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

Purple text represents the responses from measure developers.

Red text denotes developer information that has changed since the last measure evaluation review.

Brief Measure Information

NQF #: 0209

Corresponding Measures:

De.2. Measure Title: Comfortable Dying: Pain Brought to a Comfortable Level Within 48 Hours of Initial Assessment

Co.1.1. Measure Steward: National Hospice and Palliative Care Organization

De.3. Brief Description of Measure: Percentage of patients who report being uncomfortable because of pain at the initial assessment who, at the follow up assessment, report pain was brought to a comfortable level within 48 hours.

1b.1. Developer Rationale: As a patient reported outcome (PRO) the measure captures and reflects patient goals for pain management. The use of a dichotomous rating, incorporating the patient's perception of his/her own degree of comfort, provides a means of assessing provider performance of initial pain management. Consequently, this measure provides a more comprehensive picture of pain management than a measure that relies on achieving a specific score on a pain intensity rating scale or change in pain intensity rating.

While it is recognized that pain scales have intra-individual validity and that mean values have importance for population studies, the utility of numerical pain scores for a concurrently evaluated outcome measure and for program/system accountability is problematic. Not all patients mean the same thing when they give a rating – one person's ´3´ may be another patient´s ´6.´ The value of a numerical rating scale lies in comparison within subjects(comparing ratings over time) – and the fact that change is accomplished, or not, is more relevant than the absolute number achieved. However, change in scores alone does not demonstrate whether comfort was achieved. In addition, using a set numeric rating as goal loses, or at least undermines, the concept of patient self-determination. If pain is an individual experience with an individual response, then the decision of what is acceptable/comfortable

should be left up to the individual, not determined arbitrarily. It's more consistent with patient-centered care to care to ask the patient to decide how comfortable he/she wants to be. Because of its focus on comfort, the measure also allows for a broader conceptualization of pain than use of a measure that relies solely on a numeric intensity rating. The measure also has the advantage of identifying those patients who require intervention and at the same time allows the clinician to use the most appropriate means of pain assessment for each individual patient.

S.4. Numerator Statement: Patients whose pain was brought to a comfortable level (as defined by patient) within 48 hours of initial assessment.

S.6. Denominator Statement: Patients who replied "yes" when asked if they were uncomfortable because of pain at the initial assessment.

S.8. Denominator Exclusions: Patients who do not report being uncomfortable because of pain at initial assessment (i.e., patients who reply "no" to the question "Are you uncomfortable because of pain?"

Patients under 18 years of age

Patients who cannot self-report pain

Patients who are unable to understand the language of the person asking the initial and follow up questions

De.1. Measure Type: Outcome: PRO-PM

S.17. Data Source: Instrument-Based Data

S.20. Level of Analysis: Facility, Other

IF Endorsement Maintenance – Original Endorsement Date: Aug 10, 2009 **Most Recent Endorsement Date:** Oct 26, 2016

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?

Preliminary Analysis: Maintenance of Endorsement

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

Criteria 1: Importance to Measure and Report

1a. Evidence

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

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

Evidence Summary of prior review in 2016

- The developer provided a rationale and diagram illustrating the pain assessment process and how it relates to the outcome of pain being brought to a comfortable level (the focus of this Patient-Reported Outcome-based Performance Measure (PRO-PM).
- The developer also addressed a new submission question since the previous evaluation regarding demonstration that the target population values the measured PRO and finds it meaningful.

- The developer stated "The negative effect of pain on quality of life and the need for timely and effective pain management is universally accepted. Consequently, minimal investigation has been done related to the importance of pain management at end of life. One study (McMillan et al., Oncology Nursing Forum, 2002)investigating symptom distress and quality of life in patients with cancer newly admitted to hospice home care did find a strong relationship between pain and distress."

- The 2016 Committee agreed that the developers rationale from self-reported pain, to clinical and psychosocial assessment, then to intervention is an effective way of reporting alleviation of pain. The Committee also agreed with the clinical action that could influence patient – reported pain levels and that hospice patients find questions regarding level of pain to be meaningful.

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:

Question for the Committee:

- *The developer attests the underlying evidence for the measure has not changed since the last NQF endorsement review. Does the Committee agree there is no need for repeat discussion and vote on Evidence?*

Guidance from the Evidence Algorithm

Pro-based measure (Box 1) -> Relationship between the outcome and at least one healthcare action is identified and supported by the rationale (Box 2) -> PASS

Preliminary rating for evidence: ☒ Pass ☐ No Pass

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.

- For NQF maintenance of endorsement, measure stewards/developers are expected to provide current performance data. If limited data are available, data from the literature can be considered.
  - According to the developer, NHPCO has not collected data on this measure since 2015. This measure is included in CMS' Merit-based Incentive Payment System (MIPS); however, data are not available on the utilization of the measure.
- Performance data for facility scores were provided for years 2012-2015 for those hospice facilities that voluntarily submitted data. The mean and standard deviation were 66.4 (SD=21.1) in 2012 across 143 reporting hospice facilities and 64.7 (SD=24.5) in 2015 across 46 reporting hospice facilities.
- A summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement was not provided.

Disparities

- Data presented by the developer suggest there are no disparities in care by age group, sex, race, or condition (cancer vs. non-cancer).

Questions for the Committee:

- *Is the sample adequate to provide meaningful information about opportunity for improvement? Do the reporting facilities for 2012-2015 reflect U.S. hospice programs in terms of size, region, etc.?*
- *Is there a gap in care that warrants a national performance measure?*
- *Are you aware of evidence of any subgroup disparities in pain being brought to a comfortable level?*

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

RATIONALE: For NQF maintenance of endorsement, measure stewards/developers are expected to provide current performance data. Data from the literature can be considered, but were not provided.

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.

- This is one of the few (if not only) PRO-PM used in palliative care and hospice. There is good evidence that timely pain management is a strong predictor of quality of life.
- Measure 0209 Comfortable Dying: Pain Brought to a Comfortable Level Within 48 Hours of Initial Assessment is a PRO-PM that offers direct pt input on the outcomes related to management of pain.
- pass
- Not aware of any new studies/information.
- I am not sure how to answer the question posed to the Committee regarding the need for to repeat discussion and vote. I am a bit conflicted but will go along with the majority.
- Question: For this measure was a definition of "Comfortable" provided that also includes the wording "Acceptable" Pain level?
- There have been no changes in evidence.
- Meets criteria, pain remains a common symptom for dying patients. New studies/information could have been included in the submission to reflect the value of patient-reported pain assessment for dying patients.
- Data on evidence based on prior review (2016) which was also based on a prior review. Developer Rationale well-stated and clear.
- Yes
- In the previous review, developers presented data supporting the importance of this measure to patients. This is consistent with clinical evidence that patients value pain control. This measure is directly related to the outcome of controlling pain in hospice. No new evidence was presented for this review.
- There have not been any changes to the evidence since the last evaluation. No changes to the evidence base.
- Not aware of any new studies
- The evidence supports the continuation of the measure.

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?

- Performance gap date comes from prior than 2015, and agree with staff that more updated data would be idea. But in my opinion, not crucial for continued endorsement.
- "According to the developer, NHPCO has not collected data on this measure since 2015. This measure is included in CMS' Merit-based Incentive Payment System (MIPS); however, data are not available on the utilization of the measure. •A summary of data from the literature that indicates opportunity for improvement. "
- insufficient because no new data, opiate crisis and rise of palliative care programs may influence
- Would like to see additional information/evidence that there is a gap.
- So little information was provided that it is difficult to answer these questions. I want to look into the subgroup disparities issue more closely.
- Not clear that this one performance measure necessarily notes a gap in care but in conjunction with other measures for hospice care
- Performance data was from 2012-2015 and voluntarily submitted. Literature was not provided.
- Insufficient-- without more current data and without comparison of performance data by subgroups, it is not possible to determine how this measure demonstrates a gap or disparities in care.
- Data from 2012 - 2015. None since 2015. Mean was 66.4 and 64.7 suggesting an important gap. Data presented by the developer suggest there are no disparities in care by age group, sex, race, or condition (cancer vs. non-cancer). I would like to know why there aren't more current data.
- Is not risk adjusted. Researchers feel no gap in care.
- No performance or disparity data has been reported.
- Data has not been collected since 2015, it is included in CMS's MIPS no data on utilization of the measure.
- NA, previous evidence was submitted
- Criteria met to continue.

