NATIONAL QUALITY FORUM

# MEASURE WORKSHEET

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

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

## Brief Measure Information

**NQF #:** 0384

**Corresponding Measures:** 0384e

**De.2. Measure Title:** Oncology: Medical and Radiation - Pain Intensity Quantified

**Co.1.1. Measure Steward:** PCPI

**De.3. Brief Description of Measure:** Percentage of patient visits, regardless of patient age, with a diagnosis of cancer currently receiving chemotherapy or radiation therapy in which pain intensity is quantified

**1b.1. Developer Rationale:** An estimated 1.7 million new cases of cancer are diagnosed in the US each year. (1) Pain is a commonly occurring symptom for cancer patients as 30% to 50% (510,000 to 850,000 each year based on current statistics) will experience moderate to severe pain.(2) Initial and ongoing pain assessments are essential to determine the pathophysiology of pain and ensure proper pain management. According to the National Comprehensive Cancer Network, there is increasing evidence in oncology that survival is linked to symptom reporting and control and that pain management contributes to broad quality-of-life improvement.(3) Evidence has shown a positive association between higher symptom scores and higher rates of documentation and clinical actions taken. (4) A study published this year (2019) provides further evidence that symptom monitoring following treatment for cancer is associated with increased survival. (5) Cancer patients have reported that pain interferes with their mood, work, relationships with other people, sleep and overall enjoyment of life.(6) To maximize patient outcomes, pain management is an essential part of oncologic management.(3)

(1) National Cancer Institute. Cancer statistics. National Institutes of Health. 2017. https://www.cancer.gov/about-cancer/understanding/statistics

(2) Wiffen PJ, Wee B, Derry S, Bell RF, Moore RA. Opioids for cancer pain – an overview of Cochrane reviews. Cochrane Database of Systematic Reviews 2017, Issue 7. Art. No.: CD012592. DOI: 10.1002/14651858.CD012592.pub2.

(3) National Comprehensive Cancer Network (NCCN). NCCN Clinical Practice Guidelines in Oncology. Adult cancer pain Version I.2018. January 22, 2018. http://www.nccn.org

(4) Seow H, Sussman J, Martelli-Reid L, Pond G, Bainbridge D. Do high symptom scores trigger clinical actions? An audit after implementing electronic symptom screening. J Oncol Pract. 2012 Nov;8(6):e142-8. doi: 10.1200/JOP.2011.000525.

(5) Denis F, Basch E, Septans AL, Bennouna J, Urban T, Dueck AC, Letellier C. Two-Year Survival Comparing Web-Based Symptom Monitoring vs Routine Surveillance Following Treatment for Lung Cancer. JAMA. 2019 Jan 22;321(3):306-307. doi: 10.1001/jama.2018.18085.

(6) Moryl N, Dave V, Glare P, Bokhari A, Malhotra VT, Gulati A, et al. Patient-reported outcomes and opioid use by outpatient cancer patients. J Pain. 2018. Mar;19(3):278-290.

**S.4. Numerator Statement:** Patient visits in which pain intensity is quantified

**S.6. Denominator Statement:** All patient visits, regardless of patient age, with a diagnosis of cancer currently receiving chemotherapy or radiation therapy

**S.8. Denominator Exclusions:** None

**De.1. Measure Type:** Process

**S.17. Data Source:** Registry Data

**S.20. Level of Analysis:** Clinician : Group/Practice, Clinician : Individual

**IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:**

**IF this measure is included in a composite, NQF Composite#/title:**

**IF this measure is paired/grouped, NQF#/title:**

**De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** This measure is paired with NQF #0383 Oncology: Plan of Care for Pain, which assesses whether patients who report pain have a documented plan of care. These measures together represent a stepwise approach to attenuating pain that commonly results from cancer therapy. This measure requires the initial and ongoing assessment and quantification of pain which are essential to formulate the most appropriate plan with the intent of improving patient outcomes.

## Preliminary Analysis: Maintenance of Endorsement New Measure

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

### Criteria 1: Importance to Measure and Report

#### 1a. Evidence

**Maintenance measures – less emphasis on evidence unless there is new information or change in evidence** U**since the prior evaluation.**

**1a. Evidence.** The evidence requirements for a *structure, process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured. For measures derived from patient report, evidence also should demonstrate that the target population values the measured process or structure and finds it meaningful.

meaningful.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☐ **Yes** ☒ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

**Summary of prior review in 2012**

- In 2012, the Committee agreed that the measure developer presented good evidence showing the prevalence of pain; the measure will impact a large number of patients.

**Changes to evidence from last review**

☐ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☒ **The developer provided updated evidence for this measure:**

**Updates:**

- The developer provided an updated logic model tying symptom reporting and control to survival and noting that pain management contributes to broad quality-of-life improvement.
- The evidence to support the measure was updated to include the 2018 NCCN Clinical Practical Guidelines in Oncology- Adult Cancer Pain. The updated guideline states that:
    - Patients must be screened for pain at each visit
    - Pain intensity must be characterized by the patient and routinely quantified and documented.
    - Comprehensive pain assessment must be performed if pain is present and regularly performed for persistent pain.
    - Patients must be evaluated for risk factors for opioid abuse/misuse/diversion.
    - Pain reassessment must be performed at specific intervals.
- The 2018 NCCN guideline **level of evidence: Category 2A** - Based upon lower-level evidence, there is uniform NCCN consensus that the intervention is appropriate.
    - No information on the quantity of the body of evidence (total # of studies not articles or papers) and consistency provided.
- The 2018 NCCN guideline acknowledges the possible harms associated with the potential for misuse and abuse of opioids. The guideline states that initial patient evaluation should include an assessment of risk factors for abberant use of pain medications and that patients should be monitored for these behaviors as well.

**Summary of Committee review in Fall 2018**

- The Committee had a lengthy discussion about the quality of the evidence that demonstrates documenting pain leads to improved patient outcomes. Some Committee members questioned whether documenting pain intensity translated into a change in patient management. Other Committee members expressed concern about using different pain scales to quantify pain levels and the relationship to improved outcomes for cancer patients. Overall, the Committee agreed asking patients about their pain is important and likely leads to improved pain management and pain control. The Committee acknowledged that the evidence provided in the measure submission form is insufficient and does not meet current NQF Measure Evaluation Criteria for process measures. In the absence of empirical evidence demonstrating that documenting pain intensity improves patient outcomes, the Committee voted to pass the evidence criterion with an exception and determined it is beneficial to hold providers accountable for performance on this measure. The patient representatives on the Committee emphasized that asking patients about their pain is important and they value this measure.

***Questions for the Committee:***

- What is the relationship between documenting pain intensity and the following outcomes for patients with a diagnosis of cancer currently receiving chemotherapy or radiation therapy: pain management and pain control, quality of life improvement, and survival?
- How strong is the evidence for this relationship?

**Guidance from the Evidence Algorithm**

Process measure with systematic review (SR) and grading of the body of evidence (Box 3) → Incomplete Quantity/Quality/Consistency (QQC) of the body of evidence from the SR. Quantity: not provided; Quality: lower-level evidence; Consistency: not provided (Box 4) → Guideline with lower-level evidence and/or recommendation with no grading of the evidence and incomplete QQC (Box 6) → Low

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

**RATIONALE:** Guidelines with lower-level evidence and/or recommendation with no grading of the evidence and incomplete QQC submitted to support the measure focus.

## 1b. Gap in Care/Opportunity for Improvement and 1b. disparities

**Maintenance measures – increased emphasis on gap and variation**

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

- The developer provided 2016 performance data from PQRS using the measure specifications. The PQRS testing data analysis results:
  - 251 Physicians
  - Mean: 0.88
  - Median: 0.98
  - Mode 1.0
  - Standard deviation: 0.21
  - Interquartile range: 0.12 (1.0–0.88)
- The developer provided additional average performance rates from PQRS in previous years:
  - 2015: 75.9%
  - 2014: 84.8%
  - 2013: 82.7%

**Disparities**

- This measure is included in federal reporting programs, but the developer notes that those programs have not yet made disparities data available to analyze and report. The developer provided a summary of data from the literature related to cancer treatment and the management of cancer-related pain.
- Disparities data from the measure as specified is required for maintenance of endorsement.

***Questions for the Committee:***

- Does the data demonstrate considerable variation or overall less-than-optimal performance that warrants a national performance measure?
- Are you aware of evidence that disparities exist in quantifying pain intensity for cancer patients?

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

**RATIONALE:** No disparities data provided from the measure as specified that may demonstrate an opportunity for improvement/gap in care for certain sub-populations.

**Committee Pre-evaluation Comments:**
**Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)**

**1a. Evidence to Support Measure Focus: For all measures (structure, process, outcome, patient-reported structure/process), empirical data are required. How does the evidence relate to the specific structure, process, or outcome being measured? Does it apply directly or is it tangential? How does the structure, process, or outcome relate to desired outcomes? For maintenance measures –are you aware of any new**

**studies/information that changes the evidence base for this measure that has not been cited in the submission? For measures derived from a patient report: Measures derived from a patient report must demonstrate that the target population values the measured outcome, process, or structure."**

- Solid
- Although asking about pain is important, the evidence to support grading pain to improve quality of life is lacking
- The evidence provided relates directly to the process/outcome being measured with regards to patient pain.
- Same comments as placed for #0384e
- Measures accounted appropriately
- yes - this is important to patients
- Evidence was provided, no evidence showing just the documentation of pain has an impact on improved patient care
- The measure is a clinical guidelines-driven standard based on NCCN guidelines. The evidence is level 2A which is associated with low level evidence but a high degree of consensus by the expert panel. The measure is a process measure and is indirectly linked to the intended outcome of improved performance. I believe that the documentation of pain intensity supports best medical practice and can be associated with improved pain control and quality of life. I believe that the documentation of pain and the quantification of pain is valued by the target population.
- Documenting pain intensity can have a major impact on the actions taken to address the pain and resulting quality of life. I am not sure, however, of the evidence for this relationship.

**1b. Performance Gap: Was current performance data on the measure provided? How does it demonstrate a gap in care (variability or overall less than optimal performance) to warrant a national performance measure? Disparities: Was data on the measure by population subgroups provided? How does it demonstrate disparities in the care?**

- There is a gap. Disparities were reported
- performance is 75 to 82% and dropping. disparity data not available.
- Yes, the Developer provided 2016 performance data, which still have yet to make disparities data available.
- Same comments as placed for #0384e
- Performance Data on subgroups provided
- Limited data but gaps appear to exist
- Data presented is several years old with no disparities data presented
- The developers presented data from the last submission which included CMS PQRS 2016 reporting data. These data demonstrated that the mean performance rate for this measure is .88with a standard deviation of 0.21 and a range of 0.96 (0.04-1.0). The interquartile range is 0.12 (1.0–0.88). This is an improvement over the preceding three years of data. The developers also cited other studies in the literature to demonstrate a wide variation in pain intensity documentation with one study showing variability before (84%) and after intervention (43%) suggesting an ongoing need to improve documentation across the continuum of care. The developers state that this is a CMS measure and no disparity data was presented from CMS for analysis. I believe these data suggest variability and inconsistency in evaluation of pain management and follow up.
- It appears the disparities data is from literature, not from the program. It would be important to have data from the measure.

