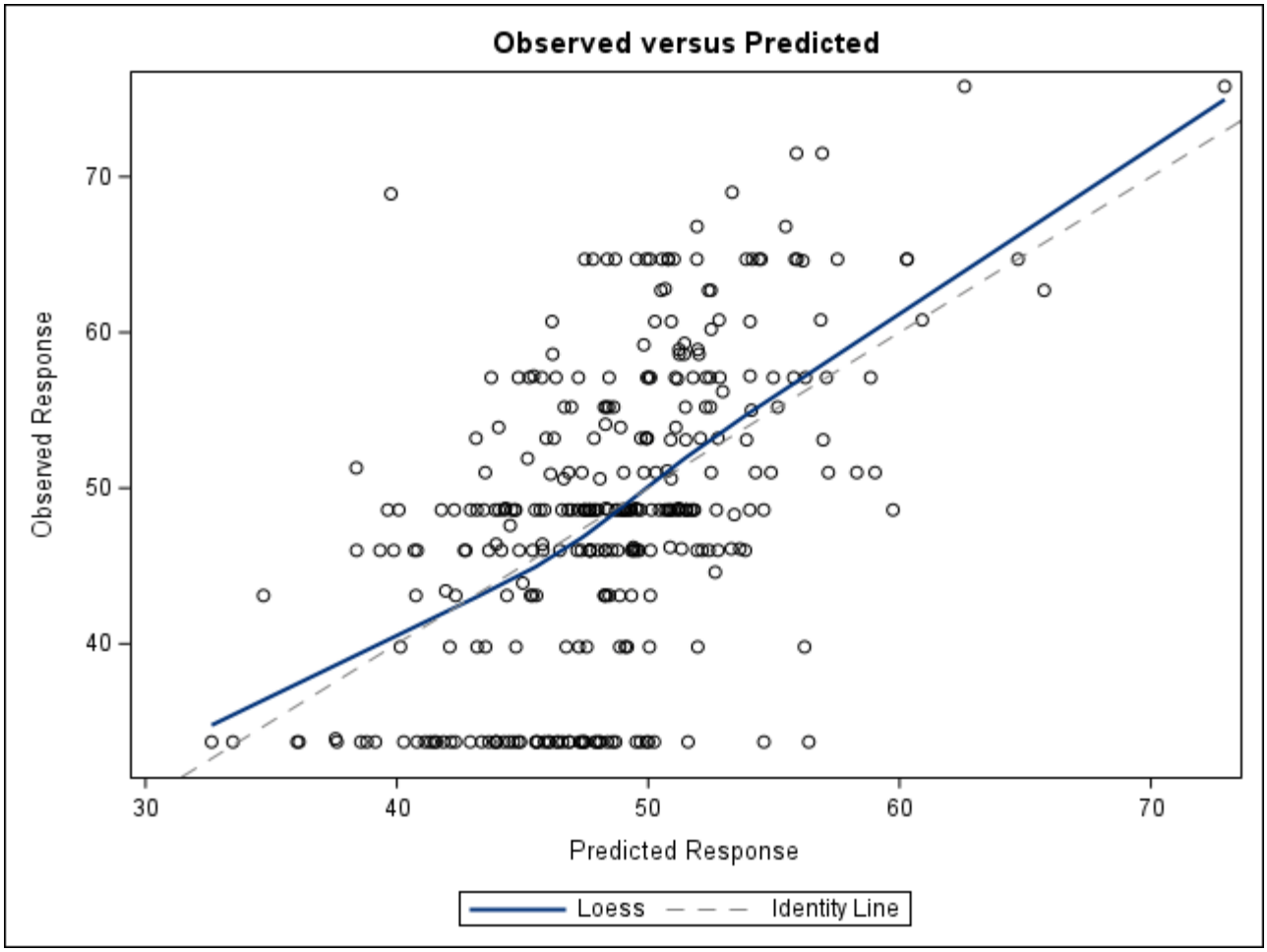
- Snijders, T.A.B. & Bosker, R. J. (2012). Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling (2nd edition). Sage.

[Response Ends]

2b.28. Provide the statistical risk model calibration statistics (e.g., Hosmer-Lemeshow statistic).

[Response Begins]

The Pearson correlation between the observed response and the predicted response is 0.55. The figure below plots the observed response with the predicted response, which includes an identity line and a Loess curve. The loess curve is in general close to the identity line, suggesting that the model in general is well specified.