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: NQF Staff Review

Reliability

- According to NQF guidance, all instrument-based measures should have both data element and score level testing conducted. However, this is challenging for single item instrument measures and the quarterly mean analysis at the facility level may be as close to this as the developer is likely to be able to get given that it is not appropriate to conduct test-retest for this population with this measure. An exception to the NQF testing guidance should be considered in the case of this measure.
- The score level analysis as described by the developer produces an appropriate reliability analysis. The primary reliability analysis should compare between facility variance as a fraction of between plus within facility variance for each facility, which is what the intraclass correlation coefficient conducted by the developer is describing. This offers an appropriate analysis of confidence that providers are ranked by performance appropriately and was separate from the between quarter intra-facility score level stability analysis.
- Reliability Testing Results:
    - Instrument Level :
        - Percent uncomfortable due to pain: F=0.42 p=0.5115
        - Percent whose pain was brought to a comfortable level: F=0.77 p=0.3814
    - Measure Level (Organization Level):
        - F-value = 4.04 ($p < 0.001$): IUR =0.75

Validity

- The analysis described by the developer (concurrent validity) is an appropriate data-element validity test. Developers compared response rates from two different wordings ("comfortable" level and "acceptable" level) for the follow-up question related to pain management. However, all instrument-based measures must also have score level validity testing. This is commonly done using convergent validity testing, calculating a Pearson's correlation coefficient between the measure of interest and an external measure of quality within a proximate quality domain. To demonstrate validity at the score level, the developer compared contemporary results from the Composite Score of the Family Evaluation of Hospice Care (NQF 0208) and the average Comfortable Dying scores. The hypothesis was, if Comfortable Dying scores are associated with high performing hospices, then they will be significantly associated with performance on another performance measure. The developer identified high performing hospices on NQF 0208 and performed a logistic regression with the hospice Comfortable Dying Score.
- Testing data included 212 of 686 patients from 9 hospice agencies who reported pain on initial assessment. Sixty percent of patients (n=127) responded their pain was brought to a comfortable level with 48 hours and 64% (n=136) responded their pain was brought to an acceptable level within 48 hours. The developers conclude that 96% of patients provided the same answer to the two wordings of the pain management question. The developers also to report a Cohen's kappa of 0.91.
- Validity testing results:
    - Logistic regression found a direct and marginally significant association between being rated a high performing hospice based on FEHC Composite Score (NQF 0208) and having a higher Comfortable Dying score ($\chi 2 = 3.352$ p=0.067). According to the developers, the results indicate that high performing hospices on the FEHC Composite Score (NQF 0208) were more likely to have a higher score on the Comfortable Dying measure. The statistical was marginal for this test but it should be noted that only 90 hospices were able to be included in the logistic regression because not all hospices participated in both measures during 2013. It is

expected, however, that with a larger sample size, statistical significance would be achieved. Developers note that the results support validity of the measure for understanding how well hospices were able to relieve patients admitted in pain.

- NHPCO conducted an analysis of potential risk factors at the facility level. Scores used in this analysis combined data from 2012 and 2013. Because the distribution of scores showed departures from normality, nonparametric statistical tests were used (Kruskal-Wallis, Wilcoxon). Factors examined included: geographic location; service area; ownership; race and ethnicity; patient age; patient gender; patient principle diagnosis; and referral source. According to the developer, none of these factors demonstrated an association with the measure scores and no trend was seen between facilities with different proportions of patients in any of the risk factor categories. **Note: specific results of these analyses were not included as part of the Fall 2020 measure submission package.**

- The developers maintained their position that risk adjustment is not needed because providers have an equal responsibility to provide timely pain management for all patients.

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- NQF staff is satisfied with the reliability testing for the measure. Does the Committee think there is a need to discuss and/or vote on reliability?

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?
- Is there any evidence that contradicts the developer's rationale and analysis underlying the decision not to risk-adjust this measure?
- NQF staff is satisfied with the validity analyses for the measure. Does the Committee think there is a need to discuss and/or vote on validity?

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

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?

- Though this is an outcome measure, agree that case-mix adjustment is not needed as the patient serves as their own control.
- It was noted that it is not appropriate to conduct test-retest for this population with this measure. An exception to the NQF testing guidance should be considered in the case of this measure.
- no concerns
- No concerns
- I have no concerns regarding the measure being consistently implemented so feel no need to discuss or vote.
- Concern over the definition provided regarding "comfortable"
- No concerns

- Specifications are sufficiently defined.
- Reliability data from between and within facilities using intraclass correlation.
- No concerns
- There is really only one data element in this measure and it is clearly defined. The numerator and denominator definitions are clear. No concerns.
- It appears that this measure can be consistently implemented, but it hasn't been reported since 2015.d
- NA, is already being implemented
- no concerns

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

- no
- None. The patient's self-report of pain is important to all aspects of quality care and is strongly supported in the literature.
- no concerns
- No
- No.
- I was not clear if tested for diverse populations of patients
- No
- No concerns, moderate rating.
- No
- No
- No Concerns. No need to vote
- no concerns discussion is not necessary
- No
- no concerns

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

- no
- None.
- no concerns
- No
- No.
- I was not clear if tested on diverse populations of patients
- No
- Moderate rating, validity testing could have been strengthened with more recent/broader organizational level testing.
- Developers compared response rates from two different wordings ("comfortable" level and "acceptable" level) for the follow-up question related to pain management. No concerns
- No
- No concerns. No need to vote
- no
- No
- No

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 concerns
- Risk adjustment data not provided as per developer not necessary; addressing pain should be for all.
- no concerns
- No concerns
- I do not see that any patient groups (except those under 18) are excluded. I also do not see any need for risk adjustments.
- No risk adjustment was included
- There was no risk adjustment.
- No adjustment for social risk factors.
- I don't believe there was any risk adjustment. Excludes people < 18 claiming most hospices have few patients under 18. However this does not help patient < 18 who do need palliative care and/or hospice care.
- No risk adjustment used: rationale acceptable
- No risk adjustment issues reported
- Exclusions seem to be consistent, the analyses indicate acceptable results although the results have not been reported since 2015.
- NA
- yes appropriate

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?

- no concerns
- No concerns.
- no concerns
- No concerns
- I do not see any threats to validity.
- Yes
- No concerns
- No threat to validity
- no
- No
- Validity calculations seem to be done according to accepted process
- There does not appear to be any missing data, statistically significant differences, or multiple data sources.

- NA
- no concerns

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.
    - Required data elements (i.e. patient response to initial query and follow-up response) are not necessarily kept electronically – some providers may need to develop and maintain a paper record system to track responses.
    - NHPCO maintains ongoing support (in the form of written materials and one-on-one guidance)for hospice providers who use the measure . There are no costs associated with the use of this measure.
    - Many hospices reported difficulty implementing the measure when it was required in the first year of reporting as part of the Hospice Quality Reporting Program. However, NHPCO notes that at that time, many hospices were unfamiliar with quality measure reporting and states that "Had 0209 been implemented later in the HQRP program and/or given more time along with education and support, hospices would likely have had more success with implementation."

Questions for the Committee:

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

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. Feasibility: 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?
    - yes, very much should be routinely collected and thus feasible
    - Challenges noted by developer that quality measurement and use of 0209 specifically was an issue when first implemented because of hospice gaps in understanding quality measurement. NHPCO provides resources on use.
    - measure has been out so long now, no issue with feasibility
    - No concerns
    - Appears that not all data are recorded or stored electronically. So it is still feasible to collect the data but could impact use of measure and results.
    - The data element ("comfort/improvement in pain") should be used during care delivery. Not sure available 100% in electronic form. To be placed into operational use would need to insure consistent application of definition of "comfortable"
    - Data is not always kept electronically--some maintain a paper record system.

- Moderate rating, data is collected as a standard part of care delivery but ease of data reporting could be enhanced with use of electronic data sources.
- Hospices reported difficulty using the measure. Requires two data points (evaluation and 48 hours after evaluation). there was a quote about implementation being more feasible with better education.
- No concerns
- The data element is apparently not routinely generated or routinely document. Not clear why this is the case given the centrality of this measure to a central goal of hospice- to control symptoms. It's not clear why the developer, who is also the professional organization of most of the users, has not done more to have this measure implemented.
- My concerns are that the data has not been reported since 2015.
- NA
- no concerns

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** ☒ **UNCLEAR**

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

OR

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

Accountability program details

- The measure is currently used for internal quality improvement.
- This measure is aligned with MIPS 342: Pain brought under control within 48 hours: https://qpp.cms.gov/docs/QPP_quality_measure_specifications/CQM-Measures/2019_Measure_342_MIPSCQM.pdf
- The level of utilization is unknown. NHPCO provides a manual for data collection but there is no comparative reporting since 2015.

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 N/A

Additional Feedback:

- The measure developer did not give additional feedback at this time.

Questions for the Committee:

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

Preliminary rating for Use: ☒ Pass ☐ No Pass

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

Improvement results

- The developer reported that the results reflect national level numbers, therefore improvements by individual providers are not reflected. Plus, the patients that self-report before receiving the pain reduction treatment might not be able to respond within the 46-72 hour time window.

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 [unexpected findings]

- The developer did not identify any unintended consequences or unexpected benefits in the measure.

Potential harms

- None

Additional Feedback:

- None

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: Data were not provided to indicate performance over time.

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?

- aligns with MIPS measure

- Extent of use not known. The measure is currently used for internal quality improvement and aligns with MIPS 342.
- pass
- Would like to see this publicly reported again.
- I think that level of utilization being known could help determine if performance results are being used to further the measure's goal.
- Unclear about publicly reported. Did not see additional feedback being considered for this measure.
- The measure developer did not give feedback. It appears that feedback is not given to those being measured.
- There is insufficient information provided about the use of this measure for accountability applications and about feedback by users on use of this measure. No support provided for alignment of patient-reported "comfort" with "pain brought under control".
- Level of use unknown. But it does say it is used for QI.
- For internal organization use; would like discussion on public reporting for this measure.
- Developer did not supply any new data on feedback
- It is unclear if the measure is being publicly reported it is in current use in an accountability program. Feedback is considered when changes are incorporated into the measure. The measure is used for internal quality improvement.
- NA
- Feedback considered

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.

- no concerns
- No concerns noted.
- unable to assess because no new data provided
- No concerns
- I cannot think of any unintended consequences.
- There would need to be follow-up to this specific measure for patient care/quality of life if a patient is not meeting the goal of "comfortable/acceptable" pain level. Plan of care to address this would be important. Does not appear to be harm caused from this measure
- The results reflect national level numbers--improvements by individual providers are not reflected. Unintended consequences are not identified.
- Insufficient information was provided about how this measure has contributed to improving quality of pain or whether it had any unintended consequences.
- There were concerns about usability. Indicate they are engaging members in a new quality program in 2021 so may revisit data collection with this measure. No harms noted. Missed people who don't have pain at evaluation but do 48 hours later.
- Improvement - would like to see this publicly reported; no harms
- It would seem that this measure could be and should be used to promote the goals of hospice on the patient, facility and national policy levels. However, the developer seems to have retired at least some of the reporting and use of this measure which raises the question for this reviewer of why the developer wants to continue supporting this measure at all.