## Criteria 2: Scientific Acceptability of Measure Properties

**2a. Reliability: Specifications and Testing**

**2b. Validity: Testing; Exclusions; Risk-Adjustment; Meaningful Differences; comparability; Missing Data**

### Reliability

**2a1. Specifications** requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. For maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures.

**2a2. Reliability testing** demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers. For maintenance measures – less emphasis if no new testing data provided.

### Validity

**2b2. Validity testing** should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For maintenance measures – less emphasis if no new testing data provided.

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

**Complex measure evaluated by Scientific Methods Panel**? ☐ **Yes** ☒ **No**

**Evaluators:** Staff

**Evaluation of Reliability and Validity:**

***Questions for the Committee regarding reliability:***

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- The level of analysis (LoA) specified are for clinician groups and individual clinicians, and therefore two sets of testing are expected. The level of analysis of the measure must align with testing; for this measure it was unclear whether testing was provided for both levels of analysis. One of these LoA may have to be dropped from the specifications, unless the developer can clarify how to interpret the testing results. Additional testing may be required if they would like the measure to be endorsed for both levels of analysis.
- Reliability testing limited to providers with 10 or more patients eligible for this measure – this minimum threshold is not included in the specifications.

***Questions for the Committee regarding validity:***

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?
- The Standing Committee should discuss the testing method used to demonstrate validity and determine if it is appropriate and if the testing results are sufficient.

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

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

### Staff Evaluation: Scientific Acceptability

Scientific Acceptability: Preliminary Analysis Form

**Measure Number:** 0384

**Measure Title:** Oncology: Plan of Care for Pain – Medical Oncology and Radiation Oncology

### Type of measure:

☒ **Process** ☐ **Process: Appropriate Use** ☐ **Structure** ☐ **Efficiency** ☐ **Cost/Resource Use**

☐ **Outcome** ☐ **Outcome: PRO-PM** ☐ **Outcome: Intermediate Clinical Outcome** ☐ **Composite**

### Data Source:

☐ **Claims** ☐ **Electronic Health Data** ☐ **Electronic Health Records** ☐ **Management Data**

☐ **Assessment Data** ☐ **Paper Medical Records** ☐ **Instrument-Based Data** ☒ **Registry Data**

☐ **Enrollment Data** ☐ **Other**

### Level of Analysis:

☒ **Clinician: Group/Practice** ☒ **Clinician: Individual** ☐ **Facility** ☐ **Health Plan**

☐ **Population: Community, County or City** ☐ **Population: Regional and State**

☐ **Integrated Delivery System** ☐ **Other**

### Measure is:

☐ **New** ☒ **Previously endorsed (**NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.)

### RELIABILITY: SPECIFICATIONS

1. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?** ☒ **Yes** ☐ **No**

   **Submission document:** "MIF_xxxx" document, items S.1-S.22

   ***NOTE****: NQF staff will conduct a separate, more technical, check of eCQM specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.*

2. **Briefly summarize any concerns about the measure specifications.**
   - The level of analysis (LoA) specified are for clinician groups and individual clinicians, and therefore two sets of testing are expected. The level of analysis of the measure must align with testing; for this measure it was unclear whether testing was provided for both levels of analysis. ***One of these LoA may have to be dropped from the specifications, unless the developer can clarify how to interpret the testing results. Additional testing may be required if they would like the measure to be endorsed for both levels of analysis.***
   - Reliability testing limited to providers with 10 or more patients eligible for this measure – this minimum threshold is not included in the specifications.

### RELIABILITY: TESTING

**Submission document:** "MIF_xxxx" document for specifications, testing attachment questions 1.1-1.4 and section 2a2

3. **Reliability testing level** ☒ **Measure score** ☐ **Data element** ☐ **Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure** ☒ **Yes** ☐ **No**
5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical <u>VALIDITY</u> testing** of <u>patient-level data</u> conducted?

☐ **Yes** ☐ **No**

6. **Assess the method(s) used for reliability testing**

   **Submission document:** Testing attachment, section 2a2.2

   Reliability of the computed measure score was measured as the ratio of signal to noise and testing was performed by using a beta-binomial model.

7. **Assess the results of reliability testing**

   **Submission document:** Testing attachment, section 2a2.3

   A reliability equal to zero implies that all the variability in a measure is attributable to measurement error. A reliability equal to one implies that all the variability is attributable to real differences in physician performance. Testing results indicated that the reliability above the minimum level of quality reporting events (10) for 251 physicians was 0.97.

8. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? NOTE: If multiple methods used, at least one must be appropriate.

   **Submission document:** Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

   ☐ **Not applicable** (score-level testing was not performed)

9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?

   **Submission document:** Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

   ☒ **Not applicable** (data element testing was not performed)

10. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and all testing results):

    ☐ **High** (NOTE: Can be HIGH only if score-level testing has been conducted)

    ☒ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has not been conducted)

    ☐ **Low** (NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete or if testing methods/results are not adequate)

    ☐ **Insufficient** (NOTE: Should rate INSUFFICIENT if you believe you do not have the information you need to make a rating decision)

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**

    - Reliability score 0.97 high; however, unclear testing included both clinician groups and individual clinicians.

### VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

12. **Please describe any concerns you have with measure exclusions.**

    **Submission document:** Testing attachment, section 2b2.

    - N/A

13. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.**

    **Submission document:** Testing attachment, section 2b4.

- Data provided is repeated from performance gap and does not demonstrate meaningful differences in performance.

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
    **Submission document:** Testing attachment, section 2b5.
    - N/A
15. **Please describe any concerns you have regarding missing data.**
    **Submission document:** Testing attachment, section 2b6.
    - Empirical analysis required. Developer states that PQRS dataset provided by CMS did not contain missing data so analysis was not performed.
16. **Risk Adjustment**

    16a. **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

    16b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

    ☐ Yes ☐ No ☒ Not applicable

    16c. **Social risk adjustment:**

    16c.1 Are social risk factors included in risk model? ☐ Yes ☐ No ☒ Not applicable

    16c.2 Conceptual rationale for social risk factors included? ☐ Yes ☐ No

    16c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☐ Yes ☐ No

    16d.**Risk adjustment summary:**

    16d.1 All of the risk-adjustment variables present at the start of care? ☐ Yes ☐ No ☒ Not applicable

    16d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐ Yes ☐ No

    16d.3 Is the risk adjustment approach appropriately developed and assessed? ☐ Yes ☐ No

    16d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☐ Yes ☐ No

    16d.5.Appropriate risk-adjustment strategy included in the measure? ☐ Yes ☐ No

    16e**. Assess the risk-adjustment approach**

## VALIDITY: TESTING

17. **Validity testing level:** ☒ **Measure score** ☐ **Data element** ☐ **Both**
18. **Method of establishing validity of the measure score:**

    ☐ **Face validity**

    ☒ **Empirical validity testing of the measure score**

    ☐ **N/A (score-level testing not conducted)**
19. **Assess the method(s) for establishing validity**

    **Submission document: Testing attachment, section 2b2.2**

    - The developer performed a correlation analysis with measure: Oncology: Medical and Radiation – Plan of Care for Pain (PQRS #144) due to the the similarities in patient population and domain. This method can demonstrate an association between patients with a diagnosis of cancer receiving chemotherapy or radiation therapy in which pain intensity is quantified (NQF # 0384) and those with a diagnosis of cancer receiving chemotherapy or radiation therapy who report having pain with a documented plan of care to address pain (PQRS #144).

20. **Assess the results(s) for establishing validity**

    **Submission document: Testing attachment, section 2b2.3**

    - The developer reports a coefficient correlation of 0.69 (P-value = > 0.001).
    - The number of providers and patients included in the correlation analysis is not clear.

21. **Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?**

    **Submission document:** Testing attachment, section 2b1.

    ☒ **Yes**

    ☐ **No**

    ☐ **Not applicable** (score-level testing was not performed)

22. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?** *NOTE that data element validation from the literature is acceptable.*

    **Submission document***: Testing attachment, section 2b1.*

    ☐ **Yes**

    ☐ **No**

    ☒ **Not applicable** (data element testing was not performed)

23. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.**

    ☐ **High** (NOTE: Can be HIGH only if score-level testing has been conducted)

    ☒ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has NOT been conducted)

    ☐ **Low** (NOTE: Should rate LOW if you believe that there <u>are</u> threats to validity and/or relevant threats to validity were <u>not assessed OR</u> if testing methods/results are not adequate)

    ☐ **Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the score level and the data element level <u>is required</u>; if not conducted, should rate as INSUFFICIENT.)

24. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**

    - The developer provided additional information on January 14, 2019 regarding the number of provider and patients include din the correlation analysis. This information will be added to the measure testing attachment at a later date.
        - Number of shared providers included in analysis based on NPI and TIN identifiers = 111
        - Number of PQRS #143 (Pain Intensity Quantified) qualifying events included in analysis = 26,635
        - Number of PQRS #144 (Plan of Care) qualifying events included in analysis = 10,785

## ADDITIONAL RECOMMENDATIONS

25. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**

    - The Standing Committee should discuss the testing method used to demonstrate validity and determine if it is appropriate and if the testing results are sufficient.

**Committee Pre-evaluation Comments:**
**Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)**

**2a1. Reliability-Specifications: Which data elements, if any, are not clearly defined? Which codes with descriptors, if any, are not provided? Which steps, if any, in the logic or calculation algorithm or other specifications (e.g., risk/case-mix adjustment, survey/sampling instructions) are not clear? What concerns do you have about the likelihood that this measure can be consistently implemented?**

- Mixed level of analysis hard to compare
- concerns regarding capture of pain discussion from EHR
- I do not have any concerns about whether the measure can be consistently implemented.
- Same comments as placed for #0384e
- No concerns at this time
- Adequate as to asked or not, less clear about pain quantification
- Data elements are defined clearly
- Data elements are clearly defined. The specifications of the measure appear to be clear and complete. I have no concerns that this measure can be consistently implemented.
- I am concerned the measure can be implemented on a consistent basis.

**2a2. Reliability - Testing: Do you have any concerns about the reliability of the measure?**

- No
- as above
- I do not have any concerns.
- Same comments as placed for #0384e
- No concerns
- no
- No
- The reliability testing data from CMS from 2016 was presented. The testing for reliability and validity was based on 5 different sites. The overall reliability was 99.9%, with the numerator value of 99.9% and the denominator value of 100%. This was associated with a kappa score consistent with a high reliability. The reliability above the minimum level of quality reporting events was 0.97. There was no risk adjustment of testing for reliability. I do not have concerns about the reliability of the measure.
- No concerns

**2b1. Validity -Testing: Do you have any concerns with the testing results?**

- Satisfactory testing
- no concerns
- I do not have any concerns.
- Same comments as placed for #0384e
- No concerns
- yes
- No
- Validity testing included previous testing from the most recent submission which was based on face validity. An expert panel of 31 experts rated the validity as high (84.21% level 4 or 5) and felt that the measure had the ability to detect meaningful differences in performance. The updated submission included an expert review of ICD-10 codes to satisfy NQF's ICD-10 conversion

requirement. Further testing included a comparison to a companion measure (PQRS#144) which measures the documentation of a plan of care for pain. The coefficient of correlation = 0.69 with a P-value = > 0.001 suggesting a high degree of correlation. This is a reasonable evaluation and continues to suggest that the measure is valid.