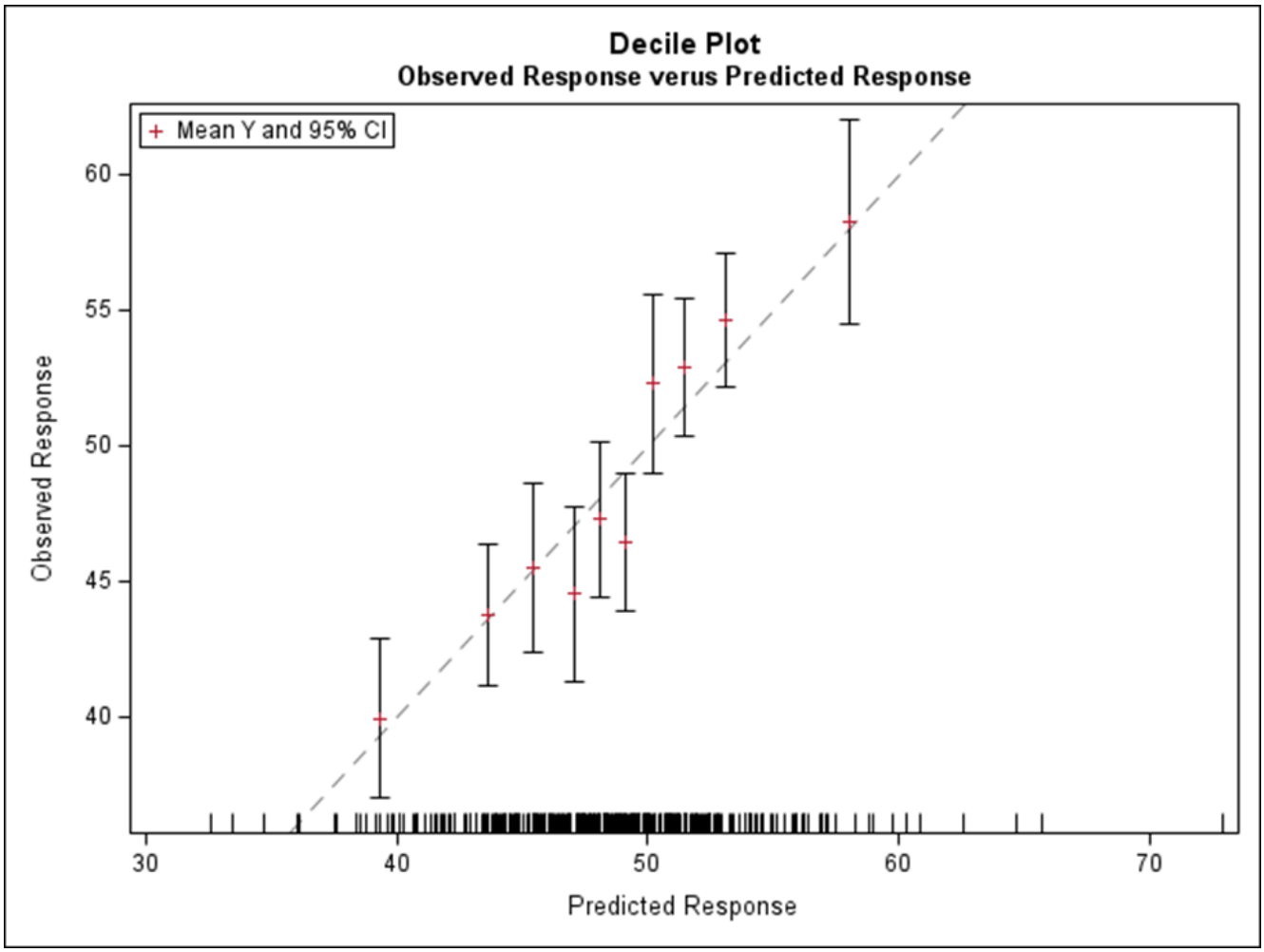
[Response Ends]

2b.29. Provide the risk decile plots or calibration curves used in calibrating the statistical risk model.

The preferred file format is .png, but most image formats are acceptable.

[Response Begins]

The figure below presents a decile plot by plotting the averaged observed response for each decile of the predicted response. The decile plot includes a diagonal line, which is the line of perfect agreement between the model and the data. We see that the 10 empirical means of the deciles fall close to the line and also vary randomly above and below the line, indicating that the model is well-specified.



[Response Ends]

2b.30. Provide the results of the risk stratification analysis.

[Response Begins]

N/A; the measure uses a statistical risk adjustment model not risk stratification.

[Response Ends]

2b.31. Provide your interpretation of the results, in terms of demonstrating adequacy of controlling for differences in patient characteristics (i.e., case mix).

In other words, what do the results mean and what are the norms for the test conducted?