- The issue of pain management is important but this measure has not been reported since 2015. Improvement by individual providers are not reflected. Plus, the patients that self-report before receiving the pain reduction treatment might not be able to respond within the 46-72 hour time window.
- NA, see no harms with this measure
- These are publicly reported and helps to educate about services

Criterion 5: Related and Competing Measures

Related or competing measures

- 0177: Percentage of home health episodes of care during which the frequency of the patient's pain when moving around improved.*[facility-level outcome measure in home health setting]*
- 0420: Percentage of patients aged 18 years and older with documentation of a pain assessment through discussion with the patient including the use of a standardized tool(s) on each visit AND documentation of a follow-up plan when pain is present *[clinician-level process measure in ambulatory setting]*
- 1637: Percentage of hospice or palliative care patients who screened positive for pain and who received a clinical assessment of pain within 24 hours of screening *[clinician-level & facility-level process measure in hospice and hospital setting]*

Harmonization

Due to differences in care setting, patient population, and measure type, there likely will not be harmonization issues; however, these should be included in the discussion of NQF's Palliative Care portfolio.

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?

- no competing outcome measures
- "Three related or competing measures noted in addition to MIPS 342: •0177: Percentage of home health episodes of care during which the frequency of the patient's pain when moving around improved.[facility-level outcome measure in home health setting] •0420: Percentage of patients aged 18 years and older with documentation of a pain assessment through discussion with the patient including the use of a standardized tool(s) on each visit AND documentation of a follow-up plan when pain is present [clinician-level process measure in ambulatory setting] •1637: Percentage of hospice or palliative care patients who screened positive for pain and who received a clinical assessment of pain within 24 hours of screening [clinician-level & facility-level process measure in hospice and hospital setting] "
- measures listed are quite different than current being evaluated
- No concerns
- None that I am aware of.
- Agree that there are other measures addressing pain but not competing
- 0177,0420, 1637; harmonization is unlikely due to differences in care settings
- As mentioned, this measure is aligned with MIPS 342 and a discussion of harmonization would be beneficial.
- 3 competing measures.
- No

- No harmonization needed at this time.
- There are 3 related measures that address differences in care setting, patient population, and measure type.
- no
- none

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 01/26/2021

- No NQF Members have submitted support/non-support choices as of this date.
- No Public or NQF Member comments submitted as of this date.

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 0209

Measure Title: Comfortable Dying

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_0209" 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.**

   No concerns noted by staff.

RELIABILITY: TESTING

Submission document: "MIF_0209" 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 VALIDITY testing** of patient-level data conducted?
   ☐ **Yes** ☐ **No**
6. **Assess the method(s) used for reliability testing**

   Submission document: Testing attachment, section 2a2.2

   The score level analysis as described by the developer produces an appropriate reliability analysis. The primary reliability analysis should compare between facility variance as a fraction of between plus within facility variance for each facility, which is what the intraclass correlation coefficient conducted by the developer is describing. This offers an appropriate analysis of confidence that providers are ranked by performance appropriately and was separate from the between quarter intra-facility score level stability analysis.
7. **Assess the results of reliability testing**

   Reliability Testing Results:
   - Instrument Level :
     - Percent uncomfortable due to pain: F=0.42 p=0.5115
     - Percent whose pain was brought to a comfortable level: F=0.77 p=0.3814
   - Measure Level (Organization Level):
     - F-value = 4.04 (p <0.001): IUR =0.75

   Submission document: Testing attachment, section 2a2.3
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 <u>all</u> 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.**

Precise specifications (Box 1) → Empiric reliability testing (Box 2) → Score-level testing (Box 4) Appropriate method (Box 5) → Moderate certainty that measure results are reliable (Box 6b)

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

    No concerns noted by staff.

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

    Submission document: Testing attachment, section 2b4.

    No concerns noted by staff.

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.

15. There is only one set of specifications for this measure. Comparability of data sources/methods is not applicable. **Please describe any concerns you have regarding missing data.**

    Submission document: Testing attachment, section 2b6.

    No concerns noted by staff.

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:**

    This measure is not risk adjusted.

    16d.1 All of the risk-adjustment variables present at the start of care? ☐ Yes ☐ No

    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 analysis described by the developer (concurrent validity) is an appropriate data-element validity test. Developers compared response rates from two different wordings ("comfortable" level and "acceptable" level) for the follow-up question related to pain management. However, all instrument-based measures must also have score level validity testing. This is commonly done using convergent validity testing, calculating a Pearson's correlation coefficient between the measure of interest and an external measure of quality within a proximate quality domain. To demonstrate validity at the score level, the developer compared contemporary results from the Composite Score of the Family Evaluation of Hospice Care (NQF 0208) and the average Comfortable Dying scores. The hypothesis was, if Comfortable Dying scores are associated with high performing hospices, then they will be significantly associated with performance on another performance measure. The developer identified high performing hospices on NQF 0208 and performed a logistic regression with the hospice Comfortable Dying Score.

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

Submission document: Testing attachment, section 2b2.3

Testing data included 212 of 686 patients from 9 hospice agencies who reported pain on initial assessment. Sixty percent of patients (n=127) responded their pain was brought to a comfortable level with 48 hours and 64% (n=136) responded their pain was brought to an acceptable level within 48 hours. The developers conclude that 96% of patients provided the same answer to the two wordings of the pain management question. The developers also to report a Cohen's kappa of 0.91.

Logistic regression found a direct and marginally significant association between being rated a high performing hospice based on FEHC Composite Score (NQF 0208) and having a higher Comfortable Dying score ($\chi 2 = 3.352$ $p=0.067$). According to the developers, the results indicate that high performing hospices on the FEHC Composite Score (NQF 0208) were more likely to have a higher score on the Comfortable Dying measure. The statistical was marginal for this test but it should be noted that only 90 hospices were able to be included in the logistic regression because not all hospices participated in both measures during 2013. It is expected, however, that with a larger sample size, statistical significance would be achieved. Developers note that the results support validity of the measure for understanding how well hospices were able to relieve patients admitted in pain.

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. **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 are threats to validity and/or relevant threats to validity were not assessed OR 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 is required; if not conducted, should rate as INSUFFICIENT.)

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

    Potential threats assessed (box 1) → Empirical validity assessed (box 2) → Computed performance measure scores (box 5) → appropriate methods (box 6) → moderate certainty (box 7b) → Moderate rating

ADDITIONAL RECOMMENDATIONS

24. **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

0209_Evidence_2016_2_29-635936604787753124-637372241260216405.docx,NQF_evidence_attachment_Sep2017-637372241260216405-637417321973185371.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)

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

Outcome

☒ Outcome: Pain brought to a comfortable level within 48 hours of initial assessment

☒ Patient-reported outcome (PRO): Symptom: pain

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*):

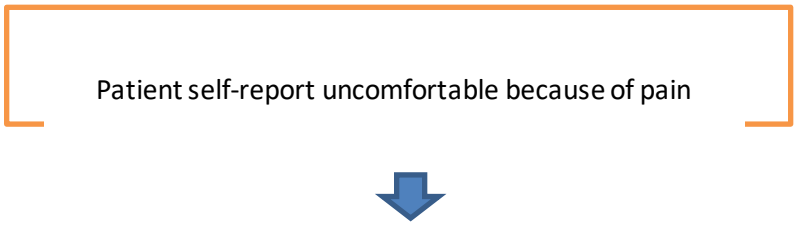
☐ Process:

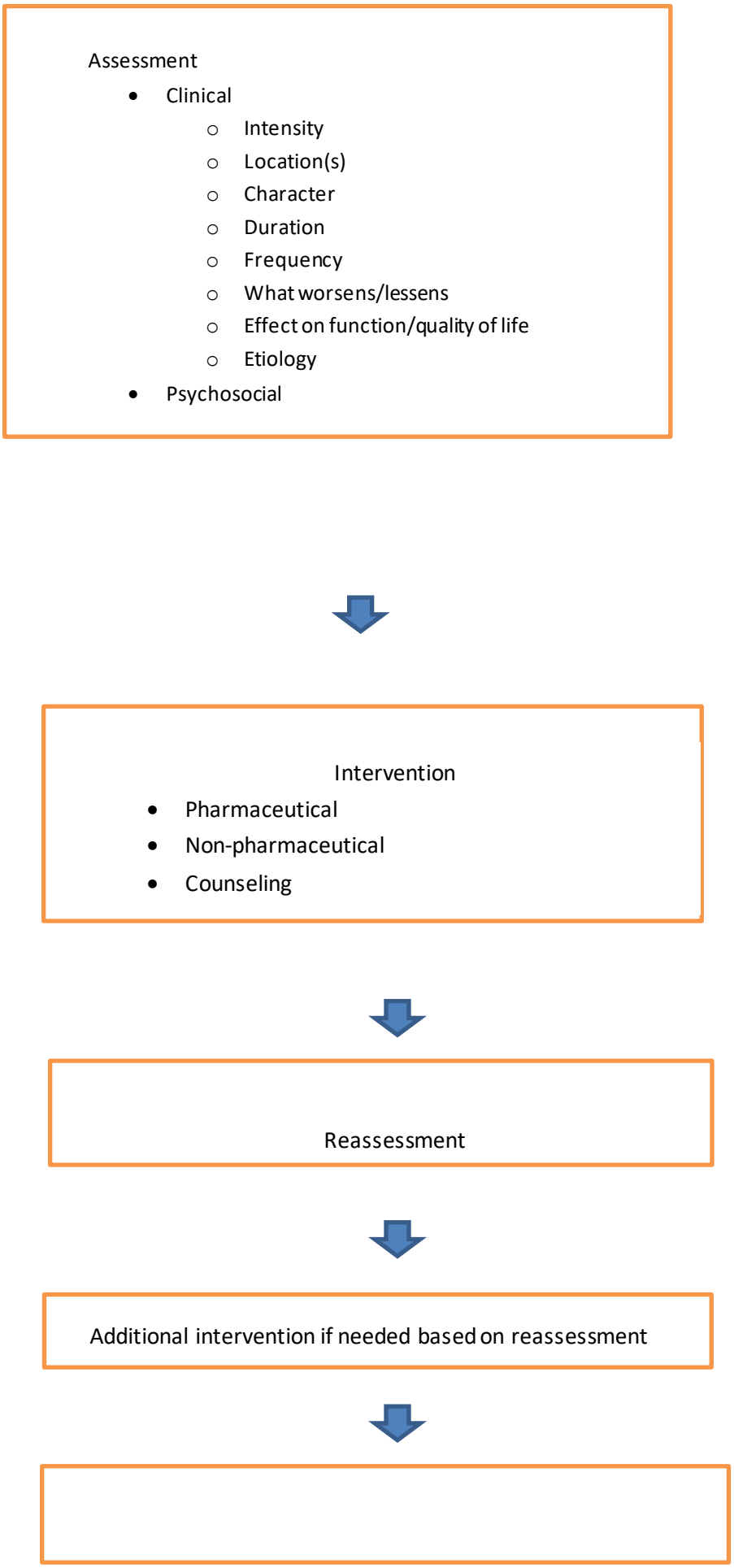
☐ Appropriate use measure:

☐ Structure:

☐ Composite:

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.




Assessment
Clinical
Intensity
Location(s)
Character
Duration
Frequency
What worsens/lessens
Effect on function/quality of life
Etiology
Psychosocial
Intervention
Pharmaceutical
Non-pharmaceutical
Counseling
Reassessment
Additional intervention if needed based on reassessment

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.)

Process: Pain Assessment. Inadequate or poorly performed pain assessment will result in unrelieved pain. A comprehensive assessment is essential to developing a pain intervention that will be effective and fully meet the needs of the patient. No objective means to assess pain exist – pain is subjective. Assessment must start with the patient's self-report of pain and proceed through careful questioning about all of the various characteristics of the patient's pain. Patients' beliefs about pain and pain management plus cognitive factors such as the ability to follow instructions affect adherence to pain interventions and assessment of these factors is key to effective pain management as well.

****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.

Data used in testing were largely drawn from the Patient Outcomes surveys that NHPCO completed. A sample of quarterly data submissions was taken covering two years (2009 and 2010) worth of submissions. The sample consisted of only those agencies that submitted multiple (=2) quarters worth of data during that period. There were 79 hospices agencies that submitted usable data for the Comfortable Dying measure covering 285 quarters (in total) worth of data and nearly 50,000 patients. Of those 79 hospice agencies, 58 (73.4%) provided multiple quarters worth of data during that period, covering data on over 38,000 patients. The two-year quarterly average percent of patients reporting being uncomfortable due to pain on admission was 20.8% (95% CI 19.5% - 22.1%). The two-year quarterly average percent of patients reporting having their pain brought to a comfortable level within 48 hours of admission was 69.3% (95% CI 66.3% - 72.3%).

Data were provided by 484 hospices, on a voluntary basis. Testing used data from 16,778 patients who qualified for the denominator of the measure. In most aspects of facility-level testing, 97 hospices were chosen whose denominators exceeded 50. A study of possible stratification by age or gender used patient-level data from two hospices. A patient-level validation test was done on a special sample of 212 patients.

Intraclass correlation was used for reliability testing for the measure. To provide evidence of measure reliability we must show that, all things being equal, hospices will reliably submit the same data over multiple quarters. Put another way, given that the proportion of patients whose pain is brought to a comfortable level within 48 hours of admission does not significantly change between quarters, the reported proportion will also remain the same.

To test this hypothesis, agency-level results were calculated from the sample hospice for the percent of hospice patients reporting being uncomfortable due to pain on admission, and the percent of patients who report having their pain brought to a comfortable level within 48 hours after admission. Univariate analysis was performed to provide the overall distribution of results for both variables results. To examine the similarity of data submitted in each quarter, an analysis of variance was performed to determine if significant differences existed in between the quarterly means for both agency level results. Next, an analysis of variance was performed to examine the differences in mean scores between and among hospices over the two years. Finally, intra-class correlations coefficients (ICC) were calculated to examine the measurements reliability over the sample years. Statistical significance was set at $P < 0.05$. All analysis completed utilizing SAS version 9.2.

NOTE: Test-retest is a frequently used method for reliability testing with single item measures and has been used with pain measures. However, the Comfortable Dying Measure assesses a characteristic that can inherently be expected to change rapidly (interventions to achieve better pain control can be and often are instituted at the time of assessment) making test-retest an inappropriate choice for reliability testing for this measure.

Analysis considered the consequences of random differences between patients by modeling changes within one facility as a binomial distribution. Tests for changes used Fisher exact or other exact statistical tests for change in proportions.

The analysis of variance of quarterly mean percentages of patients who reported being uncomfortable due to pain on admission showed no significant difference of mean scores between quarters (F-value = 1.11; P = 0.355). Variance of this measure demonstrated the expected significant difference between submitting hospices agencies (F-value = 7.48; P<0.0001). The intra-class correlation coefficient for the difference of the between and within hospice variation was 0.76 (95% CI 0.70 – 0.81).

The analysis of variance of quarterly mean percentages of patients who reported having their pain brought to a comfortable level within 48 hours of admission also showed no significant difference of mean scores between quarter (F-value = 1.7; P=0.991). The Hospice level variance analysis of this measure showed significant differences between hospice agencies (F-value = 5.87; P<0.001). The intra-class correlation coefficient for the between and within hospice 0.71 (95% CI 0.63 – 0.77).

The analysis of the data showed that indeed, over two-years of quarterly data submissions, the percent of patients reporting being uncomfortable due to pain remained relatively constant. Since the assumption of similarity between quarters was met it was then safe to examine the relative between and among variation in results for the same measure. As expected, there were significant differences in the percent of hospice patients uncomfortable due to pain on admission reported by each hospice. However, the ICC demonstrated good (over 75%) consistency of results within hospices from quarter to quarter.

Similarly, these results show that the percent of patients whose pain was brought to a comfortable level within 48 hours of admission, remained non-significantly differently. In fact, the results show that there was nearly no difference from quarter to quarter the results for this measure. The ICC for this measure also demonstrated good consistency (approximately 71%) of results within hospices from quarter to quarter. This slightly smaller ICC for measure (when compared to the percent uncomfortable due to pain on admission ICC) is not necessarily an indication of reduced reliability. Increased within hospice variation would be expected as hospices make process changes to increase their score for this measure. Indeed this expectation is observed in the variance of hospice scores for percent of patients uncomfortable due to pain on admission compared to those whose pain was brought to a comfortable level within 48 hours (F-values = 7.48 and 5.87 respectively).

It is likely that both ICC scores are conservative estimates of the true reliability of the measure. Even though there was little quarterly change in the percent of patients uncomfortable due to pain on admission (and likewise having their pain brought to a comfortable level), common sense dictates that real differences actually occurred at the hospice level. Since we know that the assumption of consistency of the base data can't be exactly true, we know that the true ICC's for these measures must be higher than what was observed.

In conclusion, this analysis provides statistical evidence that the NHPCO Comfortable Dying measure has good reliability.

This measure is concerned with newly admitted patients. It cannot be repeated on the same population of patients because each patient had only one initial period of 48 hours after hospice admission. There are also some very real constraints about how many times a patient can be asked whether comfort was attained in the first 48 hours. If the patient reports that comfort was not attained, the clinician may react by immediately increasing the dose of analgesics. The patient's subjective recollection of pain at 48 hours could change as a result.

Reliability of this measure may, however, be considered by two other avenues. Basic probability theory considerations give a guideline for how precise and repeatable the measure can be during random variations in the characteristics of individual patients admitted. Additionally, experience with real data gives an impression about whether the measure tends to stay constant between successive time periods.

Variations between individual patients will cause the numerator of this measure to fluctuate in accordance with the usual binomial distribution, even if a hospice keeps a completely constant pain management strategy and continues to admit patients with the same average characteristics. The measure itself, numerator/denominator, will show less random variability as the denominator increases.

We consider here the behavior of this measure when the denominator is at least 50.

In a 2013 nationwide survey involving over 16,000 patients who qualified for the denominator, we found 58% of those patients qualified for the numerator. A guideline for the random variability of the measure is provided by supposing that a hospice had admitted 50 patients, chosen at random from all the patients in the nationwide denominator, and that the measure was computed with the outcomes that occurred in the national sample. Such a hospice would have an average comfortable dying measure of 58%. It would have an 80% chance of being assigned a measure between 48% and 68%, and a less than 1% chance of being assigned a measure less than 38% or greater than 78%.