- No concerns

**2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data)2b4. Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality? 2b5. Comparability of performance scores: If multiple sets of specifications: Do analyses indicate they produce comparable results? 2b6. Missing data/no response: Does missing data constitute a threat to the validity of this measure?**

- Acceptable
- EHR extraction of pain quantitation
- As far as I can tell, there appears to be no threat to the validity of this measure.
- Same comments as placed for #0384e
- No
- yes
- N/A
- The developers presented data that demonstrates that the range of performance among 5 different sites varied from 0.04-1. This suggests that there is evidence that the measure can detect meaningful differences in performance. No risk adjustment was done to analyze differences. No missing data was identified. There is no evidence to suggest that there are significant threats to validity.
- No input

**2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment)2b2. Exclusions: Are the exclusions consistent with the evidence? Are any patients or patient groups inappropriately excluded from the measure?2b3. Risk Adjustment: If outcome (intermediate, health, or PRO-based) or resource use performance measure: Is there a conceptual relationship between potential social risk factor variables and the measure focus? How well do social risk factor variables that were available and analyzed align with the conceptual description provided? Are all of the risk-adjustment variables present at the start of care (if not, do you agree with the rationale provided)? Was the risk adjustment (case-mix adjustment) appropriately developed and tested? Do analyses indicate acceptable results? Is an appropriate risk-adjustment strategy included in the measure?**

- No risk adjustment
- no
- Risk adjustment variables are not present at the start of care which makes it not applicable.
- Same comments as placed for #0384e
- Yes - appropriate strategy included
- No threats I see.
- N/A
- The developers presented data that demonstrates that the range of performance among 5 different sites varied from 0.04-1. This suggests that there is evidence that the measure can detect meaningful differences in performance. No risk adjustment was done to analyze differences. No missing data was identified. There is no evidence to suggest that there are significant threats to validity.
- No input

## Criterion 3. Feasibility

**Maintenance measures – no change in emphasis – implementation issues may be more prominent**

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

- The developer states that all data elements are in defined fields in a combination of electronic data sources.
- Data are generated and used by healthcare personnel during provision of care. Data are coded by someone other than the individual obtaining the original information.
- The developer reports no areas of concern or measure modification as a result of feasibility testing.
- The measure is copyrighted, but can be reproduced and distributed without modification for noncommercial purposes. Commercial use of the measure requires a license agreement between the user and the PCPI Foundation or the American Medical Association (AMA).

***Questions for the Committee:***

- Are the required data elements routinely generated and used during care delivery?
- Are the required data elements available in electronic form, e.g., EHR or other electronic sources?

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

**Committee Pre-evaluation Comments:**
**Criteria 3: Feasibility**

**3. Feasibility: Which of the required data elements are not routinely generated and used during care delivery? Which of the required data elements are not available in electronic form (e.g., EHR or other electronic sources)? What are your concerns about how the data collection strategy can be put into operational use?**

- It's feasible
- no concerns
- I don't have any concerns at this time.
- Same comments as placed for #0384e
- No concerns
- Limited data collection issues
- Metric is feasible in most EHRs due to the easy collection of the data through built in functionality
- The data elements required for the measure are routinely generated in the course of care. Many of the data elements are available in electronic form. This measure has been used for several years. I have no concerns that there would be difficulty implementing this measure for clinical use.
- No concerns or input on whether the data is routinely generated & used.

## Criterion 4: Usability and Use

**Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact/improvement and unintended consequences**

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

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

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

**Current uses of the measure**

**Publicly reported?** ☒ **Yes** ☐ **No**

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

**OR**

**Planned use in an accountability program?** ☐ **Yes** ☐ **No**

**Accountability program details**

- This measure is currently used in the Merit-based Incentive Payment System (MIPS). The measure was previously used in the Physician Quality Reporting System (PQRS).
- The measure is not currently publicly reported, but data will be available for public reporting in Physician Compare beginning in late 2019

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

**Feedback on the measure by those being measured or others**

- The developer received feedback from Joint Commission stating that standards already require accredited hospitals to establish policies and procedures that address comprehensive clinical assessment of pain.
    - The developer responded that while pain assessment has been adopted by the Joint Commission as a requirement for hospital accreditation, at least one-third of radiation therapy services are provided at free-standing centers and the majority of chemotherapy administration is provided in non-hospital settings. The PCPI measure is assessed at the physician level rather than the hospital level.
- The developer received recommendations to consider potential denominator exclusions.
    - The developer responded that the expert panel that designed the measure did so specifically without exclusions since addressing pain is such a critical aspect of care for all cancer patients.

**Additional Feedback:**

***Questions for the Committee:***

- How have 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**

**RATIONALE:**

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

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

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

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

**Improvement results**

- The developer did not discuss any progress on improvement.

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

**Unexpected findings (positive or negative) during implementation**

- The developer reports no unexpected findings during implementation

**Potential harms**

- The developer reports no potential harms for the measure.

**Additional Feedback:**

***Questions for the Committee:***

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

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

**RATIONALE:** The developer did not discuss any progress on improvement.

**Committee Pre-evaluation Comments:**
**Criteria 4: Usability and Use**

**4a1. Use - Accountability and Transparency: How is the measure being publicly reported? Are the performance results disclosed and available outside of the organizations or practices whose performance is measured? For maintenance measures - which accountability applications is the measure being used for? For new measures - if not in use at the time of initial endorsement, is a credible plan for implementation provided?4a2. Use - Feedback on the measure: Have those being measured been given performance results or data, as well as assistance with interpreting the measure results and data? Have those being measured or other users been given an opportunity to provide feedback on the measure performance or implementation? Has this feedback has been considered when changes are incorporated into the measure?**

- It's usable
- in use and publically reported. feedback considered.
- The measure provides greater benefit then harm.
- Same comments as placed for #0384e
- Feedback has been considered and incorporated
- Unsure
- Measure is used in multiple reporting programs

- The measure is already in use for MIPs and has been able to generate data suggesting that it is useable.
- No input on use of the performance results.

**4b1. Usability – Improvement: How can the performance results be used to further the goal of high-quality, efficient healthcare? If not in use for performance improvement at the time of initial endorsement, is a credible rationale provided that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations?4b2. Usability – Benefits vs. harms: Describe any actual unintended consequences and note how you think the benefits of the measure outweigh them.**

- Low risk of unintended consequences
- no potential harms. benefits less clear.
- The performance results can improve the goal of high quality.
- Same comments as placed for #0384e
- None
- Potentially very beneficial
- Documentation of pain may be used to identify patients for performance improvements on other measures
- The developers have identified potential harms associated with the quantification and reporting of pain which include the potential misuse or abuse of opioids and the need to assure that pain. Appropriate selection of analgesics and other interventions is necessary for the optimal outcome. I believe that the benefits of quantifying and documenting pain outweighs the risks.
- I believe the benefits outweigh any negative consequences of this measure.

## Criterion 5: Related and Competing Measures

**Related measures**

- 0177 : Improvement in pain interfering with activity
- 0192 : Residents who experience moderate to severe pain during the 7-day assessment period (risk-adjusted)
- 0420 : Pain Assessment and Follow-Up
- 0523 : Pain Assessment Conducted
- 0676 : Percent of Residents Who Self-Report Moderate to Severe Pain (Short Stay)
- 0677 : Percent of Residents Who Self-Report Moderate to Severe Pain (Long Stay)
- 1628 : Patients with Advanced Cancer Screened for Pain at Outpatient Visits
- 1637 : Hospice and Palliative Care -- Pain Assessment

**Harmonization**

- There are several NQF-endorsed measures related to this measure. Most related measures are assessed within different settings and at distinct levels of analysis.
- The developer notes that these measures are harmonized to the extent possible.

**Committee Pre-evaluation Comments: Criterion 5: Related and Competing Measures**

**5. Related and Competing: Are there any related and competing measures? If so, are any specifications that are not harmonized? Are there any additional steps needed for the measures to be harmonized?**

- Patients may be a subset of another measure
- interelated with a number of pain measures that could quite likely be grouped.
- No additional steps necessary at this time.
- Same comments as placed for #0384e
- No additional steps at this time.
- Similar measure to prior one from Medical Oncology - that one is is quantify pain, vs. care plan for this one.
- There are related measures, there does not appear to be overlap
- There are 8 related or competing measures and all will require harmonization.
- I do not have any concerns with related measures.

## Public and Member Comments

**Comments and Member Support/Non-Support Submitted as of: 02/14/2020**

- **Of the one NQF members who have submitted a support/non-support choice:**
  - One supports the measure:
    - "Thank you for the opportunity to provide comment on NQF Measure # 0384: Oncology: Medical and Radiation - Pain Intensity Quantified. We support the committee's recommendation for continued NQF endorsement, however we have a question that we hope the committee and measure developer will consider. Specifically, we are interested in understanding if the measure considers any role for cancer immunotherapy agents in the measure denominator. While patients may be treated with chemotherapy and immunotherapy in combination, some may be treated with immunotherapy only. In such cases, pain management for patients undergoing only cacner immunotherapy may be missed within this measure."

### ADDITIONAL RECOMMENDATIONS

26. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**

## 1. Evidence and Performance Gap – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***

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

NQF_evidence_attachment_0384_registry.docx

**1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission?**

Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

No

### 1a. Evidence (subcriterion 1a)

**Measure Number** (*if previously endorsed*)**:** 0384

**Measure Title**: Oncology: Medical and Radiation – Pain Intensity Quantified

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

**Date of Submission**: 4/13/2018

**1a.1.This is a measure of**: (*should be consistent with type of measure entered in De.1*)

Outcome

☐ Outcome: Click here to name the health outcome

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

*PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)*

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

☒ Process: Assessment and quantification of pain intensity at each visit for patients receiving chemotherapy or radiation therapy

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

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

**1a.2 LOGIC MODEL** Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

Initial and ongoing pain assessments, the focus of the measure, are essential to ensure proper pain

management among patients with cancer. "Failure to adequately assess pain frequently leads to poor control."(1) "Unrelieved pain denies [patients] comfort and greatly affects their activities, motivation, interactions with family and friends, and overall quality of life." (1)

(1) National Comprehensive Cancer Network (NCCN). Clinical Practice Guidelines in Oncology: Adult Cancer Pain. Version 2, 2011. Available at: http://www.nccn.org.

Updated guideline: There is increasing evidence in oncology that survival is linked to symptom reporting and control and that pain management contributes to broad quality-of-life improvement. To maximize patient outcomes, pain management is an essential part of oncologic management.

National Comprehensive Cancer Network (NCCN). NCCN Clinical Practice Guidelines in Oncology. Adult cancer pain Version I.2018. January 22, 2018. http://www.nccn.org

**1a.3 Value and Meaningfulness: IF** this measure is derived from patient report, provide evidence that the target population values the measured ***outcome, process, or structure*** and finds it meaningful. (Describe how and from whom their input was obtained.)