[Response Begins]

The findings in Sections 2b.26 to 2b.29 support the use of risk adjustment. The Kendall's tau results (comparing scores between null and multivariate model adjustments) suggest a moderate effect of the adjustment model. The proportion reduction in unexplained variance is nears .20. The model provides a good fit to the data as shown in the plot comparing observed with predicted responses, and the decile plot. Together, these results suggest that risk adjustment model is in general well-specified, and the risk adjustment has a modest effect, but one that is likely to be important for groups with unusual patient mix.

[Response Ends]

2b.32. Describe any additional testing conducted to justify the risk adjustment approach used in specifying the measure.

Not required but would provide additional support of adequacy of the risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed.

[Response Begins]

N/A

[Response Ends]

Criterion 3. Feasibility

3.01. Check all methods below that are used to generate the data elements needed to compute the measure score.

[Response Begins]

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

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

Other (Please describe)

[Other (Please describe) Please Explain]

Patient survey

[Response Ends]

3.02. Detail to what extent the specified data elements are available electronically in defined fields.

In other words, indicate whether data elements that are needed to compute the performance measure score are in defined, computer-readable fields.

[Response Begins]

Some data elements are in defined fields in electronic sources

[Response Ends]

3.03. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using data elements not from electronic sources.

[Response Begins]

With the exception of the PROMIS survey items, all data elements required for the PROMOnc PRO-PMs should be captured in structured fields within an oncologist's electronic health record. During testing, some documentation continued in provider notes instead of available, structured fields; however, this practice is changing. Certain electronic

health records, including Epic and Cerner, now include PROMIS surveys, and leading EHRs allow for creation of patient surveys.

[Response Ends]

3.04. Describe any efforts to develop an eCQM.

[Response Begins]

Not applicable.

[Response Ends]

3.06. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

[Response Begins]

The PROMOnc measure developers acknowledge the impact of the COVID public health emergency on our testing efforts. The unfortunate overlap of the public health emergency with some of the PROMOnc testing period caused significant oncology practice disruption and resulted in less robust testing data than anticipated. The delayed and disrupted normal clinical schedules during the public health emergency impacted our test sites' ability to administer the patient surveys. Test sites were required to implement alternative strategies and modes for survey implementation than were planned prior to the pandemic. This resulted in fewer baseline and follow-up surveys than anticipated. We did, however, have sufficient testing data to complete the full analysis presented.

Except for the PROMIS survey items, all data elements required for the PROMOnc PRO-PMs should be captured in structured fields within an oncologist's electronic health record. Testing analyses included analyses of data missingness for all variables including clinical and demographic variables; certain data elements were removed during testing due to feasibility and reliability issues. Based on the clinical expertise and feasibility assessment of our TEP, and knowledge of the literature in oncology practice trends, PROMOnc believes the required data are in fact present in the medical record for the majority cases for which they were reported as missing. Throughout the field of oncology, there is increasing attention on ensuring that critical data elements such as those used in PROMOnc are captured in structured fields that can be easily retrieved from an EHR so feasibility of automated data capture is increasing rapidly. Moreover, certain electronic health records, including Epic and Cerner, now include PROMIS surveys, and leading EHRs allow for creation of patient surveys. When the measure is implemented in the context of a reporting program, we anticipate that missing data will be reduced and survey completion will be increased.

As in many measure testing projects, PROMOnc will expand and refine testing analyses during implementation for maintenance submission. We anticipate that when the measure is implemented outside of the COVID public health emergency and in the context of a reporting program, many of the implementation challenges we faced during PROMOnc testing will be minimized.

Collecting the baseline survey within the originally defined timeframe from patients taking oral chemotherapy was challenging. While oncology providers have full visibility into the oral chemotherapy prescription date, the actual start date may not be known if there are delays due to authorizations, pharmacy delays, or patient timeliness and preferences. In their deliberations regarding this uncertainty, the TEP broadened the PROMIS administration window for oral chemotherapy to promote patient capture. Another consideration is that most side effects and toxicities of common breast cancer oral chemotherapy agents do not interfere with the measures we collected until after the first week of administration with rare exception. The implementation guide for PROMOnc explicitly recognized these challenges with oral chemotherapy. Users were instructed to prioritize PROMIS administration prior to administration and only extend beyond if necessary.

During testing, we fielded a questionnaire to assess burden and feasibility related to data abstraction as well as implementation and patient-related activities. Seven ADCC sites and two MOQC sites responded to the burden

questionnaire. The majority 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 which test sites mitigated by building EHR reports to facilitate patient identification.

PROMOnc also fielded a survey to patients to assess their understanding of the survey and ease of us. Twelve patients provided feedback. Feedback indicated that 75% of respondents reported that it took them less than 10 minutes to complete the PROMOnc survey; 92% reported that they understood the survey instructions; 83% reported that they didn't have any technical issues completing the survey; and 83% felt that the time that it took to complete the survey was reasonable.