A survey covering both 2013 and 2014 gives some experience about the variability of the measure between successive time periods. Data was considered from 22 hospices whose denominators exceeded 50 in both of those years. Only 32% of those hospices had a measure that changed by more than 15 percentage points from 2013 to 2014, and only three of those hospices had a measure that changed by more than 20 percentage points. Changes were somewhat larger than those expected by pure random variation, but still consistent with good reliability for the measure. We expect more change than random variation in some subset of the hospices. Some hospices may be changing their strategy for pain management or may be taking referrals from different sources.

Results are consistent with good reliability for the test when the denominator exceeds 50.

Initial testing of measure performed with 686 patients in 9 hospices. Of those, 212 (31%) indicated that they were uncomfortable because of pain at the initial assessment on admission to hospice services. Criterion (concurrent) validity was tested by using two different wordings for the follow up question related to whether pain was managed. Patients first were asked if their pain was brought to a comfortable level within 48 hours and then they were asked if their pain was brought to an acceptable level within 48 hours. These two forms of the follow-up question were judged by the expert panel for the Comfortable Dying Measure to be equivalent in that they equally reflected patient preference and level of effectiveness achieved for pain management.

A special test dataset was created by using data from nine hospices. The dataset included 212 patients who qualified for the denominator of the measure. 127 of those patients stated that pain was brought to a comfortable level within 48 hours, and 136 of the 212 stated that pain was brought to an acceptable level within 48 hours. Because 96% of patients gave the same answer to the two wordings of the follow up question, the results indicate good concurrent criterion validity for the measure.

Sixty percent (N = 127) of the patients who initially responded that they were uncomfortable because of pain responded that their pain was brought to a comfortable level within 48 hours. Of those same patients, 64% (N = 136) responded that their pain was brought to an acceptable level within 48 hours. The two questions elicited very little difference in the proportion of patients replying that their pain was brought under control, indicating acceptable concurrent criterion validity of the measure.

Cohen's kappa = 0.91.

The results support validity of the measure for understanding how well hospices were able to relieve patients admitted in pain.

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

Systematic Review	Evidence
Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	*
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.	*
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	*
Provide all other grades and definitions from the evidence grading system	*
Grade assigned to the **recommendation** with definition of the grade	*
Provide all other grades and definitions from the recommendation grading system	*
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	*
Estimates of benefit and consistency across studies	*
What harms were identified?	*

Systematic Review	Evidence
Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	*
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	*

*cell intentionally left blank

_____

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.

As a patient reported outcome (PRO) the measure captures and reflects patient goals for pain management. The use of a dichotomous rating, incorporating the patient's perception of his/her own degree of comfort, provides a means of assessing provider performance of initial pain management. Consequently, this measure provides a more comprehensive picture of pain management than a measure that relies on achieving a specific score on a pain intensity rating scale or change in pain intensity rating.

While it is recognized that pain scales have intra-individual validity and that mean values have importance for population studies, the utility of numerical pain scores for a concurrently evaluated outcome measure and for program/system accountability is problematic. Not all patients mean the same thing when they give a rating – one person's ´3´ may be another patient´s ´6.´ The value of a numerical rating scale lies in comparison within subjects (comparing ratings over time) – and the fact that change is accomplished, or not, is more relevant than the absolute number achieved. However, change in scores alone does not demonstrate whether comfort was achieved. In addition, using a set numeric rating as goal loses, or at least undermines, the concept of patient self-determination. If pain is an individual experience with an individual response, then the decision of what is acceptable/comfortable

should be left up to the individual, not determined arbitrarily. It's more consistent with patient-centered care to care to ask the patient to decide how comfortable he/she wants to be. Because of its focus on comfort, the measure also allows for a broader conceptualization of pain than use of a measure that relies solely on a numeric intensity rating. The measure also has the advantage of identifying those patients who require intervention and at the same time allows the clinician to use the most appropriate means of pain assessment for each individual patient.

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.*

Years

	2012	2013	2014	2015
Mean	66.4	61.4	61.4	64.7
Std. Dev.	21.1	20.2	20.4	24.5
n (facilities)	143	292	74	46
no. of patients	9077	16522	3750	2072

Quartiles of the facility scores

	2012	2013	2014	2015
min	0	0	20	0
1st	57	50	46	50
median	66	60	60	65
3rd	80	74	75	81
max	100	100	100	100

Deciles of the facility scores

	2012	2013	2014	2015
min	0%	0%	20%	0%
10 %ile	40%	37%	33%	31%
20 %ile	51%	46%	43%	48%
30 %ile	60%	53%	50%	51%
40 %ile	63%	58%	52%	57%
50 %ile	66%	62%	60%	65%
60 %ile	70%	65%	64%	69%
70 %ile	75%	70%	74%	74%
80 %ile	84%	78%	79%	87%
90 %ile	97%	88%	86%	100%
max	100%	100%	100%	100%

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.

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.

The initial testing included a total of 1409 patients, 463 (32.86 %) of whom responded that they were uncomfortable because of pain. On follow up,(13%) indicated their pain was not brought to a comfortable level; 87 (18.8%) were unable to self-report; and 44 (9.5%) had missing data. Data were collected over a 6 month period from all patients on initial assessment enrolled in the hospices participating in the testing of the measure.

Of those patients in the sample who had a primary diagnosis of cancer, 81% had pain brought to a comfortable level and 19% did not. Of those patients in the sample who had a non-cancer primary diagnosis, 84.8% had pain brought to a comfortable level and 15.2% did not. There was no statistically significant difference (p 0.52) in the ethnic distribution of patients whose pain was not brought to a comfortable level compared to those who achieved comfort.

Subsequent, more recent (2014) testing used a sample of 2329 patients to examine possible disparities by age, gender, and race. 383 of those patients qualified for the denominator of the measure. The measure did not seem to show a tendency with age. Patients younger than 75 had a similar score to those aged 75 and older (difference not statistically significant, p =0.54). Patients younger than 65 also had a similar score to that of the rest (41% vs 46%, p=0.68). The two genders had almost identical scores on the measure (45% vs 44%, p=0.92). There was not a statistically significant difference between the comfortable dying measures in the Caucasian and other-than-Caucasian portions of the sample (p=0.29). Thus there was no evidence in the sample for disparity by age, gender, or race.

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

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, Cardiovascular, Gastrointestinal (GI), Infectious Diseases (ID), Musculoskeletal, Neurology, Palliative Care and End-of-Life Care, Renal, Respiratory : Chronic Obstructive Pulmonary Disease (COPD)

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

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

Populations at Risk

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

https://www.nhpco.org/pom/

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

This is not an eMeasure **Attachment:**

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

No data dictionary **Attachment:**

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

Attachment**:** ComfortableDyingWorkbook-637372241257560122.xls

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.

Patient

S.3.1. For maintenance of endorsement: 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.

No

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

No changes to specifications. Explanatory phrase in parentheses removed: (after admission to hospice).

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.*

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

Patients whose pain was brought to a comfortable level (as defined by patient) within 48 hours of initial assessment.

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)*

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

Number of patients who replied "yes" when asked if their pain was brought to a comfortable level within 48 hours of initial assessment.

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

Patients who replied "yes" when asked if they were uncomfortable because of pain at the initial assessment.

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).

Patients who are able to self-report pain information and replied "yes" when asked if they were uncomfortable because of pain at the initial assessment.

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

Patients who do not report being uncomfortable because of pain at initial assessment (i.e., patients who reply "no" to the question "Are you uncomfortable because of pain?"

Patients under 18 years of age

Patients who cannot self-report pain

Patients who are unable to understand the language of the person asking the initial and follow up questions

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.)*

Patients who replied ´No" to initial question: "Are you uncomfortable because of pain?"

Patients under 18 years of age

Patients who are unable to understand the language of the person asking the initial and follow up questions

Patients who cannot self-report pain

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.)*

None

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.)*

Calculation of measure score:

1. Identify number of patients admitted to hospice services during the timeframe of interest (e.g., CY quarter).
2. Identify number of admitted patients who were able to respond to the question "Are you uncomfortable because of pain?" during the initial assessment and were not excluded because they met the exclusion criteria.
3. Identify the number of patients who responded "yes" to the question "Are you uncomfortable because of pain?" during the initial assessment.

4. Identify the number of patients who were contacted between 48 and 72 hours of the initial assessment and responded "yes" to the question: "Was your pain brought to a comfortable level within 48 hours of the start of hospice services?" This number is the numerator.
5. Divide the number of patients whose pain was brought to a comfortable level within 48 hours after initial assessment by the number of patients who reported they were uncomfortable because of pain at the initial assessment.
6. Multiply this number by 100 to get the hospice's score as a percent. This is the proportion of patients who reported being uncomfortable because of pain at initial assessment whose pain was brought to a comfortable level within 48 hours of the start of hospice services.

NOTE: A Problem Score may also calculated as a complement to the measure score The Problem Score is calculated by dividing the number of patients whose pain was NOT brought to a comfortable level within 48 hours after the initial assessment by the number of patients who were uncomfortable on admission. Multiply this number by 100 to get the hospice's score as a percent. A lower score/percentile = better performance. The Problem Score is useful for assessing the proportion of patients for whom comfort was not achieved and subsequent root cause analysis for quality improvement purposes.

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.

No sampling methodology required. All patients are assessed for eligibility for inclusion in the measure at the initial assessment.

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.

STEP 1: AT INITIAL ASSESSMENT

Prior to performing a comprehensive pain assessment, the nurse first determines if the patient is eligible for inclusion in the measure.

If the patient meets the eligibility criteria, the nurse asks the question "Are you uncomfortable

because of pain?"