Not applicable

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

**1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES - Provide empirical data demonstrating the relationship between the outcome (or PRO) to at least one healthcare structure, process, intervention, or service.**

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

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

☒ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

| | |
|---|---|
| **Source of Systematic Review:**<br>• **Title**<br>• **Author**<br>• **Date**<br>• **Citation, including page number**<br>• **URL** | • **Title: Clinical Practice Guidelines in Oncology: Adult Cancer Pain. Version 2, 2011**<br>• **Author: National Comprehensive Cancer Network (NCCN).**<br>• **Date: 2011**<br>• **Citation: National Comprehensive Cancer Network (NCCN). Clinical Practice Guidelines in Oncology: Adult Cancer Pain. Version 2, 2011.**<br>• **URL: http://www.nccn.org**<br><br>• **Title: American Pain Society Recommendations for Improving the Quality of Acute and Cancer Pain Management**<br>• **Author: Gordon DB; Dahl JL, Miaskowski C, et al**<br>• **Date: July 25, 2005**<br>• **Citation: Gordon DB; Dahl JL, Miaskowski C, et al. American Pain Society Recommendations for Improving the Quality of Acute and Cancer Pain Management: American Pain Society Quality of Care Task Force. *Arch Intern Med.* 2005;165 (14):1574-1580.**<br>• **URL: https://jamanetwork.com/journals /jamainternalmedicine/fullarticle/486669**<br><br>**Updated 2018 NCCN guideline**<br>• **Title: NCCN Clinical Practice Guidelines in Oncology – Adult Cancer Pain**<br>• **Author: Swarm RA, Paice JA, Anghelescu DL, et al; NCCN Guidelines Panel**<br>• **Date: January 22, 2018**<br>• **Citation: National Comprehensive Cancer Center. Adult Cancer Pain: Version I.2018.2018: PAIN-1**<br>• **URL: http://www.nccn.org** |
| Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR. | **2011 Clinical Practice Guidelines in Oncology: Adult Cancer Pain**<br><br>This algorithm begins with the premise that all patients with cancer should be screened for pain during the initial evaluation, at regular intervals, and whenever new therapy is initiated. If pain is present |

| | |
|---|---|
| | on a screening evaluation, the pain intensity must be quantified, by the patient (whenever possible). Since pain is inherently subjective, patient's self-report to pain is the current standard of care for assessment. Intensity of pain should be quantified using a 0-10 numerical rating scale, a categorical scale, or a pictorial scale (e.g., The Faces Pain Rating Scale). The Faces Pain Rating Scale may be successful with patients who have difficulty with other scales, for example, children, the elderly, and patients with language or cultural differences or other communication barriers.<br><br>**2005 American Pain Society Recommendations for Improving the Quality of Acute and Cancer Pain Management**<br><br>All patients should be routinely screened for pain, and when it is present, pain intensity should be recorded in highly visible ways that facilitate regular review by health care providers. A standard for pain assessment and documentation should be established in each setting to ensure that pain is recognized, documented, and treated promptly.<br><br>**2018 NCCN Clinical Practice Guidelines in Oncology – Adult Cancer Pain**<br><br>• All patients must be screened for pain at each contact.<br>• Pain intensity must be routinely quantified and documented, and quality must be characterized by the patient (whenever possible based on patient communication capacity). Also include patient reporting of breakthrough pain, treatments used and their impact on pain, patient reporting of adequate comfort, patient reporting of satisfaction with pain relief, provider assessment of impact on function, and any special issues for the patient relevant to pain treatment. If necessary, get additional information from |

| | |
|---|---|
| | family/caregiver regarding pain and impact on function.<br>• Comprehensive pain assessment must be performed if new or worsening pain is present and regularly performed for persisting pain. |
| Grade assigned to the **evidence** associated with the recommendation with the definition of the grade | **2011 Clinical Practice Guidelines in Oncology: Adult Cancer Pain:** Category 2A; Based upon lower-level evidence, there is uniform NCCN consensus that the intervention is appropriate.<br><br>**2005 American Pain Society Recommendations for Improving the Quality of Acute and Cancer Pain Management:** The body of evidence in the APS guideline has not been graded. However, the APS indicates that recommendations result from literature reviews, expert experience, and consensus.<br><br>**2018 NCCN Clinical Practice Guidelines in Oncology – Adult Cancer Pain**<br>Category 2A; Based upon lower-level evidence, there is uniform NCCN consensus that the intervention is appropriate |
| Provide all other grades and definitions from the evidence grading system | NCCN Categories of Evidence and Consensus:<br><br>• Category 1: Based upon high-level evidence, there is uniform NCCN consensus that the intervention is appropriate.<br>• Category 2A: Based upon lower-level evidence, there is uniform NCCN consensus that the intervention is appropriate.<br>• Category 2B: Based upon lower-level evidence, there is NCCN consensus that the intervention is appropriate.<br>• Category 3: Based upon any level of evidence, there is major NCCN disagreement that the intervention is appropriate.<br><br>The body of evidence in the APS guideline has not been graded. However, the APS indicates that recommendations result from literature reviews, expert experience, and consensus. |

| | |
|---|---|
| Grade assigned to the **recommendation** with definition of the grade | **2011 Clinical Practice Guidelines in Oncology: Adult Cancer Pain:** Category 2A; Based upon lower-level evidence, there is uniform NCCN consensus that the intervention is appropriate.<br><br>**2005 American Pain Society Recommendations for Improving the Quality of Acute and Cancer Pain Management:** The body of evidence in the APS guideline has not been graded. However, the APS indicates that recommendations result from literature reviews, expert experience, and consensus.<br><br>**2018 NCCN Clinical Practice Guidelines in Oncology – Adult Cancer Pain**<br>Category 2A; Based upon lower-level evidence, there is uniform NCCN consensus that the intervention is appropriate |
| Provide all other grades and definitions from the recommendation grading system | NCCN Categories of Evidence and Consensus:<br><br>• Category 1: Based upon high-level evidence, there is uniform NCCN consensus that the intervention is appropriate.<br>• Category 2A: Based upon lower-level evidence, there is uniform NCCN consensus that the intervention is appropriate.<br>• Category 2B: Based upon lower-level evidence, there is NCCN consensus that the intervention is appropriate.<br>• Category 3: Based upon any level of evidence, there is major NCCN disagreement that the intervention is appropriate.<br><br>The body of evidence in the APS guideline has not been graded. However, the APS indicates that recommendations result from literature reviews, expert experience, and consensus. |
| Body of evidence:<br>• Quantity – how many studies?<br>• Quality – what type of studies? | • **Quantity:** The description of the evidence review in the NCCN guideline did not address the overall quantity of studies in the body of evidence. However, 105 articles are cited.<br><br>The 2018 NCCN guideline also does not provide a description of the body of evidence. However, the |

| | |
|---|---|
| | overview and pain assessment section within the guideline cites 36 articles in support of the pain assessment concept and recommendations.<br><br>Similarly, the description of the evidence review in the APS guideline did not address the overall quantity of studies in the body of evidence. However, 82 articles are cited.<br><br>• **Quality:** The quality of the body of evidence supporting the NCCN guideline recommendations are summarized according to the NCCN categories of evidence and consensus as being based on "lower-level evidence". Lower-level evidence is later described as evidence that may include non-randomized trials; case series; or when other data are lacking, the clinical experience of expert physicians.<br><br>The quality of the body of evidence supporting the APS guideline recommendation is not provided. |
| Estimates of benefit and consistency across studies | Although there is no explicit statement regarding the overall consistency of results across studies in the NCCN guidelines supporting the measure, the recommendation received uniform NCCN consensus that the intervention is appropriate and initial and ongoing pain assessments are essential to ensure proper pain management. |
| What harms were identified? | Initial and ongoing pain assessments are essential to ensure proper pain management.<br><br>The NCCN guideline states that "…attempts should be made to determine the underlying pain mechanism and diagnose the pain syndrome." The selection of analgesic for optimal outcomes will depend on the intensity of pain, current analgesic therapy and existing comorbidities.<br><br>The NCCN acknowledges the potential for misuse and abuse of opioids and states that initial patient evaluation should include the routine assessment of |

| | |
|---|---|
| | risk factors for aberrant use of pain medications by detailed patient evaluation and/or the use of screening tools. The guideline also recommends monitoring for aberrant drug-taking behaviors or evidence of diversion. |
| Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR? | |

**1a.4 OTHER SOURCE OF EVIDENCE**

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

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

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

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

## 1b. Performance Gap

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

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

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

*If a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.*

An estimated 1.7 million new cases of cancer are diagnosed in the US each year. (1) Pain is a commonly occurring symptom for cancer patients as 30% to 50% (510,000 to 850,000 each year based on current statistics) will experience moderate to severe pain.(2) Initial and ongoing pain assessments are essential to determine the pathophysiology of pain and ensure proper pain management. According to the National Comprehensive Cancer Network, there is increasing evidence in oncology that survival is linked to symptom reporting and control and that pain management contributes to broad quality-of-life improvement.(3) Evidence has shown a positive association between higher symptom scores and higher rates of documentation and clinical actions taken. (4) A study published this year (2019) provides further evidence that symptom monitoring following treatment for cancer is associated with increased survival. (5) Cancer patients have reported that pain interferes with their mood, work, relationships with other people, sleep and overall enjoyment of life.(6) To maximize patient outcomes, pain management is an essential part of oncologic management.(3)

(1) National Cancer Institute. Cancer statistics. National Institutes of Health. 2017. https://www.cancer.gov/about-cancer/understanding/statistics

(2) Wiffen PJ, Wee B, Derry S, Bell RF, Moore RA. Opioids for cancer pain – an overview of Cochrane reviews. Cochrane Database of Systematic Reviews 2017, Issue 7. Art. No.: CD012592. DOI: 10.1002/14651858.CD012592.pub2.

(3) National Comprehensive Cancer Network (NCCN). NCCN Clinical Practice Guidelines in Oncology. Adult cancer pain Version I.2018. January 22, 2018. http://www.nccn.org

(4) Seow H, Sussman J, Martelli-Reid L, Pond G, Bainbridge D. Do high symptom scores trigger clinical actions? An audit after implementing electronic symptom screening. J Oncol Pract. 2012 Nov;8(6):e142-8. doi: 10.1200/JOP.2011.000525.

(5) Denis F, Basch E, Septans AL, Bennouna J, Urban T, Dueck AC, Letellier C. Two-Year Survival Comparing Web-Based Symptom Monitoring vs Routine Surveillance Following Treatment for Lung Cancer. JAMA. 2019 Jan 22;321(3):306-307. doi: 10.1001/jama.2018.18085.

(6) Moryl N, Dave V, Glare P, Bokhari A, Malhotra VT, Gulati A, et al. Patient-reported outcomes and opioid use by outpatient cancer patients. J Pain. 2018. Mar;19(3):278-290.

**1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis**. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

CMS provided 2016 PQRS reporting data for analysis. Based on the 251 included physicians from 2016 PQRS registry reporting, the mean performance rate for this measure is 0.88 the median performance rate is 0.98 and the mode is 1.0. The standard deviation is 0.21. The range of the performance rate is 0.96, with a minimum rate of 0.04 and a maximum rate of 1.0. The interquartile range is 0.12 (1.0–0.88).