[Response Ends]

Consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

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

Attach the fee schedule here, if applicable.

[Response Begins]

There are no fees, licensing or other requirements for the survey. PROMIS measures are free and publicly available for use.

[Response Ends]

Criterion 4: Use and Usability

4a. Use

4a.01. Check all current uses. For each current use checked, please provide:

- **Name of program and sponsor**
- **URL**
- **Purpose**
- **Geographic area and number and percentage of accountable entities and patients included**
- **Level of measurement and setting**

[Response Begins]

Not in use

[Not in use Please Explain]

This PRO-PM is fully tested and will be submitted to the MUC List for the CMS Quality Payment Program. Thus, the measure is not publicly reported or used in an accountability application at this time. This is the first submission to NQF for endorsement.

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]

Public reporting

Payment Program

Professional Certification or Recognition Program

Quality Improvement with Benchmarking (external benchmarking to multiple organizations)

Quality Improvement (internal to the specific organization)

[Response Ends]

4a.03. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing), explain why the measure is not in use.

For example, do policies or actions of the developer/steward or accountable entities restrict access to performance results or block implementation?

[Response Begins]

This PRO-PM is fully tested and will be submitted to the MUC List for the CMS Quality Payment Program. Thus, the measure is not publicly reported or used in an accountability application at this time. This is the first submission to NQF for endorsement.

[Response Ends]

4a.04. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes: used in any accountability application within 3 years, and publicly reported within 6 years of initial endorsement.

A credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.

[Response Begins]

The goal of this project is to produce quality measures that can be used by providers eligible for CMS' Merit-Based Incentive Payment System (MIPS) who provide oncology care services to their patients, so that the patient experience of care components of high-quality care can be attributed to their providers and used to incentivize quality improvement. Medicare providers now choose one of two payment tracks – alternative payment models (APMs) and MIPS – which offer different combinations of incentives and requirements to encourage high-quality, low-cost care. PROMOnc measures will be submitted to the 2023 MUC List for inclusion in CMS' Quality Payment Programs, including MIPS and APMs. If the measure is added to the CMS MUC List in December 2023, we will support the MAP process through February 2024 and then support the adaptation of the measures for specification for the QPP during September and October 2024. The determination of whether the measure is accepted in the QPP should be in December 2024. For implementation in payment programs such as the CMS Quality Payment Program (QPP), providers will need to submit data to a third-party vendor to aggregate the data and calculate risk-adjusted scores.

[Response Ends]

4a.05. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

Detail how many and which types of measured entities and/or others were included. If only a sample of measured entities were included, describe the full population and how the sample was selected.

[Response Begins]

Ten practice sites were included in PROMOnc testing. PROMOnc project managers met with practice administrators and oncologists from the test sites on a twice-monthly basis. Any issues that were identified were escalated to the PROMOnc project team and resulted in additional training, definitional clarifications and revisions to the testing implementation guide. The PROMOnc project team also met with each test site to review response rates and discuss implementation issues and elicit best practices.

During testing, we received feedback from the PROMOnc Technical Expert Panel and PROMOnc Steering Committee. Feedback was obtained via 14 zoom meetings with the PROMOnc TEP, 6 zoom meetings with the PROMOnc Steering Committee, and 2 meetings with the MOQC Patient and Caregiver Council. Workgroups of the TEP were convened to address targeted areas for discussion as needed, including 5 meetings with the PROMOnc Clinician Workgroup and 2 meetings with the Methods Workgroup. Refer to Table 4a.1 for topics addressed in each TEP meeting; refer to Table 4a.2 for topics addressed at each Steering Committee meeting; refer to Table 4a.3 for topics addressed at each Clinician Workgroup meeting; and refer to Table 4a.4 for topics addressed at each Methods Workgroup meeting.

Table 4a.1: Technical Expert Panel Meeting Dates & Topics

Date	Meeting Topics
12-20-2018	Orientation Webinar
1-08-2019	Review PRO Instrument Landscape and Discuss Criteria for Selection
2-19-2019	• Review Landscape of Potential PROMs and Approach for PRO-PM Project • Review Ranking of Potential Questions and Subscales; TEP Input to Approach and PROM Selection
3-19-2019	• Confirm Selection of PROMs for PRO-PMs • Discuss Measure Rationale and Refine Measure Specifications • Brainstorm about Risk Adjustment Variables
4-16-2019	• Review and Discuss Options for Numerators • Discuss Timing of Survey Administration • Brainstorm about Risk Adjustment
5-21-2019	• Discuss Updated Measure Specifications • Discuss Reliability and Validity Testing • Discuss Plans to Assess Burden & Feasibility • Confirm Risk Adjustment Variables for Testing
9-10-2019	• Review Key Findings from Alpha Testing (Data Quality Assurance, Missing Data, Data Quality) • Discuss and Approve Recommendations for Beta Testing (Modifications to Data Dictionary)
2-25-2020	• Input from Beta Testing • Review Comments Received During Public Comment Period
10-28-2020	• Review Input from Beta Midpoint Testing • Review Plans for Burden and Feasibility Assessment
12-15-2020	Review Results of Feasibility & Burden Assessment