If the patient responds "yes," the patient is included in the measure.

If the patient responds "no" the patient is not included in the measure.

The nurse documents the patient's response and proceeds with the comprehensive pain assessment using whatever pain scale or assessment tools are appropriate for the patient. Pain management strategies and interventions are instituted based on the pain assessment.

STEP 2: FOLLOW-UP

Between 48 and 72 hours after the initial assessment, the patient is contacted and asked: "Was your pain brought to a comfortable level within 48 hours of the start of hospice care?"

The patient's yes or no response to the question is then documented.

If the patient is unable to self-report, that should be documented. For quality improvement purposes, it is also desirable to document the reason that the patient is unable to self-report(discharged due to death, discharged alive, disease progression/unable to communicate, other reasons).

The follow-up assessment can be completed in person or by telephone, but the patient must self-report his/her own response to the question by answering "yes" or "no. The follow up assessment does not need to

be done by the nurse who performed the initial assessment and can be done by any staff member who has experience communicating with patients.

If the patient seems to have difficulty understanding the 48 hour timeframe for achieving comfort, reframing the question using language that is more natural for the patient is permissible, as long as the question of achieving comfort within the prescribed timeframe of 48 hours of the initial assessment is kept

intact.

Patient responses to the initial measure question and the follow up measure question should be recorded in the patient medical record.

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

If other, please describe in S.18.

Instrument-Based 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.

Data specific to measure (initial question on admission and follow-up question asked between 48 and 72 hours of admission) recorded by hospice. Data can be part of patient record or recorded and tracked separately.

Data are aggregated and submitted quarterly by hospices to NHPCO which maintains a national data repository. NHPCO analyzes the data and produces a quarterly national level report for hospices as a source of comparative data for use in performance improvement initiatives.

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)*

Available at measure-specific web page URL identified in S.1

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

Facility, Other

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

Home Care

If other:

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.)*

2. Validity – See attached Measure Testing Submission Form

NQF_testing_attachment_Sep2017_-1--637372241262404932_-1-.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.

No

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.

No

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*)**:** 0209
Measure Title: Comfortable Dying
Date of Submission: 7/31/2020

Type of Measure:

Measure	Measure (continued)
☒ Outcome (*including PRO-PM*)	☐ Composite – *STOP – use composite testing form*
☐ Intermediate Clinical Outcome	☐ Cost/resource
☐ Process *(including Appropriate Use)*	☐ Efficiency
☐ Structure	*

*cell intentionally left blank

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:	☐ other:

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*). Data used in testing were largely drawn from the Patient Outcomes surveys that NHPCO does on a continuing basis.

1.3. What are the dates of the data used in testing? 2009-2014

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

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.20*)	Measure Tested at Level of:
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other:	☐ other:

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*)

A sample of quarterly data submissions was taken covering two years (2009 and 2010) worth of submissions. The sample consisted of only those agencies that submitted multiple (=2) quarters worth of data during that period. There were 79 hospices agencies that submitted usable data for the Comfortable Dying measure covering 285 quarters (in total) worth of data and nearly 50,000 patients. Of those 79 hospice agencies, 58 (73.4%) provided multiple quarters worth of data during that period, covering data on over 38,000 patients. The two-year quarterly average percent of patients reporting being uncomfortable due to pain on admission was 20.8% (95% CI 19.5% - 22.1%). The two-year quarterly average percent of patients reporting having their pain brought to a comfortable level within 48 hours of admission was 69.3% (95% CI 66.3% - 72.3%).

Data were provided by 484 hospices, on a voluntary basis.

For the test sample, 181 hospice agencies were selected who provided Comfortable Dying data to NHPCO during the 2013. Agency characteristics of the sample are included in the table that follows.

Agency Characteristics	N = 181	Percent
Average Daily Census	*	*
1. 1 to 49	37	20.4%
2. 50 to 150	17	9.39%
3. 151 to 500	61	33.7%
4. 501 to 1,500	49	27.1%
6. >1,500	17	9.39%
Geographic Area Served	*	*
Primarily Rural	28	23.3%
Primarily Urban	22	18.3%
Mixed Urban and Rural	70	58.3%
Agency Tax Status	*	*
Not-for-profit	123	75.9%

Agency Characteristics	N = 181	Percent
For profit	34	21.0%
Government	5	3.1%

*cell intentionally left blank

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*)

Testing used data from 16,778 patients who qualified for the denominator of the measure.

The sample included 90,453 admissions to the hospices in 2013. There were 9,059 patients that met the inclusion criteria and indicated that they were uncomfortable due to pain so were, thus, included in the measure.

NHPCO only collects aggregate hospice-level data for the Comfortable Dying measure and does not routinely obtain patient-level data. To obtain patient-level characteristics needed for testing, NHPCO selected two hospices that regularly participate in the Comfortable Dying measure to provide patient-level characteristics on all patients admitted over a six-month period. Characteristics of this sub-sample are as follows:

Patient Characteristics	N =	Percent
Gender	*	*
Male	1,049	45.0%
Female	1,281	55.0%
Age at Death		
80 years or older	1,318	56.6%
65 - 79 years	642	27.6%
20 - 64 years	359	15.4%
19 years or younger	11	0.5%
Ethnicity	*	*
Hispanic	50	2.2%
Non-Hispanic	2,277	97.9%
Race	*	*
White	2,159	94.8%
Black or African-American	102	4.5%
Another race or multiracial	16	0.7%

*cell intentionally left blank

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.

In most aspects of facility-level testing, 97 hospices were chosen. A study of possible stratification by age or gender used patient-level data from two hospices. A patient-level validation test was done on a special sample of 212 patients.

Measure level testing was conducted with data collected by NHPCO as part of its quarterly Patient Outcomes and Measures protocol. These data are collected and reported as aggregate numbers (total admissions, total participating in the pain measure, etc.…), then summarized and reported quarterly for participating hospice to use while interpreting their own Comfortable Dying Results.

Instrument level testing was conducted using a sub-sample of two hospice organizations with a total of eight hospice locations. Six months of complete patient-level data related to the Comfortable Dying measure was provided on all admissions by the two organizations for use in testing.

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.

Geography, ethnicity, and race

NHPCO conducted an analysis of potential risk factors at the facility level. Scores used in this analysis combined data from 2012 and 2013. Because the distribution of scores showed departures from normality, nonparametric statistical tests were used (Kruskal-Wallis, Wilcoxon).

Factors examined included: geographic location; service area; ownership; race and ethnicity; patient age; patient gender; patient principle diagnosis; and referral source. None of these factors demonstrated an association with the measure scores and no trend was seen between facilities with different proportions of patients in any of the risk factor categories.

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*)

Intraclass correlation was used for reliability testing for the measure. To provide evidence of measure reliability we must show that, all things being equal, hospices will reliably submit the same data over multiple quarters. Put another way, given that the proportion of patients whose pain is brought to a comfortable level within 48 hours of admission does not significantly change between quarters, the reported proportion will also remain the same.

To test this hypothesis, agency-level results were calculated from the sample hospice for the percent of hospice patients reporting being uncomfortable due to pain on admission, and the percent of patients who report having their pain brought to a comfortable level within 48 hours after admission. Univariate analysis was performed to provide the overall distribution of results for both variables results. To examine the similarity of data submitted in each quarter, an analysis of variance was performed to determine if significant differences existed in between the quarterly means for both agency level results. Next, an analysis of variance was performed to examine the differences in mean scores between and among hospices over the two years. Finally, intra-class correlations coefficients (ICC) were calculated to examine the measurements reliability over the sample years. Statistical significance was set at $P < 0.05$. All analysis completed utilizing SAS version 9.2.

NOTE: Test-retest is a frequently used method for reliability testing with single item measures and has been used with pain measures. However, the Comfortable Dying Measure assesses a characteristic that can inherently be expected to change rapidly (interventions to achieve better pain control can be and often are instituted at the time of assessment) making test-retest an inappropriate choice for reliability testing for this measure.

Analysis considered the consequences of random differences between patients by modeling changes within one facility as a binomial distribution. Tests for changes used Fisher exact or other exact statistical tests for change in proportions.

Instrument Level:

To provide evidence of measure reliability we must show that, all things being equal, hospices will reliably submit the same data over time. Put another way, given that the proportion of patients whose pain is brought to a comfortable level within 48 hours of admission does not significantly change between quarters, the reported proportion will also remain the same.

To test this hypothesis, an analysis of variance was performed to determine if significant differences existed in between the quarterly means for the percent uncomfortable and the percent whose pain level was brought to a comfortable level at agency level results. If no significant difference between quarters then we will conclude that the agencies are implementing the Comfortable Dying Measure consistently over time.

Measure Level:

To assess reliability of the computed performance measure on the hospice organization level, a signal to noise analysis was performed. An analysis of variance (ANOVA) using a random effects model was performed with the percent whose pain was brought to a comfortable level as the dependent variable and the Facility ID as the class variable. Taking the generated F-Value from the ANOVA, we then calculated the signal to noise ratio. The signal to noise ratio is an estimate of the score variance that is due to true differences between hospice organizations (signal) to the amount of variance caused by random error (noise). This measure is also known as interunit reliability (IUR).

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*)

The analysis of variance of quarterly mean percentages of patients who reported being uncomfortable due to pain on admission showed no significant difference of mean scores between quarters (F-value = 1.11; P = 0.355). Variance of this measure demonstrated the expected significant difference between submitting hospices agencies (F-value = 7.48; P<0.0001). The intra-class correlation coefficient for the difference of the between and within hospice variation was 0.76 (95% CI 0.70 – 0.81).