The CMS PQRS Experience report provides these additional average performance rates for previous years:

Average performance rate:

2015: 75.9%

2014: 84.8%

2013: 82.7%

It is important to note that both PQRS and now the Merit-based Incentive Payment System (MIPS), have been and remain a voluntary reporting program. Participants are allowed to self-select measures and may choose those that will result in high performance rates. As a result, performance rates may not be nationally representative.

**1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**

Despite the availability of clinical guidelines and increased attention on its assessment and management, pain continues to be a commonly occurring symptom in patients with cancer and may not be appropriately assessed.

One retrospective analysis sought to evaluate compliance with pain assessment recommendations from the NCCN guidelines. The study found an 84% compliance with pain intensity documentation at hospital admission (1). Furthermore, the study found that pain characteristics were documented in 69% of patients at pain onset and pain reassessed only 43% of the time after opioids were administered.(1) These results suggest that pain is not assessed appropriately and therefore not optimally managed. Based on the current prevalence of cancer-related pain, assessment is essential to decrease the impact of pain and its sequelae.

Two meta-analyses confirmed that the prevalence of cancer pain has not changed significantly in the past decades.

The meta-analysis published in 2007 reported significant rates of pain in cancer patients:

- 64% in advanced, metastatic or terminal disease
- 59% during anticancer treatment
- 33% after curative cancer treatment

The analysis found that 1/3 of patients graded their pain as moderate or severe. Patients with head/neck cancer had the highest prevalence of pain (70%), followed by gastrointestinal cancer (59%), gynecological (60%), lung/bronchus (55%), breast (54%), and urogenital (52%).(2)

An updated systematic review and meta-analysis was published in 2016 and reported prevalence of pain:

- 66% in advanced, metastatic or terminal disease
- 55% during anticancer treatment
- 39% after curative cancer treatment

This analysis found that 38% or all patients graded their pain as moderate to severe. A higher pain prevalence was associated with lung, gastrointestinal, head and neck, and breast cancer.(3)

A recent analysis of registry data for cancer patients with pain found average pain intensity reported as mild (24.6% of patients), moderate (41.5%), and severe (33.9%). The study also indicated that patient report of pain relief is inversely related to the average pain intensity reported.(4) These data suggest that assessing and managing a cancer patient's pain is critical and there remains significant room for improvement in assessing and mitigating cancer-related pain.

A prospective study of changes in pain severity of cancer patients found that, at initial assessment, 47% of patients reported pain. At follow-up, the patients with pain at initial assessment reported reduced pain (32.2%), stable pain (48.2%) and worse pain (19.6%). Of the 53% of patients reporting no pain at initial assessment, 82.6% reported stable pain and 17.4% reported worse pain at follow-up assessment.(4) This study highlights the importance of initial and ongoing assessments of pain to identify gaps and ensure proper pain management.

(1) El Rahi C, Murillo JR, Zaghloul H. Pain assessment practices in patients with cancer admitted to the oncology floor. J Hematol Oncol Pharm. 2017;7(3):109-113.

(2) van den Beuken-van Everdingen MH, de Rijke JM, Kessels AG, Schouten HC, van Kleef M, Patijn J. Prevalence of pain in patients with cancer: a systematic review of the past 40 years. Ann Oncol. 2007 Sep;18(9):1437-49.

(3) van den Beuken-van Everdingen MH, Hochstenbach LM, Joosten EA, Tjan-Heijnen VC, Janssen DJ. Update on prevalence of pain in patients With cancer: systematic review and meta-analysis. J Pain Symptom Manage. 2016 Jun;51(6):1070-1090.

(4) Moryl N, Dave V, Glare P, Bokhari A, Malhotra VT, Gulati A, et al. Patient-reported outcomes and opioid use by outpatient cancer patients. J Pain. 2018. Mar;19(3):278-290.

(5) Zhao F, Chang VT, Cleeland C, Cleary JF, Mitchell EP, Wagner LI, Fisch MJ. Determinants of pain severity changes in ambulatory patients with cancer: an analysis from Eastern Cooperative Oncology Group trial E2Z02. J Clin Oncol. 2014.Feb 1;32(4):312-9.

**1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

While this measure is included in federal reporting programs, those programs have not yet made disparities data available for us to analyze and report.

**1b.5. If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4**

A number of studies have documented disparities in cancer treatment and in the management of cancer-related pain.(1-3). One prospective observational study of patients with breast, prostate, colon/rectum, or lung cancer found that 67% reported pain upon initial assessment. An estimated 33% of patients did not receive adequate analgesic treatment. Furthermore, the study results estimate that the odds of inadequate pain treatment for non-Hispanic whites were half those of minorities.(4)

Another study highlights the importance of ongoing pain assessments in cancer patients, and particularly in minorities, to determine pain intensity at follow up. Despite analgesic use at initial assessment, Hispanics had 3.4 times higher odds of moderate to severe pain at follow-up as compared to Whites. The study also reports that as compared to Whites, Blacks had 2 times higher odds of moderate to severe pain at follow-up.

(1) Gorin SS, Heck JE, Cheng B, Smith SJ. Delays in breast cancer diagnosis and treatment by racial/ethnic group. Arch Intern Med. 2006 Nov 13;166(20):2244-52.

(2) Payne R, Medina E, Hampton JW. Quality of life concerns in patients with breast cancer: evidence for disparity of outcomes and experiences in pain management and palliative care among African-American women. Cancer. 2003 Jan 1;97(1 Suppl):311-7.

(3) Anderson KO, Green CR, Payne R. Racial and ethnic disparities in pain: causes and consequences of unequal care. J Pain. 2009 Dec;10(12):1187-204.

(4) Fisch MJ, Lee JW, Weiss M, Wagner LI, Chang VT, Cella D, Manola JB, Minasian LM, McCaskill-Stevens W, Mendoza TR, Cleeland CS. Prospective, observational study of pain and analgesic prescribing in medical oncology outpatients with breast, colorectal, lung, or prostate cancer. J Clin Oncol. 2012 Jun 1;30(16):1980-8.

(5) Zhao F, Chang VT, Cleeland C, Cleary JF, Mitchell EP, Wagner LI, Fisch MJ. Determinants of pain severity changes in ambulatory patients with cancer: an analysis from Eastern Cooperative Oncology Group trial E2Z02. J Clin Oncol. 2014.Feb 1;32(4):312-9.

## 2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

**2a.1. Specifications** The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

**De.5. Subject/Topic Area** *(check all the areas that apply):*

Cancer

**De.6. Non-Condition Specific***(check all the areas that apply):*

Person-and Family-Centered Care

**De.7. Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*

Elderly

**S.1. Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

The measure specifications are included with this form. Additional measure details may be found at: http://www.thepcpi.org/?page=PCPIMeasures

**S.2a. <u>If this is an eMeasure</u>**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)

This is not an eMeasure **Attachment:**

**S.2b. Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*

Attachment **Attachment:** NQF0384_I9toI10_conversion_2018Nov.xlsx

**S.2c.** Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

No, this is not an instrument-based measure **Attachment:**

**S.2d.** Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

Not an instrument-based measure

**S.3.1. <u>For maintenance of endorsement:</u>** Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

Yes

**S.3.2. <u>For maintenance of endorsement,</u>** please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.

Beginning with 2019 implementation, the measure was revised to have two submission criteria: 1.) All patient visits for patients with a diagnosis of cancer currently receiving chemotherapy OR 2.) All patient visits for patients with a diagnosis of cancer currently receiving radiation therapy. This change was made to more clearly delineate the denominator requirements to promote accurate implementation. Based on feedback we heard regarding how vendors have implemented the measure, there was an inconsistent approach to applying the measure criteria. Therefore, we decided to split this measure out into two populations, based on the type of treatment the patient is receiving, which can be implemented in both the eCQM and registry versions of this measure. Though the measure is split into two, the measure still requires only one performance rate for reporting.

**S.4. Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*

*<u>IF an OUTCOME MEASURE</u>, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).*

Patient visits in which pain intensity is quantified

**S.5. Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

*<u>IF an OUTCOME MEASURE,</u> describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).*

Time Period for Data Collection: At each visit within the measurement period

Guidance: Pain intensity should be quantified using a standard instrument, such as a 0-10 numerical rating scale, visual analog scale, a categorical scale, or pictorial scale. Examples include the Faces Pain Rating Scale and the Brief Pain Inventory (BPI).

The Oncology: Medical and Radiation - Pain Intensity Quantified measure is specified for both registry (this measure) and for EHR (NQF #384e) implementation. The registry version has two submission criteria to capture 1) patients undergoing chemotherapy and 2) patients undergoing radiation therapy, and to align with the specifications for the EHR version of this measure.

For the Submission Criteria 1 and Submission Criteria 2 numerators, report one of the following CPT Category II codes to submit the numerator option for patient visits in which pain intensity was quantified:

1125F: Pain severity quantified; pain present

OR

1126F: Pain severity quantified; no pain present

**S.6. Denominator Statement** *(Brief, narrative description of the target population being measured)*

All patient visits, regardless of patient age, with a diagnosis of cancer currently receiving chemotherapy or radiation therapy

**S.7. Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

*IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).*

Time Period for Data Collection: 12 consecutive months

The registry version has two submission criteria to capture 1) patients undergoing chemotherapy and 2) patients undergoing radiation therapy, and to align with the specifications for the EHR version of this measure.

Guidance: For patients receiving radiation therapy, pain intensity should be quantified at each radiation treatment management encounter where the patient and physician have a face-to-face interaction. Due to the nature of some applicable coding related to the radiation therapy (eg, delivered in multiple fractions), the billing date for certain codes may or may not be the same as the face-to-face encounter date. For patients receiving chemotherapy, pain intensity should be quantified at each face-to-face encounter with the physician while the patient is currently receiving chemotherapy. For purposes of identifying eligible encounters, patients "currently receiving chemotherapy" refers to patients administered chemotherapy within 30 days prior to the encounter AND administered chemotherapy within 30 days after the date of the encounter.

Submission Criteria 1 denominator: Patient visits for patients with a diagnosis of cancer currently receiving chemotherapy

Diagnosis for cancer (ICD-10-CM) - Due to character limitation, please see codes in the attached Excel file in S.2b.

AND

Patient encounter during the performance period (CPT) – to be used to evaluate remaining denominator criteria and for numerator evaluation: 99201, 99202, 99203, 99204, 99205, 99212, 99213, 99214, 99215

WITHOUT

Telehealth Modifier: GQ, GT, 95, POS 02

AND

Patient procedure within 30 days before denominator eligible encounter: 51720, 96401, 96402, 96405, 96406, 96409, 96411, 96413, 96415, 96416, 96417, 96420, 96422, 96423, 96425, 96440, 96446, 96450, 96521, 96522, 96523, 96542, 96549

AND

Patient procedure within 30 days after denominator eligible encounter: 51720, 96401, 96402, 96405, 96406, 96409, 96411, 96413, 96415, 96416, 96417, 96420, 96422, 96423, 96425, 96440, 96446, 96450, 96521, 96522, 96523, 96542, 96549

Submission Criteria 2 denominator: Patient visits for patients with a diagnosis of cancer currently receiving radiation therapy

DENOMINATOR NOTE: For the reporting purposes for this measure, in instances where CPT code 77427 is reported, the billing date, which may or may not be the same date as the face-to-face encounter with the physician, should be used to pull the appropriate patient population into the denominator. It is expected, though, that the numerator criteria would be performed at the time of the actual face-to-face encounter during the series of treatments.