Date	Meeting Topics
6-08-2021	• Review Initial Data Analysis • Review and Confirm Recommendations related to Dropping/Modifying Certain Data Elements, Denominator Exclusions and Selection of Numerator Option
7-13-2021	• Review and Confirm Recommendations related to Categorization of Surgery and Chemotherapy • Obtain Input on Survey 3 Time Window • Obtain Input about Measures to Use for Validity Testing
11-16-2021	• Review and Confirm Recommendations related to Categorization of Surgery and Chemotherapy • Review Decision to Remove Survey 2 from Measure Specifications • Obtain Input on Survey 3 Time Window • Obtain Input about Measures to Use for Validity Testing • Review Which Sites to Include for Performance Measure Scoring, e.g., Sites with 5 of More Follow-Up Surveys
12-14-2021	• Review Updated Analyses • Review Candidate Risk Adjustors • Review Surgery and Chemotherapy Categorization into Variables for Risk Adjustment • Selection of Risk Adjustment Variables • Review Performance Measure Scores (Risk Unadjusted and Adjusted) • Obtain TEP Input on Survey Respondents vs. Non-Respondents

Table 4a.2: Steering Committee Meeting Dates & Topics

Date	Meeting Topics
12-17-2018	Orientation Webinar
2-25-2019	• Criterion for PROM Instrument Selection • TEP & Patient Panel Input on PROM Questions • Approach to Selecting PROM Instrument
4-29-2019	Design Decisions for PRO-PMs in Accountability Programs
11-21-2019	• Alpha Testing Results • Measure Specifications • Design Decisions for Assessing Burden & Feasibility • Discuss Options for Implementing PRO-PMs in Payment Models; Perspectives on CMS Oncology Care First Model
12-10-2020	• Review Feasibility & Burden Assessment Methodology & Results • Discuss Recommendations for Increasing Adoption of PROMs

Table 4a.3: Clinician Workgroup Meeting Dates & Topics

Date	Meeting Topics
5-03-2019	• Discuss Timing of Survey Administration • Develop Hypotheses for Expected Change for Each Domain Between Timepoints • Refine Risk Adjustment Variables

Date	Meeting Topics
6-07-2019	• Review Data Dictionary Questions • What Comorbidities or Indices Should Be Used • How to Collect Smoking Status • Date of Cancer Diagnosis • How to Define Concurrent Cancer Diagnoses • AJCC Clinical and Pathologic Stage • How to Define Performance Status • Chemotherapy Regimen Collection Timing and Whether to Group • Treatment Data Element Questions
11-22-2019	Input on Numerator Options
6-23-2021	Recommendations for Categorization of Surgery and Treatment Regimens
11-16-2021	• Changes to Survey Time Windows (Remove Survey 2; Expand Time Window for Survey 3) • Recommendation to Include Sites with 5 or More Survey 3 for Performance Measure Scoring • Selection of Risk Adjustment Variables

Table 4a.4: Methods Workgroup Meeting Dates & Topics

Date	Meeting Topics
8-20-2019	• Input on numerator options • Input on risk adjustment model • Input on missing data analyses • Approach to validity testing
1-10-2020	• Discuss numerator options • Review PROMIS symptom severity thresholds and minimal important differences to consider interpretability and use

See Additional (2) for members of these committees. We also received feedback during the measure development public comment period and reached out to the American Society of Clinical Oncologists (ASCO) and the Community Oncology Alliance (COA) to encourage public comment.

Patients and caregivers were engaged throughout the PROMOnc testing process. PROMOnc engaged the Patient and Caregiver Oncology Quality Council from the Michigan Oncology Quality Consortium (MOQC) to provide input into the selection of PROMIS scales for assessing patient-reported outcomes. Two representatives from the MOQC Patient and Caregiver Oncology Quality Council also participated on the PROMOnc Steering Committee. And, PROMOnc collaborated with the Seattle Cancer Care Alliance (SCCA) Patient Family Advisory Council (PFAC) on implementation of a patient burden questionnaire during testing.