The analysis of variance of quarterly mean percentages of patients who reported having their pain brought to a comfortable level within 48 hours of admission also showed no significant difference of mean scores between quarter (F-value = 1.7; P=0.991). The Hospice level variance analysis of this measure showed significant differences between hospice agencies (F-value = 5.87; P<0.001). The intra-class correlation coefficient for the between and within hospice 0.71 (95% CI 0.63 – 0.77).

The analysis of the data showed that indeed, over two-years of quarterly data submissions, the percent of patients reporting being uncomfortable due to pain remained relatively constant. Since the assumption of similarity between quarters was met it was then safe to examine the relative between and among variation in results for the same measure. As expected, there were significant differences in the percent of hospice patients uncomfortable due to pain on admission reported by each hospice. However, the ICC demonstrated good (over 75%) consistency of results within hospices from quarter to quarter.

Similarly, these results show that the percent of patients whose pain was brought to a comfortable level within 48 hours of admission, remained non-significantly differently. In fact, the results show that there was nearly no difference from quarter to quarter the results for this measure. The ICC for this measure also demonstrated good consistency (approximately 71%) of results within hospices from quarter to quarter. This slightly smaller ICC for measure (when compared to the percent uncomfortable due to pain on admission ICC) is not necessarily an indication of reduced reliability. Increased within hospice variation would be expected as hospices make process changes to increase their score for this measure. Indeed this expectation is observed in the variance of hospice scores for percent of patients uncomfortable due to pain on admission compared to those whose pain was brought to a comfortable level within 48 hours (F-values = 7.48 and 5.87 respectively).

It is likely that both ICC scores are conservative estimates of the true reliability of the measure. Even though there was little quarterly change in the percent of patients uncomfortable due to pain on admission (and likewise having their pain brought to a comfortable level), common sense dictates that real differences actually occurred at the hospice level. Since we know that the assumption of consistency of the base data can't be exactly true, we know that the true ICC's for these measures must be higher than what was observed.

In conclusion, this analysis provides statistical evidence that the NHPCO Comfortable Dying measure has good reliability.

This measure is concerned with newly admitted patients. It cannot be repeated on the same population of patients because each patient had only one initial period of 48 hours after hospice admission. There are also some very real constraints about how many times a patient can be asked whether comfort was attained in the first 48 hours. If the patient reports that comfort was not attained, the clinician may react by immediately increasing the dose of analgesics. The patient's subjective recollection of pain at 48 hours could change as a result.

Reliability of this measure may, however, be considered by two other avenues. Basic probability theory considerations give a guideline for how precise and repeatable the measure can be during random variations in the characteristics of individual patients admitted. Additionally, experience with real data gives an impression about whether the measure tends to stay constant between successive time periods.

Variations between individual patients will cause the numerator of this measure to fluctuate in accordance with the usual binomial distribution, even if a hospice keeps a completely constant pain management strategy and continues to admit patients with the same average characteristics. The measure itself, numerator/denominator, will show less random variability as the denominator increases.

We consider here the behavior of this measure when the denominator is at least 50.

In a 2013 nationwide survey involving over 16,000 patients who qualified for the denominator, we found 58% of those patients qualified for the numerator. A guideline for the random variability of the measure is provided by supposing that a hospice had admitted 50 patients, chosen at random from all the patients in the nationwide denominator, and that the measure was computed with the outcomes that occurred in the national sample. Such a hospice would have an average comfortable dying measure of 58%. It would have an 80% chance of being assigned a measure between 48% and 68%, and a less than 1% chance of being assigned a measure less than 38% or greater than 78%.

A survey covering both 2013 and 2014 gives some experience about the variability of the measure between successive time periods. Data was considered from 22 hospices whose denominators exceeded 50 in both of those years. Only 32% of those hospices had a measure that changed by more than 15 percentage points from 2013 to 2014, and only three of those hospices had a measure that changed by more than 20 percentage points. Changes were somewhat larger than those expected by pure random variation, but still consistent with good reliability for the measure. We expect more change than random variation in some subset of the hospices. Some hospices may be changing their strategy for pain management or may be taking referrals from different sources.

Instrument Level:

Percent uncomfortable due to pain: F=0.42 p=0.5115

Percent whose pain was brought to a comfortable level: F=0.77 p=0.3814

Measure Level (Organization Level):

F-value = 4.04 ($p < 0.001$): IUR =0.75

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?*)

Instrument Level:

The analysis of variance p-values for both the percent of patients being reported as uncomfortable due to pain and the percent of those who had their pain brought to a comfortable level within 48 hours were less than 0.05. The values indicate that no significant difference between location level scores across quarters. Assuming that there were no significant differences in patient populations between quarters, we would thus conclude that the measure was being implemented consistently at the location level.

Measure Level (Organization Level):

The analysis of variance showed a statistically significant difference of the organization score values (F = 4.04; p<0.001). The IUR of 0.75 indicates that the vast majority of variance across provider scores is due to true differences being measured. As such, we can conclude that the computed performance measure is very reliable.

Results are consistent with good reliability for the test when the denominator exceeds 50.

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)*

Initial testing of measure performed with 686 patients in 9 hospices. Of those, 212 (31%) indicated that they were uncomfortable because of pain at the initial assessment on admission to hospice services.

Criterion (concurrent) validity was tested by using two different wordings for the follow up question related to whether pain was managed. Patients first were asked if their pain was brought to a comfortable level within 48 hours and then they were asked if their pain was brought to an acceptable level within 48 hours. These two forms of the follow-up question were judged by the expert panel for the Comfortable Dying Measure to be equivalent in that they equally reflected patient preference and level of effectiveness achieved for pain management.

A special test dataset was created by using data from nine hospices. The dataset included 212 patients who qualified for the denominator of the measure. 127 of those patients stated that pain was brought to a comfortable level within 48 hours, and 136 of the 212 stated that pain was brought to an acceptable level within 48 hours. Because 96% of patients gave the same answer to the two wordings of the follow up question, the results indicate good concurrent criterion validity for the measure.

Instrument Level:

<<correlation analysis of the relationship between FEHC scores on pain management being sufficient (B2) an 0209 scores failed to find a relationship. As such, this text is simply to remind you that we discussed using "Criterion Validity" to address the instrument level validity>>

Measure Level (Organization Level):

To assess validity at the measure scale we compared contemporary results from the Composite Score of the Family Evaluation of Hospice Care (NQF 0208) and the average Comfortable Dying scores. The hypothesis is, if Comfortable Dying scores are associated with high performing hospices, then they will be significantly associated with performance on another performance measure. To test this hypothesis we identified high performing hospices on NQF 0208 and performed a logistic regression with the hospice Comfortable Dying Score.

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

Instrumental level

Sixty percent (N = 127) of the patients who initially responded that they were uncomfortable because of pain responded that their pain was brought to a comfortable level within 48 hours. Of those same patients, 64% (N = 136) responded that their pain was brought to an acceptable level within 48 hours. The two questions

elicited very little difference in the proportion of patients replying that their pain was brought under control, indicating acceptable concurrent criterion validity of the measure.

Cohen's kappa = 0.91.

Measure Level (Organization Level):

Logistic regression found a direct and marginally significant association between being rated a high performing hospice based on FEHC Composite Score (NQF 0208) and having a higher Comfortable Dying score ($\chi^2 = 3.352$ p=0.067).

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?*)

Instrument Level :

N/A

Measure Level (Organization Level):

The results indicate that high performing hospices on the FEHC Composite Score (NQF 0208) were more likely to have a higher score on the Comfortable Dying measure. The statistical was marginal for this test but it should be noted that only 90 hospices were able to be included in the logistic regression because not all hospices participated in both measures during 2013. It is expected, however, that with a larger sample size, statistical significance would be achieved.

The results support validity of the measure for understanding how well hospices were able to relieve patients admitted in pain.

_____

2b2. EXCLUSIONS ANALYSIS

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

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*)

To evaluate how exclusions might affect validity we examined the frequency of *included* and *excluded* responses across respondent demographics and organizational characteristics utilizing the sub-sample discussed earlier.

Frequency Distribution across the Target Population:

To evaluate exclusions on the target population we examined the frequency of *included* and *excluded* responses across patient demographic characteristics. Specifically, we looked to determine if there were significant and meaningful differences in the proportion of excluded responses across the demographic characteristics of ethnicity, race, gender, and age. To perform this analysis we created frequency distribution tables then performed a chi-square goodness of fit test. The chi-square tests the null hypothesis that there are no significant differences in the amount of included or excluded surveys between demographic groups. If the test is significant to a p-value of 0.05 or less then we reject the null hypothesis and conclude there are significant differences between groups.

Where significant differences are found, we also calculated "expected" cell frequencies. The expected cell frequency represent the expected frequency of responses should the null hypothesis be true. This allows us to evaluate the departure from the expected number of excluded responses under the null hypothesis. This is an important additional step since the analysis is likely to be overpowered with such a large sample size.

Frequency Distribution across the Measured Entities:

To evaluate the effect of exclusions across organizations, we calculated the percent of patients who were at a Comfortable Dying for each hospice location in the sub-sample. We then examined the distributions of the mean percentage of included responses across provider characteristics to determine if meaningful differences exist between means.