Diagnosis for cancer (ICD-10-CM) - Due to character limitation, please see codes in the attached Excel file in S.2b.

AND

Patient procedure during the performance period (CPT) – Procedure codes: 77427, 77431, 77432, 77435

**S.8. Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*

None

**S.9. Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

Not applicable

**S.10. Stratification Information** *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*

Consistent with the CMS Measures Management System Blueprint and recent national recommendations put forth by the IOM and NQF to standardize the collection of race and ethnicity data, we encourage the results of this measure to be stratified by race, ethnicity, administrative sex, and payer.

**S.11. Risk Adjustment Type** (Select type. Provide specifications for risk stratification in measure testing attachment)

No risk adjustment or risk stratification

If other:

**S.12. Type of score:**

Rate/proportion

If other:

**S.13. Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*

Better quality = Higher score

**S.14. Calculation Algorithm/Measure Logic** (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*

This measure is comprised of two submission criteria but is intended to result in one reporting rate. The reporting rate is the aggregate of Submission Criteria 1 and Submission Criteria 2, resulting in a single performance rate. For the purposes of this measure, the single performance rate can be calculated as follows:

Performance Rate = (Numerator 1 + Numerator 2)/ (Denominator 1 + Denominator 2)

Calculation algorithm for Submission Criteria 1: Patient visits for patients with a diagnosis of cancer currently receiving chemotherapy

1. Find the patients who meet the initial population (ie, the general group of patients that a set of performance measures is designed to address).

2. From the patients within the initial population criteria, find the patients who qualify for the denominator (ie, the specific group of patients for inclusion in a specific performance measure based on defined criteria). Note: in some cases the initial population and denominator are identical.

3. From the patients within the denominator, find the patients who meet the numerator criteria (ie, the group of patients in the denominator for whom a process or outcome of care occurs). Validate that the number of patients in the numerator is less than or equal to the number of patients in the denominator

If the patient does not meet the numerator, this case represents a quality failure.

Calculation algorithm for Submission Criteria 2: Patient visits for patients with a diagnosis of cancer currently receiving radiation therapy

1. Find the patients who meet the initial population (ie, the general group of patients that a set of performance measures is designed to address).

2. From the patients within the initial population criteria, find the patients who qualify for the denominator (ie, the specific group of patients for inclusion in a specific performance measure based on defined criteria). Note: in some cases the initial population and denominator are identical.

3. From the patients within the denominator, find the patients who meet the numerator criteria (ie, the group of patients in the denominator for whom a process or outcome of care occurs). Validate that the number of patients in the numerator is less than or equal to the number of patients in the denominator

If the patient does not meet the numerator, this case represents a quality failure.

**S.15. Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*

IF an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.

Not applicable. The measure does not require sampling or a survey.

**S.16. Survey/Patient-reported data** *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*

Specify calculation of response rates to be reported with performance measure results.

Not applicable

**S.17. Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*

*If other, please describe in S.18.*

Registry Data

**S.18. Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data are collected.)*

IF instrument-based, identify the specific instrument(s) and standard methods, modes, and languages of administration.

**S.19. Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

No data collection instrument provided

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

Clinician : Group/Practice, Clinician : Individual

**S.21. Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*

Other, Outpatient Services

If other: Oncology/Outpatient Clinic; Radiation Oncology Dept/Clinic

**S.22. COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*

Not applicable. The measure is not a composite.

**2. Validity – See attached Measure Testing Submission Form**

0384_nqf_testing_attachment_registry_7.1-636687217586704045.docx

**2.1 For maintenance of endorsement**

*Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.*

Yes

**2.2 For maintenance of endorsement**

*Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.*

Yes

**2.3 For maintenance of endorsement**

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

No - This measure is not risk-adjusted

## Measure Testing (subcriteria 2a2, 2b1-2b6)

**Measure Number** (*if previously endorsed*)**:** 0384
**Measure Title**: Oncology: Pain Intensity Quantified – Medical Oncology and Radiation Oncology
**Date of Submission**: 8/1/2018

**Type of Measure:**

| | |
|---|---|
| ☐ **Outcome (*including PRO-PM*)** | ☐ **Composite – *STOP – use composite testing form*** |

| | |
|---|---|
| ☐ Intermediate Clinical Outcome | ☐ Cost/resource |
| ☒ Process *(including Appropriate Use)* | ☐ Efficiency |
| ☐ Structure | |

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

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

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

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

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

The data source is 2016 Registry data from the PQRS program, provided by the Center for Medicare & Medicaid Services (CMS).

**1.3. What are the dates of the data used in testing**?

The measurement period (data collected from patients seen) was 1/1/2010 through 12/31/2010. Chart abstraction was performed between 8/8/2011 and 11/3/2011.

The data are for the time period January 2016 through December 2016 and cover the entire United States. Given the required conversion to ICD-10 in late 2015, the testing was completed on the ICD-10 specified measure.

**1.4. What levels of analysis were tested**? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

| **Measure Specified to Measure Performance of:** | **Measure Tested at Level of:** |
|---|---|

| (*must be consistent with levels entered in item S.20*) | |
|---|---|
| ☒ individual clinician | ☒ individual clinician |
| ☒ group/practice | ☒ group/practice |
| ☐ hospital/facility/agency | ☐ hospital/facility/agency |
| ☐ health plan | ☐ health plan |
| ☐ other: Click here to describe | ☐ other: Click here to describe |

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

Data abstracted from patient records were used to calculate inter-rater reliability for the measure.

862 patient visits were reviewed.

We received data from 371 physicians reporting on this measure through the registry option for CMS's PQRS in 2016. Of those, 251 physicians had all the required data elements and met the minimum number of quality reporting events (10) for a total of 44,795 quality events. For this measure, 73 percent of physicians are included in the analysis, and the average number of quality reporting events are 179 for the remaining 44,795 events. The range of quality reporting events for 251 physicians included is from 10 to 3,811. The average number of quality reporting events for the remaining 27 percent of physicians that aren't included is 4.

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

There were 44,795 patients included in this reliability testing and analysis. These were the patients that were associated with physicians who had 10 or more patients eligible for this measure.

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

Five practice sites representing various types, locations and sizes were identified to participate in testing the PCPI/ASCO/ASTRO-developed measures.

- Site A: hospital, multi-practice sites in urban, rural and suburban settings; 21 physicians; average 9600 oncology/prostate cancer patient visits per month for MD/NP assessment, chemotherapy; submitted PQRS claims for one measure and utilized a full-fledged EHR.
- Site B: physician owned private practice, suburban setting; 4 physicians; average 48 oncology/prostate cancer patients seen per day; submitted PQRS claims for one measure and utilized paper medical records.
- Site C: physician owned private practice, urban setting; 41 physicians; average 2500 oncology/prostate cancer patients seen per month; submitted PQRS claims for two measures and utilized a full-fledged EHR.

- Site D: academic, suburban setting; 9 physicians; average 240 oncology/prostate cancer patients seen per month; submitted PQRS claims for one measure and utilized paper and EHR.
- Site E: academic, urban setting; 14 physicians; average 250 oncology/prostate cancer patients seen per month; collected PQRS data on 3 measures and utilized a full-fledged EHR.

The same data samples were used for reliability testing and exceptions analysis.

**1.8 What were the social risk factors that were available and analyzed**? For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

Patient-level socio-demographic (SDS) variables were not captured as part of the testing.

---

**2a2. RELIABILITY TESTING**

***Note****: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

**2a2.1. What level of reliability testing was conducted**? (*may be one or both levels*)
☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)
☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

**2a2.2. For each level checked above, describe the method of reliability testing and what it tests** (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Data analysis included:

- Percent agreement; and
- Kappa statistic to adjust for chance agreement.

Reliability of the computed measure score was measured as the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in physician performance and the noise is the total variability in measured performance. Reliability at the level of the specific physician is given by:

Reliability = Variance (physician-to-physician) / [Variance (physician-to-physician) + Variance (physician-specific-error]

Reliability is the ratio of the physician-to-physician variance divided by the sum of the physician-to-physician variance plus the error variance specific to a physician. A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to real differences in physician performance.

Reliability testing was performed by using a beta-binomial model. The beta-binomial model assumes the physician performance score is a binomial random variable conditional on the physician's true value that

comes from the beta distribution. The beta distribution is usually defined by two parameters, alpha and beta. Alpha and beta can be thought of as intermediate calculations to get to the needed variance estimates.

Reliability is evaluated by averaging over physician specific reliabilities for all providers that meet the minimum number of quality reporting events for the measure. Each provider must have at least 10 eligible reporting events to be included in this calculation.

A reliability equal to zero implies that all the variability in a measure is attributable to measurement error. A reliability equal to one implies that all the variability is attributable to real differences in physician performance. A reliability of 0.70 – 0.80 is generally considered the acceptable threshold for reliability, 0.80 – 0.90 is considered high reliability, and 0.90 – 1.0 is considered very high. [1]

1. Adams JL, Mehrotra A, McGlynn EA, Estimating Reliability and Misclassification in Physician Profiling, Santa Monica, CA: RAND Corporation, 2010. www.rand.org/pubs/technical_reports/TR863. (Accessed on February 24, 2012.)

**2a2.3. For each level of testing checked above, what were the statistical results from reliability testing**? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

**N, % Agreement, Kappa (95% Confidence Interval)**

Overall Reliability: 862, **99.9%,** 0.990 (0.970-1.000)

Denominator Reliability: 862, **100.0%,** Kappa is noncalculable*

Numerator Reliability: 862, **99.9%,** 0.990 (0.970-1.000)

*Kappa Statistics cannot be calculated because of complete agreement. Confidence intervals cannot be calculated because to do so would involve dividing by zero which cannot be done.

The reliability above the minimum level of quality reporting events was 0.97.

**2a2.4 What is your interpretation of the results in terms of demonstrating reliability**? (i.*e., what do the results mean and what are the norms for the test conducted?*)

This measure demonstrates almost perfect reliability, as shown in results from the above analysis.

This measure has very high reliability when evaluated above the minimum level of quality reporting events.

---

**2b1. VALIDITY TESTING**

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

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

☒ **Performance measure score**

☒ **Empirical validity testing**

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

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

All PCPI performance measures are assessed for content validity by a panel of expert work group members during the development process. Additional input on the content validity of draft measures is obtained through a 30-day public comment period and by also soliciting comments from a panel of consumer, purchaser, and patient representatives convened by the PCPI specifically for this purpose. All comments received are reviewed by the expert work group and the measures adjusted as needed. Other external review groups (eg, focus groups) may be convened if there are any remaining concerns related to the content validity of the measures.

The expert panel was used to assess face validity of the measure. This panel consisted of 31 members, with representation from the following specialties: oncology, radiation oncology, surgical oncology, urology, gastroenterology, hematology, pathology, colon and rectal surgery, otolaryngology, and pain medicine.

The aforementioned panel was asked to rate their agreement with the following statement:

> The scores obtained from the measure as specified will accurately differentiate quality across providers.

The scale was 1-5, where 1=Strongly Disagree; 3=Neither Disagree nor Agree; 5=Strongly Agree.