[Response Ends]

4a.06. Describe the process for providing measure results, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

[Response Begins]

The PROMOnc TEP received testing and measure results at multiple points during the testing period: after Alpha testing, at Beta testing mid-point, and multiple iterations of final Beta analyses.

Beta testing results were reviewed with all PROMOnc test sites during two meetings which included review of the measure specifications, including the risk adjustment variables, review of unadjusted and adjusted performance results, and the distribution of performance across test sites.

[Response Ends]

4a.07. Summarize the feedback on measure performance and implementation from the measured entities and others. Describe how feedback was obtained.

[Response Begins]

Feedback was obtained from the PROMOnc test sites and others using the processes described in 4a.05. Committee members contributed to multiple specification refinement decisions; the PROMOnc TEP included 11 oncology clinicians, and the Steering Committee included 3. Members provided guidance regarding methods to integrate survey administration into clinical workflows to minimize burden. They determined numerator analytic options (based on the PROMIS survey data) to maximize clinical meaningfulness and enhance reliability. They established denominator exclusions for testing, and determined final exclusions based on the testing results. They selected the candidate risk adjustment variables, which were tested, and the final variables in the model. When reviewing measure performance data, they evaluated various ways to report the data to maximize meaningfulness for improvement.

The time windows for survey administration were established with direction from the TEP, which included 11 practicing oncology clinicians. Over the course of 5 meetings, the TEP carefully considered balancing clinical meaningfulness of the PROMIS scores with the norms of clinic schedules and workflows. Important differences were discussed between parenteral chemotherapy, administered in the practice infusion setting, and oral chemotherapy, taken in the patients' homes. Oncology providers have full visibility into the oral chemotherapy prescription date; however, the actual start date can be influenced by authorizations, pharmacy delays, and patient timeliness and preferences. Oncology providers are often not able to ascertain the actual start date until the patient returns for a check-in visit. In their deliberations regarding this uncertainty, the TEP broadened the PROMIS administration window for oral chemotherapy to promote patient capture. Another consideration is that most side effects and toxicities of common breast cancer oral chemotherapy agents do not interfere with the measures we collected until after the first week of administration with rare exception.

PROMOnc also fielded a survey to patients to assess whether patients felt the PROMOnc survey was meaningful; 83% reported that the survey responses would help the doctor and care team.

[Response Ends]

4a.08. Summarize the feedback obtained from those being measured.

[Response Begins]

In addition to the feedback described in 4a.07, throughout testing, we engaged clinical and data leads at our test sites through twice-monthly check-in calls with PROMOnc project managers and two check-in calls with the PROMOnc project team. Similar to feedback described in 4a.07, feedback from test sites during the check-in calls included challenges with identifying eligible patients, narrow survey administration window at baseline for patients taking oral therapy, and patient engagement due to COVID. Moreover, in a survey conducted among 8 clinicians at our test sites, respondents stated that clinicians from their cancer centers would support use of the PROMOnc survey to better understand patient symptoms, function and quality of life (4 "Yes, definitely" and 4 "Yes, somewhat"); to better manage patient symptoms, enhance function and improve quality of life (3 "Yes, definitely" and 5 "Yes, somewhat"); and to measure the quality of care (3 "Yes, definitely" and 5 "Yes, somewhat").

[Response Ends]

4a.09. Summarize the feedback obtained from other users.

[Response Begins]

In addition to the feedback described in 4a.07, PROMOnc fielded a public comment period and received twenty comments from two specialty societies, two provider organizations, two individuals and one consumer organization. In public comment, we received comments about the numerator options, survey collection timepoints, stratification, case mix adjustment variables, selection of the PROM instrument, workflow challenges and clinical use of the patient-reported outcomes. PROMOnc also fielded a survey to patients to assess whether patients felt the PROMOnc survey was meaningful. Twelve patients provided feedback. 83% reported that the survey responses would help the doctor and care team.

[Response Ends]

4a.10. Describe how the feedback described has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

[Response Begins]

As previously reviewed, the PROMOnc committees were instrumental in defining and refining the specifications of the PRO-PM, including time windows, denominator exclusions, numerator definitions, and risk adjustment. The MOQC Patient and Caregiver Oncology Quality Council influenced selection of the PROMIS survey instruments. Questions and issues raised by the PROMOnc test sites led to definition and implementation guide refinements. The feedback that we received from public comment was discussed with the TEP and it informed specification refinement.