The expert panel consists of 31 members, whose specialties include oncology, radiation oncology, surgical oncology, urology, gastroenterology, hematology, pathology, colon and rectal surgery, otolaryngology, and pain medicine.

The panel members are as follows:

Patricia Ganz, MD (Co-Chair) (Clinical Oncology) University of California – Los Angeles, Los Angeles, CA

James Hayman, MD (Co-Chair) (Radiation Oncology) American Society for Therapeutic Radiology and Oncology (ASTRO), Ann Arbor MI

Joseph Bailes, MD (Clinical Oncology) A Society for Clinical Oncology, The Woodlands, TX

Nancy Baxter, MD, PhD (Colorectal Surgery) American Society of Colon and Rectal Surgery Toronto, Ontario Canada

Joel V. Brill, MD (Gastroenterology) AGA, Phoenix, AZ

Steven B. Clauser, PhD (Outcomes Research) National Cancer Institute, Bethesda, MD

Charles Cleeland, PhD (Oncology) McCullough Professor of Cancer Research, Houston, TX

J. Thomas Cross, Jr. MD, MPH (Oncology) Colorado Spring, CO

Chaitanya R. Divgi, MD (Nuclear Medicine) Professor of Radiology & Chief University of Pennsylvania, Philadelphia, PA

Stephen B. Edge, MD (Surgical Oncology) Roswell Park Cancer Institute, Buffalo, NY

Patrick L. Fitzgibbons, MD (Oncology) St Jude Medical Center, Fullerton, CA

Myron Goldsmith, MD (Oncology) Huntington Beach, CA

Joel W. Goldwein, MD (Oncology) IMPAC Medical Systems, Inc., Merion Station, PA

Alecia Hathaway, MD, MPH (Oncology) Fort Worth, TX

Kevin P. Hubbard, DO (Oncology) Kansas City, MO

Nora Janjan, MD, MPSA (Radiation Oncology) University of Texas, Houston, TX

Maria Kelly, MB, BCh (Radiation Oncology) ASTRO, Earlysville, VA

Wayne Koch, MD (Head and Neck surgery) American Academy of Otolaryngology, Columbia, MD

Andre Konski, MD (Radiation Oncology) Fox Chase Cancer Center, Philadelphia, PA

Len Lichtenfeld, MD (Oncology) Deputy Chief Medical Officer, American Cancer Society, Atlanta, GA

Norman J. Marcus, MD (Anesthesiology and Psychiatry) New York University, New York, NY

Catherine Miyamoto, RN, BSN (Oncology) Cancer Center of North Dakota, Grand Forks, ND

Michael Neuss, MD (Oncology, Hematology) Oncology Hematology Care, Inc., Cincinnati, OH

David F. Penson, MD, MPH (Urology) Associate Professor of Urology and Preventive Medicine, Vanderbilt University Medical Center, Nashville, TN

Louis Potters, MD (Radiation Oncology) Chairman of Radiation Medicine, North Shore-NIJ, New Hyde Park, NY

John M. Rainey, MD (Medical Oncology) ASCO, Lafayette, LA

Christopher M. Rose, MD (Radiation Therapy) Radiation Therapy Center – Beverly Hills, El Segundo, CA

Lee Smith, MD (Oncology) Washington Hospital Center, Washington, DC

Lawrence A. Solberg, MD, PhD (Oncology) Mayo Clinic, Jacksonville, FL

Paul E. Wallner, MD (Radiation Oncology) Willingboro, NJ

J. Frank Wilson, MD (Radiation Oncology) Medical College of Wisconsin, Milwaukee, WI

To satisfy NQF's ICD-10 Conversion Requirements, we are providing the information below:

- NQF ICD-10-CM Requirement 1: Statement of intent related to ICD-10 CM
  Goal was to convert this measure to a new code set, fully consistent with the original intent of the measure.

- NQF ICD-10-CM Requirement 2: Coding Table
  See attachment in S.2b

- NQF ICD-10-CM Requirement 3: Description of the process used to identify ICD-10 codes
  The PCPI uses the General Equivalence Mappings (GEMs) as a first step in the identification of ICD-10 codes. We then review the ICD-10 codes to confirm their inclusion in the measure is consistent with the measure intent, making additions or deletions as needed. We have an RHIA-credentialed professional on our staff who reviews all ICD-10 coding. For measures included in CMS' Quality Payment Program (QPP), the ICD-10 codes have also been reviewed and vetted by the CMS contractor. Comments received from stakeholders related to ICD-10 coding are first reviewed internally. Depending on the nature of the comment received, we also engage clinical experts to advise us as to whether a change to the specifications is warranted.

Oncology: Medical and Radiation – Plan of Care for Pain (PQRS #144) was chosen as a suitable candidate for correlation analysis due to the similarities in patient population and domain. We hypothesize that there exists a positive association between patients with a diagnosis of cancer receiving chemotherapy or radiation therapy in which pain intensity is quantified (NQF # 0384) and those with a diagnosis of cancer receiving

chemotherapy or radiation therapy who report having pain with a documented plan of care to address pain (PQRS #144). Providers included in the analysis met the minimum number of quality reporting events (10) and were cleaned in the same process as the PQRS dataset.

Datasets were reviewed to identify shared providers based on NPI and TIN identifiers. Correlation analysis was then performed to evaluate the association between performance scores of these shared providers.

We use the following guidance to describe correlation[1]:

| Correlation | Interpretation |
|---|---|
| > 0.40 | Strong |
| 0.20 - 0.40 | Moderate |
| < 0.20 | Weak |

1. Shortell T. An Introduction to Data Analysis & Presentation. Sociology 712. http://www.shortell.org/book/chap18.html. Accessed July 13, 2018.

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

The results of the expert panel rating of the validity statement were as follows: N = 19; Mean rating = 4.32

Percentage in the top two categories (4 and 5): 84.21%

Frequency Distribution of Ratings
1- 0
2- 1
3- 2
4- 6
5- 10

Oncology: Pain Intensity Quantified – Medical Oncology and Radiation Oncology was positively correlated with Oncology: Medical and Radiation – Plan of Care for Pain (PQRS #144).

**PQRS #144**

Coefficient of correlation = 0.69

P-value = > 0.001

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

The measure is valid, as specified.

Oncology: Pain Intensity Quantified – Medical Oncology and Radiation Oncology has a strong positive correlation with another evidence-based process of care. The correlation demonstrates the criterion validity of the measure.

---

**2b2. EXCLUSIONS ANALYSIS**

**NA ☒ no exclusions — *skip to section 2b3***

**2b2.1. Describe the method of testing exclusions and what it tests** (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Not applicable

**2b2.2. What were the statistical results from testing exclusions**? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Not applicable

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

Not applicable

---

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

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

☒ **No risk adjustment or stratification**

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

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

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

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

Not applicable

**2b3.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities**.

Not applicable

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

Not applicable

**2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:**

- ☐ **Published literature**
- ☐ **Internal data analysis**
- ☐ **Other (please describe)**

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

Not applicable

**2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors** *(e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

Not applicable

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

Not applicable

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

**2b3.6. Statistical Risk Model Discrimination Statistics** (*e.g., c-statistic, R-squared*)**:**

Not applicable

**2b3.7. Statistical Risk Model Calibration Statistics** (*e.g., Hosmer-Lemeshow statistic*):

Not applicable

**2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves**:

Not applicable

**2b3.9. Results of Risk Stratification Analysis**:

Not applicable

**2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)?** (i.*e., what do the results mean and what are the norms for the test conducted*)

Not applicable

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

Not applicable

---

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

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

Data analysis performed on the measure included:

Average measure performance rate overall and by site, performance rate range by site and overall standard deviation for the measure.

Measures of central tendency, variability, and dispersion were calculated.

**2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities?** (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some*

*benchmark, different from expected; how was meaningful difference defined*)

Measure rate without exceptions: N= 862 Mean = 94.0% Standard Deviation= 0.2382

The performance rate by site is as follows, where n is the number of performance events by site:

| | | |
|---|---|---|
| A | 0.9780 | n=183 |
| B | 0.9740 | n=189 |
| C | 0.9730 | n=186 |
| D | 0.9730 | n=188 |
| E | 0.7160 | n=116 |

Based on the sample of 251 included physicians, the mean performance rate is 0.88 the median performance rate is 0.98 and the mode is 1.0. The standard deviation is 0.21. The range of the performance rate is 0.96, with a minimum rate of 0.04 and a maximum rate of 1.0. The interquartile range is 0.12 (1.0–0.88).

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

The performance rate range is .2620. Although this study captured performance on 862 events, the data were not captured at the physician level, restricting reporting of variation in performance to the organization level only. Additionally, we are unable to present a meaningful calculation of variation in performance across organizations due to the small sample size of sites (n=5) in this study.

The range of performance from 0.04 to 1 suggests there's clinically meaningful variation across physicians' performance.

---

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

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

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

**2b5.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications** (*describe the steps—do not just name a method; what statistical analysis was used*)

This test was not performed for this measure.

**2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications?** (*e.g., correlation, rank order*)

This test was not performed for this measure.

**2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications?** (i.*e., what do the results mean and what are the norms for the test conducted*)

This test was not performed for this measure.

---

**2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS**

**2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased** due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

The PQRS dataset provided to us by CMS did not contain missing data so this test was not performed. Nevertheless, missing data may have been rejected when submitted to CMS in which case those values would not be counted towards measure performance. There is no indication that this missing data was systematic, thus their omission would lead to unbiased performance results.

**2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data?** (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

This test was not performed for this measure. There was no missing data.

**2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased** due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

The PQRS dataset provided to us by CMS did not contain missing data so this test was not performed. Nevertheless, missing data may have been rejected when submitted to CMS in which case those values would not be counted towards measure performance. There is no indication that this missing data was systematic, thus their omission would lead to unbiased performance results.

## 3. Feasibility

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

**3a. Byproduct of Care Processes**

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

**3a.1. Data Elements Generated as Byproduct of Care Processes.**

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

If other:

**3b. Electronic Sources**

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

**3b.1. To what extent are the specified data elements available electronically in defined fields** (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*) Update this field for **maintenance of endorsement**.

ALL data elements are in defined fields in a combination of electronic sources

**3b.2. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.** For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

**3b.3. If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.**

**Attachment:**

**3c. Data Collection Strategy**

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

**3c.1. Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

**IF instrument-based, consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.**

We have not identified any areas of concern or made any modifications as a result of feasibility testing and operational use of the measure in relation to data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, and other feasibility issues unless otherwise noted.

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

The Measure, while copyrighted, can be reproduced and distributed, without modification, for noncommercial purposes, eg, use by health care providers in connection with their practices. Commercial use is defined as the sale, license, or distribution of the Measure for commercial gain, or incorporation of the Measure into a product or service that is sold, licensed or distributed for commercial gain.

Commercial uses of the Measure require a license agreement between the user and the PCPI Foundation® (PCPI®) or the American Medical Association (AMA). Neither the AMA, nor the former AMA-convened Physician Consortium for Performance Improvement® (AMA-PCPI), nor PCPI, nor their members shall be responsible for any use of the Measure.