[Response Ends]

4b. Usability

4b.01. You may refer to data provided in Importance to Measure and Report: Gap in Care/Disparities, but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included). If no improvement was demonstrated, provide an explanation. If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

[Response Begins]

The measure just completed testing and has not been used for performance improvement at the time for submission of endorsement. Briefly, the rationale for the measure is as follows: Many patients who undergo chemotherapy with curative intent experience persistent detriments following treatment. Common persistent symptoms include pain, fatigue and detriments to health-related quality of life. Evidence-based practices can manage these symptoms during treatment and position patients better for the survivorship phase. As a result of oncologists assessing and actively managing symptoms during chemotherapy, patients will experience lower symptom burden, less suffering, and will be better prepared and have lower persistent symptom interference as they enter the survivorship phase. Group-level PRO-PM data are useful to inform practice improvement. Payers can promote these practice changes that improve patient outcomes by rewarding high-performing physicians and practices.

[Response Ends]

4b.02. Explain any unexpected findings (positive or negative) during implementation of this measure, including unintended impacts on patients.

[Response Begins]

To date, we have not encountered any unintended adverse consequences from measuring cancer patients' pain, fatigue or detriments to health-related quality of life. We did not expect to as PROMIS survey development included multiple levels of patient input (see 1a.02). Also, prior to implementation, PROMOnc engaged the Patient and Caregiver Oncology Quality Council from the Michigan Oncology Quality Consortium (MOQC) to provide input into the selection of PROMIS scales for assessing patient-reported outcomes. The council found the PROMIS surveys to be highly acceptable.

[Response Ends]

4b.03. Explain any unexpected benefits realized from implementation of this measure.

[Response Begins]

While the PROMIS survey implementation during PROMOnc testing was used to inform testing analyses for the PRO-PM, we encouraged test sites to use the data collected to identify and address concerns during routine clinical care. In a survey conducted among 8 clinicians at our test sites, respondents stated that clinicians from their cancer centers would support use of the PROMOnc survey to better understand patient symptoms, function and quality of life (4 "Yes, definitely" and 4 "Yes, somewhat"); to better manage patient symptoms, enhance function and improve quality of life (3 "Yes, definitely" and 5 "Yes, somewhat"); and to measure the quality of care (3 "Yes, definitely" and 5 "Yes, somewhat").

[Response Ends]

Criterion 5: Related and Competing Measures

5.01. Search and select all NQF-endorsed related measures (conceptually, either same measure focus or target population).

(Can search and select measures.)

[Response Begins]

0220: Adjuvant hormonal therapy is recommended or administered within 1 year (365 days) of diagnosis for women with AJCC T1cN0M0 or Stage IB – Stage III hormone receptor positive breast cancer

0387e: Breast Cancer: Hormonal Therapy for Stage I (T1b)-IIIC Estrogen Receptor/Progesterone Receptor (ER/PR) Positive Breast Cancer

[Response Ends]

5.02. Search and select all NQF-endorsed competing measures (conceptually, the measures have both the same measure focus or target population).

(Can search and select measures.)

[Response Begins]

[Response Ends]

5.03. If there are related or competing measures to this measure, but they are not NQF-endorsed, please indicate the measure title and steward.

[Response Begins]

There are no NQF-endorsed measures with the same focus. NQF measures 0220 and 0387e have overlapping target populations: women receiving curative breast cancer treatment.

Regarding non-NQF endorsed measures, the Minnesota Community Measurement group (MNCM) has undertaken an initiative to develop PRO-PMs for oncology, but these measures are complementary, not competing. The MNCM measures assess symptom control (pain, nausea and constipation) during days 5 – 15 of the chemotherapy treatment

cycle (MNCM 2021). The PROMOnc and MNCM measure are complimentary in that the MNCM symptom control measures are focused on the window during the chemotherapy cycle (Day 5 to Day 15) with a goal of symptoms being in control (rated as none or mild) using the PRO-CTCAE tool for all adult patients undergoing chemotherapy regardless of cancer type. The PROMOnc measures are collected at different timepoints (start of chemotherapy treatment and 3 months after completion of chemotherapy) with the PROMIS tool which does not overlap with measures under development by MNCM.

Recent PCORI research conducted by Stover et al. (2022) tested PROMs to see if the PROMs could detect differences in how well cancer centers control patients' treatment side effects. The PROMs, which included question items from the PRO-CTCAE and PROMIS (and outcomes that included nausea, constipation, diarrhea, neuropathy, pain, fatigue, insomnia, anxiety, depression and physical function) detected differences between centers. Based on the 12 PROMs, one cancer center performed better than others, and one performed worse. However, not enough patients completed the surveys to consistently compare the quality of care across cancer centers. (Stover et al. 2022) Similar to the MNCM measures, these measures were based on the symptom severity during days 5 – 15 of the chemotherapy cycle so do not overlap with PROMOnc measures.