AMA and PCPI encourage use of the Measure by other health care professionals, where appropriate.

THE MEASURE AND SPECIFICATIONS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND.

## 4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

**4a. Accountability and Transparency**

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

**4.1. Current and Planned Use**

*NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.*

| Specific Plan for Use | Current Use (for current use provide URL) |
|---|---|
| Public Reporting | Payment Program<br>Merit-based Incentive Payment System (MIPS)<br>https://qpp.cms.gov/mips/quality-measures |

**4a1.1 For each CURRENT use, checked above (update for maintenance of endorsement), provide:**

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

Merit-based Incentive Payment System (MIPS)-Sponsored by the Centers for Medicare and Medicaid Services (CMS)

Prior to 2016, this measure was used for Eligible Providers (EPs) in the Physician Quality Reporting System (PQRS). As of 2017, PQRS has been replaced by the MIPS program. MIPS is a national performance-based payment program that uses performance scores across several categories to determine payment rates for EPs. MIPS takes a comprehensive approach to payment by basing consideration of quality on a set of evidence-based measures that were primarily developed by clinicians, thus encouraging improvement in clinical practice and supporting advances in technology that allow for easy exchange of information.

**4a1.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

According to the CY 2019 Quality Payment Program final rule, Physician Compare has continued to pursue a phased approach to public reporting under MACRA. CMS intends to make all measures under MIPS quality

performance category available for public reporting on Physician Compare. These measures include those reported via all available submission methods for MIPS-eligible clinicians and groups. Because this measure has been in use for at least one year and meets the minimum sample size requirement for reliability, this measure meets criteria for public reporting but has not yet been included in Physician Compare.

**4a1.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

Because this measure has been in use for at least one year and meets the minimum sample size requirement for reliability, this measure meets criteria for public reporting. 2018 data will be available for public reporting on Physician Compare in late 2019.

**4a2.1.1. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.**

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

The PCPI measure development process is a rigorous, evidence-based process that has been refined and standardized over the past fifteen years, since the PCPI's inception. Throughout its tenure, several key principles have guided the development of performance measures by the PCPI, including the following which underscore the role those being measured have played in the development process and later through implementation feedback:

Collaborative Approach to Measure Development

PCPI measures have been developed through cross-specialty, multi-disciplinary technical expert panels. Representatives of all relevant disciplines of medicine and other health care professionals are invited to participate as equal contributors to the measure development process. In addition, the PCPI strives to include on its panels, individuals representing the perspectives of patients, consumers, private health plans, and employers. Liaisons from key measure development organizations, including The Joint Commission and NCQA at times, participate in the PCPI's measure development process to ensure harmonization of measures; measure methodologists, coding and informatics experts also are considered important members of the expert panel. This broad-based approach to measure development maximizes measure buy-in from stakeholders and minimizes bias toward any individual specialty or stakeholder group.

Conduct Public Comment Period

Input from multiple stakeholders is integral to the measure development process. In particular, feedback is critical from those clinicians who will implement these measures. To that end, all measures are released for a 30-day public and PCPI member comment period. All comments are reviewed by the technical expert panel to determine whether measure modifications are needed based on comments received.

Feedback Mechanism

The PCPI has a dedicated process set up to receive comments and questions from implementers. As comments and questions are received, they are shared with appropriate staff for follow up. If comments or questions require expert input, these are shared with the PCPI's technical expert panels to determine if measure modifications may be warranted. Additionally, for PCPI measures included in federal reporting programs, there is a system that has been set up to elicit timely feedback and responses from PCPI staff in consultation with technical expert panel members, as appropriate.

**4a2.1.2. Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.**

See description in 4a2.1.1 above.

**4a2.2.1. Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.**

**Describe how feedback was obtained.**

In addition to the feedback obtained from a cross-specialty, multi-disciplinary technical expert panel during the measure development and maintenance process, the PCPI obtains feedback via a public comment period and an email-based process set up to receive measure inquiries from implementers. The public comment period feedback is provided via an online survey tool.

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

We received feedback stating that Joint Commission standards already require accredited hospitals to establish policies and procedures that address comprehensive clinical assessment of pain.

We also received recommendations to consider potential denominator exclusions.

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

See summary in 4a2.2.2 above.

**4a2.3. Describe how the feedback described in 4a2.2.1 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.**

While pain assessment has been adopted by the Joint Commission as a requirement for hospital accreditation, at least one-third of radiation therapy services are provided at free-standing centers and the majority of chemotherapy administration is provided in non-hospital settings. The PCPI measure is assessed at the physician level rather than the hospital level. Existing evidence suggests that cancer pain is not being optimally assessed or managed. This gap in care is likely to be more pronounced in private practices where Joint Commission standards do not apply.

The Oncology expert panel specifically designed this measure without denominator exclusions since addressing pain is such a critical aspect of care for all cancer patients.

**Improvement**

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

**4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)**

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

While the PCPI creates measures with an ultimate goal of improving the quality of care, measurement is a mechanism to drive improvement but does not equate with improvement. Measurement can help identify opportunities for improvement with actual improvement requiring making changes to health care processes and structure. In order to promote improvement, quality measurement systems need to provide feedback to front-line clinical staff in as close to real time as possible and at the point of care whenever possible. (1)

1. Conway PH, Mostashari F, Clancy C. The future of quality measurement for improvement and accountability. JAMA. 2013 Jun 5;309(21):2215-6.

**4b2. Unintended Consequences**

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

**4b2.1. Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.**

We are not aware of any unintended consequences related to this measurement.

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

We are not aware of any unexpected benefits from implementation of this measure.

## 5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

**5. Relation to Other NQF-endorsed Measures**

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

Yes

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

0177 : Improvement in pain interfering with activity

0192 : Residents who experience moderate to severe pain during the 7-day assessment period (risk-adjusted)

0420 : Pain Assessment and Follow-Up

0523 : Pain Assessment Conducted

0676 : Percent of Residents Who Self-Report Moderate to Severe Pain (Short Stay)

0677 : Percent of Residents Who Self-Report Moderate to Severe Pain (Long Stay)

1628 : Patients with Advanced Cancer Screened for Pain at Outpatient Visits

1637 : Hospice and Palliative Care -- Pain Assessment

**5.1b. If related or competing measures are not NQF endorsed please indicate measure title and steward.**

**5a. Harmonization of Related Measures**

The measure specifications are harmonized with related measures;

**OR**

The differences in specifications are justified

**5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**

**Are the measure specifications harmonized to the extent possible?**

Yes

**5a.2. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**

There are several NQF-endorsed measures related to measure #384 Oncology: Medical and Radiation – Pain Intensity Quantified. Most related measures are assessed within different settings and at distinct levels of analysis. NQF measure #177 assesses the percentage of home health episodes with improvements in the frequency of a patient's pain. The measure is assessed at the facility level and within the home care setting. NQF measure #192 assesses the percentage of nursing home residents or patients within skilled nursing facilities who experience moderate to severe pain. In contrast to the PCPI measure, measure #192 is assessed at the facility level. NQF measure #523 is also assessed at the facility level and focuses on whether home health patients are assessed for pain. NQF measures #676 and 677 are facility-based measures and assess

whether patients report moderate or severe pain while in post-acute care as short-stay or long stay patients, respectively. Measure #1628 is limited to patients with Stage IV diagnosis and is identified as a measure to be assessed at the facility, health plan or integrated delivery system level of analysis. NQF measure #1637 is also a facility level measure and assesses whether hospice or palliative care patients are assessed for pain. NQF measure #420 is also related to the PCPI measure but is a claims-based measure. Measure #420 generally assesses pain whereas the PCPI measure assesses cancer treatment-related pain which represents a current gap in care.

**5b. Competing Measures**

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
**OR**
Multiple measures are justified.

**5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
**Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)**
Not applicable

## Appendix

**A.1 Supplemental materials may be provided in an appendix.** All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.

No appendix **Attachment:**

## Contact Information

**Co.1 Measure Steward (Intellectual Property Owner):** PCPI

**Co.2 Point of Contact:** Samantha, Tierney, samantha.tierney@thepcpi.org, 312-224-6071-

**Co.3 Measure Developer if different from Measure Steward:** PCPI

**Co.4 Point of Contact:** Elvia, Chavarria, elvia.chavarria@thepcpi.org, 312-224-6064-

## Additional Information

**Ad.1 Workgroup/Expert Panel involved in measure development**

**Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.**

PCPI measures are developed through cross-specialty, multi-disciplinary technical expert panels (TEPs). Representatives of all relevant disciplines of medicine and other health care professionals are invited to participate. In addition, the PCPI strives to include on its TEPs individuals representing the perspectives of patients, consumers, private health plans, and employers. Measure methodologists, and coding and informatics experts also are considered important members of the TEP. All TEP members participate as equal contributors to the measure development process. This broad-based approach to measure development

ensures buy-in on the measures from all stakeholders and minimizes bias toward any individual specialty or stakeholder group. The Oncology measures were developed in 2007 and are maintained and updated by the TEP which was last reconvened in December 2017 to review the measure and ensure its currency.

The Cancer TEP members include:

Paul Wallner, DO (Chair)

Kerin Adelson, MD

Peter Albertsen, MD

Nancy Baxter, MD, PhD

Joel Brill, MD

David Cella, PhD

Andrea Cheville, MD

Charles Cleeland, PhD

John Gore, MD, MS

James Hayman, MD, MBA

Jerry Hussong, MD, DDS, MS

Arif Kamal, MD, MBA, MHS

W. Robert Lee, MD, MEd, MS

David Penson, MD, MPH

Louis Potters, MD

Howard Sandler, MD, MS

Eric Wisotsky, MD

**Measure Developer/Steward Updates and Ongoing Maintenance**

**Ad.2 Year the measure was first released:** 2007

**Ad.3 Month and Year of most recent revision:** 10, 2018

**Ad.4 What is your frequency for review/update of this measure?** Supporting guidelines, specifications, and coding for this measure are reviewed annually

**Ad.5 When is the next scheduled review/update for this measure?** 12, 2019

**Ad.6 Copyright statement:** 

**Ad.7 Disclaimers:** The Measure is not a clinical guideline, does not establish a standard of medical care, and has not been tested for all potential applications. The Measure, while copyrighted, can be reproduced and distributed, without modification, for noncommercial purposes, eg, use by health care providers in connection with their practices. Commercial use is defined as the sale, license, or distribution of the Measure for commercial gain, or incorporation of the Measure into a product or service that is sold, licensed or distributed for commercial gain.

Commercial uses of the Measure require a license agreement between the user and the PCPI Foundation® (PCPI®) or the American Medical Association (AMA). Neither the AMA, nor the former AMA-convened Physician Consortium for Performance Improvement® (AMA-PCPI), nor PCPI, nor their members shall be responsible for any use of the Measure.

AMA and PCPI encourage use of the Measure by other health care professionals, where appropriate.

THE MEASURE AND SPECIFICATIONS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND.

Limited proprietary coding is contained in the Measure specifications for convenience. Users of the proprietary code sets should obtain all necessary licenses from the owners of these code sets. The AMA, the PCPI and its

members and former members of the AMA-PCPI disclaim all liability for use or accuracy of any Current Procedural Terminology (CPT[R]) or other coding contained in the specifications.



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