References:

- Minnesota Community Measurement (MNCM). 2021. https://helpdesk.mncm.org/helpdesk/KB/View/40816577-oncology-measures-symptom-control-during-chemotherapy
- Stover AM, Urick BY, Jansen J, Carr P, Deal A, Spears PA, Smith ML, Geoghegan C, Basch EM. (2022) Developing Patient-Reported Outcome Measures to Assess Side Effects of Cancer Treatment. Patient-Centered Outcomes Research Institute (PCORI) https://doi.org/10.25302/09.2021.ME.150732079

[Response Ends]

5.04. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s), indicate whether the measure specifications are harmonized to the extent possible.

[Response Begins]

Yes

[Response Ends]

5.05. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

[Response Begins]

The denominators of the two NQF measures with overlapping target populations are below. Both measures are assessing use of hormonal therapy in the numerator, and thus limit the denominator to tumors that are estrogen receptor positive or progesterone receptor positive, which is not relevant to the PROMOnc PRO-PM target population. Otherwise, the measures include similar populations, when denominator inclusion and exclusion criteria are considered.

NQF# 0220 denominator: Include if all of the following characteristics are identified:

Women

Age = 18 at time of diagnosis

Known or assumed to be first or only cancer diagnosis

Epithelial malignancy only

Invasive tumors

Primary tumors of the breast

AJCC T1cN0M0 or Stage IB – IIIC

Primary tumor is estrogen receptor positive or progesterone receptor positive

All or part of 1st course of treatment performed at the reporting facility

Known to be alive within 1 year (365 days) of date of diagnosis

Surgical procedure of the primary site

NQF # 0387e denominator: All female patients aged 18 years and older with a diagnosis of breast cancer with Stage I (T1b) through IIIC, estrogen receptor (ER) or progesterone receptor (PR) positive breast cancer

[Response Ends]

5.06. Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality). Alternatively, justify endorsing an additional measure.

Provide analyses when possible.

[Response Begins]

The National Quality Forum has also noted large gaps in cancer-focused outcome measures. Within the NQF Cancer Standing Committee's Portfolio of Measures (18 measures), there are no outcome measures for breast cancer (NQF 2021a). NQF's Global Positioning System reports 22 endorsed cancer process measures and 4 endorsed cancer outcome measures. There are no endorsed cancer PRO-PMs (NQF 2021b). Notably, the Minnesota Community Measurement group (MNCM) has undertaken an initiative to develop PRO-PMs for oncology but these measures are complementary, not competing. The MNCM measures assess symptom control (pain, nausea and constipation) during days 5 – 15 of the chemotherapy treatment cycle. The PROMOnc and MNCM measures are complimentary in that the MNCM symptom control measures are focused on the window during the chemotherapy cycle (Day 5 to Day 15) with a goal of symptoms being in control (rated as none or mild) using the PRO-CTCAE tool for all adult patients undergoing chemotherapy regardless of cancer type (MNCM 2021). The PROMOnc measures are collected at different timepoints (start of chemotherapy treatment and 3 months after completion of chemotherapy) with the PROMIS tool which does not overlap with measures under development by MNCM.

Recent PCORI research conducted by Stover et al. (2022) tested PROMs to see if the PROMs could detect differences in how well cancer centers control patients' treatment side effects. The PROMs, which included question items from the PRO-CTCAE and PROMIS (and outcomes that included nausea, constipation, diarrhea, neuropathy, pain, fatigue, insomnia, anxiety, depression and physical function) detected differences between centers. Based on the 12 PROMs, one cancer center performed better than others, and one performed worse. However, not enough patients completed the surveys to consistently compare the quality of care across cancer centers. (Stover et al. 2022) Similar to the MNCM measures, these measures were based on the symptom severity during days 5 – 15 of the chemotherapy cycle so do not overlap with PROMOnc measures.

References:

- Minnesota Community Measurement (MNCM). 2021. https://helpdesk.mncm.org/helpdesk/KB/View/40816577-oncology-measures-symptom-control-during-chemotherapy
- National Quality Forum (NQF 2021a). Cancer, Spring 2020 Cycle: CDP Report. Technical Report, February 22, 2021.
- National Quality Forum (NQF 2021 b). Global positioning system. Available at: https://www.qualityforum.org/QPS Accessed December 9, 2021.
- Stover AM, Urick BY, Jansen J, Carr P, Deal A, Spears PA, Smith ML, Geoghegan C, Basch EM. (2022) Developing Patient-Reported Outcome Measures to Assess Side Effects of Cancer Treatment. Patient-Centered Outcomes Research Institute (PCORI) https://doi.org/10.25302/09.2021.ME.150732079

[Response Ends]