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*)

Frequency Distribution across the Target Population:

The following table details the distribution across patient characteristics of included and excluded from the Comfortable Dying measure:

Demographic Category	Included: (1)Frequency (2) (Expected freq) (3) Column %	Excluded: (1)Frequency (2) (Expected freq) (3) Column %
Ethnicity (P= 0.068)	*	*
Hispanic (1)	37	13
Hispanic (2)	(43)	(7)
Hispanic (3)	1.9%	3.9%%
Non-Hispanic (1)	1,953	324
Non-Hispanic (2)	(1,947)	(330)
Non-Hispanic (3)	98.1%	96.1%
Race (P=0.046)	*	*
All Other Races (1)	14	2
All Other Races (2)	(13.7)	(2.3)
All Other Races (3)	0.7%	0.6%
Black (1)	96	6
Black (2)	(87.5)	(14.5)
Black (3)	4.9%	1.9%
White (1)	1,843	316
White (2)	(1,851.8)	(307.2)
White (3)	94.4%	97.5%
Gender (P = 0.003)	*	*
Male (1)	922	127
Male (2)	(897.3)	(151.7)
Male (3)	46.3%	37.7%
Female (1)	1,071	210
Female (2)	(1095.7)	(185.3)
Female (3)	53.7%	62.3%
Age (P=0.001)	*	*
80 years or older (1)	1,093	225

Demographic Category	Included: (1)Frequency (2) (Expected freq) (3) Column %	Excluded: (1)Frequency (2) (Expected freq) (3) Column %
80 years or older (2)	(1,127.4)	(190.6)
80 years or older (3)	54.8%	66.8%
65 - 79 years (1)	564	78
65 - 79 years (2)	(549.1)	(92.9)
65 - 79 years (3)	28.3%	23.2%
20 - 64 years (1)	332	27
20 - 64 years (2)	(307.1)	(51.9)
20 - 64 years (3)	16.7%	8.0%
19 years or younger (1)	4	7
19 years or younger (2)	(9.4)	(1.6)
19 years or younger (3)	0.2%	2.1%

*cell intentionally left blank

Frequency Distribution across the Measured Entity:

The following table details the mean percentage of included responses for each agency characteristic:

Hospice Characteristic	N	Mean	Upper CL	Lower CL	Std Dev
Geographic Area Served	*	*	*	*	*
Rural	32	26.2%	30.1%	22.3%	10.8%
Urban	27	32.9%	37.7%	28.1%	12.0%
Mixed	82	34.2%	36.7%	31.7%	11.3%
Average Daily Census	*	*	*	*	*
< 100	24	25.0%	29.8%	20.1%	11.6%
101 to 200	13	30.5%	35.7%	25.4%	8.6%
> 200	19	35.6%	39.0%	32.1%	7.2%
Tax Status	*	*	*	*	*
Not-for-profit	125	32.1%	34.0%	30.3%	10.5%
For profit	34	35.3%	40.8%	29.7%	16.0%
Government	5	20.2%	36.8%	3.6%	13.4%

*cell intentionally left blank

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*)

Frequency Distribution across the Target Population:

Frequency analysis showed significant differences in the likeliness of being excluded or included based on the characteristics of race, gender, and age. Closer analysis of race and gender frequency distributions show that, where differences are found, the groups are more likely to be included than expected. For example, the only

major difference between expected and observed frequency of inclusions in the race category were for those identified as African American, where were substantially more frequently included than would expected. A similar result was observed for wen, where they were included more frequently than would be expected.

The driver behind the significant difference based on age was found in the 80 years or older group. Individuals in this group were observed to be excluded substantially more often than would expected (225 observed vs. 197 expected). This is likely a reflection of the types of diseases more often experienced by older populations (e.g., Alzheimer's disease, dementia, etc.) which would prohibit the ability to self-report.

Distribution across the Measured Entities:

Statistically significant but not substantial differences were found between the percent of respondents excluded from the Comfortable Dying measure and provider demographic characteristics over the course of the year. The significant differences were found in the follow groups; rural and mixed urban and rural providers, providers with an ADC <100 and those >200, and for-profit and government agencies.

There were three exclusions for this measure: patients less than 18 years old; patients who were unable to self-report at the time of admission; and patients who are unable to communicate and understand the language of the person asking the question.

Most hospices have few patients younger than 18, and so that exclusion has little impact. Patients truly unable to self-report must be excluded simply because this is a patient-reported measure. The same is true for patients who are unable to understand and communicate the language of the person asking the question. Because this is a patient reported measure, the responses to the initial and the follow-up measure questions must be from the patient and not a proxy. The use of an interpreter, however, is permitted if the patient cannot understand the language of the clinician conducting the assessment. Use of a qualified interpreter will suffice to surmount the language barrier and include the patient in the NQF #0209 measure. The same standard regarding use of an interpreter for the comfort question(s) as for any regular assessment or visit.

There was no exclusion for patients who were reported to become unable to self-report for the follow up question, after they were considered able to self-report at admission. Those patients were counted for the measure just the same as if they had responded to the follow up question by saying that they had not been made comfortable at 48 hours. The measure was designed in that way to give providers an incentive to persist with attempting to ask the follow up question. Because some patients actually do become completely unable to self-report, the result is that the comfortable dying measure will report slightly less than the true percentage of patients who are made comfortable at 48 hours.

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 2b5.

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

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with risk factors**

☐ **Stratification by risk categories**

☐ **Other**

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.

N/A

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.

The standard of care for hospices is to provide timely and effective pain management based on patient preferences for all patients regardless of primary diagnosis, underlying mechanism for pain, or other patient

characteristics, including pain intensity rating. Because the measure is based on the patient´s statement of comfort/discomfort no adjustment is necessary (e.g., for patients who report a high pain intensity but refuse intervention aimed at lowering pain intensity levels).

For this measure, no clear effect has been demonstrated for readily identifiable patient characteristics.

No risk adjustment strategy has been included in this measure.

A sample of 2,329 patients was used to investigate possible stratification by age or gender. Of those patients, 383 qualified for the denominator of this measure. Of those patients, the measure showed no statistically significant difference between the patients younger than 65 and the patients aged 65 and over. There was, likewise, no statistically significant difference between male patients and female patients.

We maintain that risk adjustment for NQF measure #0209 is not necessary.

We contend that the customary approach and factors utilized in risk adjustment are not appropriate for pain management in hospice and palliative care. The standard of care in hospice is to provide timely and effective pain management based on patient preferences for all patients regardless of primary diagnosis, underlying mechanism for pain, or other patient characteristics, including pain intensity rating. Hospice providers should be equally responsible for optimizing pain management for all patients who state they are uncomfortable on the initial pain screening.

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?

Factors examined included: geographic location; service area; ownership; race and ethnicity; patient age; patient gender; patient principle diagnosis; and referral source. None of these factors demonstrated an association with the measure scores and no trend was seen between facilities with different proportions of patients in any of the risk factor categories.

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?

We conducted an analysis of potential risk factors at the facility level. Scores used in this analysis combined data from 2012 and 2013. Because the distribution of scores showed departures from normality, nonparametric statistical tests were used (Kruskal-Wallis, Wilcoxon).

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.**

Risk adjustment for measure NQF #0209 is not needed because providers have an equal responsibility to provide timely pain management for all patients. The lack of need for risk adjustment is supported by the evidence in the supplementary analyses provided. Moreover, this measure fills a critical gap in the field of palliative and hospice practice. The Comfortable Dying measure is one of the very few patient reported outcome measures which NQF, and most importantly patients, consider highly desirable.

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*)

N/A

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*)**:**

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

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

2b3.9. Results of Risk Stratification Analysis:

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*)

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*)

_____

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)*

To examine the ability of the performance measure to identify differences among hospice providers, we performed three analyses. First, we examined the distribution of Comfortable Dying scores across hospice organizations. Then we assigned each organization to a performance category based on the quartile where its Comfortable Dying Score fell (first quartile = low performance; second quartile = medium-low performance; third quartile = medium-high performance; fourth quartile = high performance). An analysis of variance was then performed to determine if significant differences existed between Comfortable Dying scores of providers based on their performance rating. A Tukey's test was performed to determine if the Comfortable Dying scores varied significantly across groups.

From 2004 through 2010 the National Hospice and Palliative Care Organization has collected aggregate data from hospices for the Comfortable Dying Measure. Data collected during that time provide evidence for an overall less-than-optimal performance by participant providers. Deviation from the national mean and the presence of providers with substantially higher (better) scores indicate that performance of individual hospices and the industry as a whole can be significantly improved.

The data were obtained through an ongoing collection effort by the NHCPO and submitted by hospices voluntarily providing their aggregated data. From 2004 through 2007, hospices submitted data annually through the NHPCO Data Analysis and Reporting Tools (DART) system and by manual submission of raw data files (e.g., CSV files). From 2008 to the present, participating hospices voluntarily submit data on a quarterly basis only through the DART system.

After collecting data for the specified period of time (one year / one quarter), hospices reported to NHPCO their aggregated numerator and denominator totals. The numerator represents the total number of hospice patients who reported being uncomfortable due to pain on admission and were made comfortable with regards to pain within 48 hours after admission. The denominator value represents the total number of patients admitted to the hospice during the time period who self-report being uncomfortable due to pain on admission. Hospices also reported time-period totals for admissions, patients self-reported comfort level due to pain (uncomfortable, not uncomfortable, not able to participate), and patient's comfort level due to pain after admissions (limited to patients reporting being uncomfortable due to pain on admission).

After the submission period ends, agency-level data are aggregated to the national-level to produce the national mean percent of; admissions participating in the pain measure protocol, patients uncomfortable due to pain on admission, and patients whose pain was brought to a comfortable level within 48 hours after

admission to hospice. National means as well as agency quartile scores are reported in a National Summary Report for hospices use to compare to their own results.

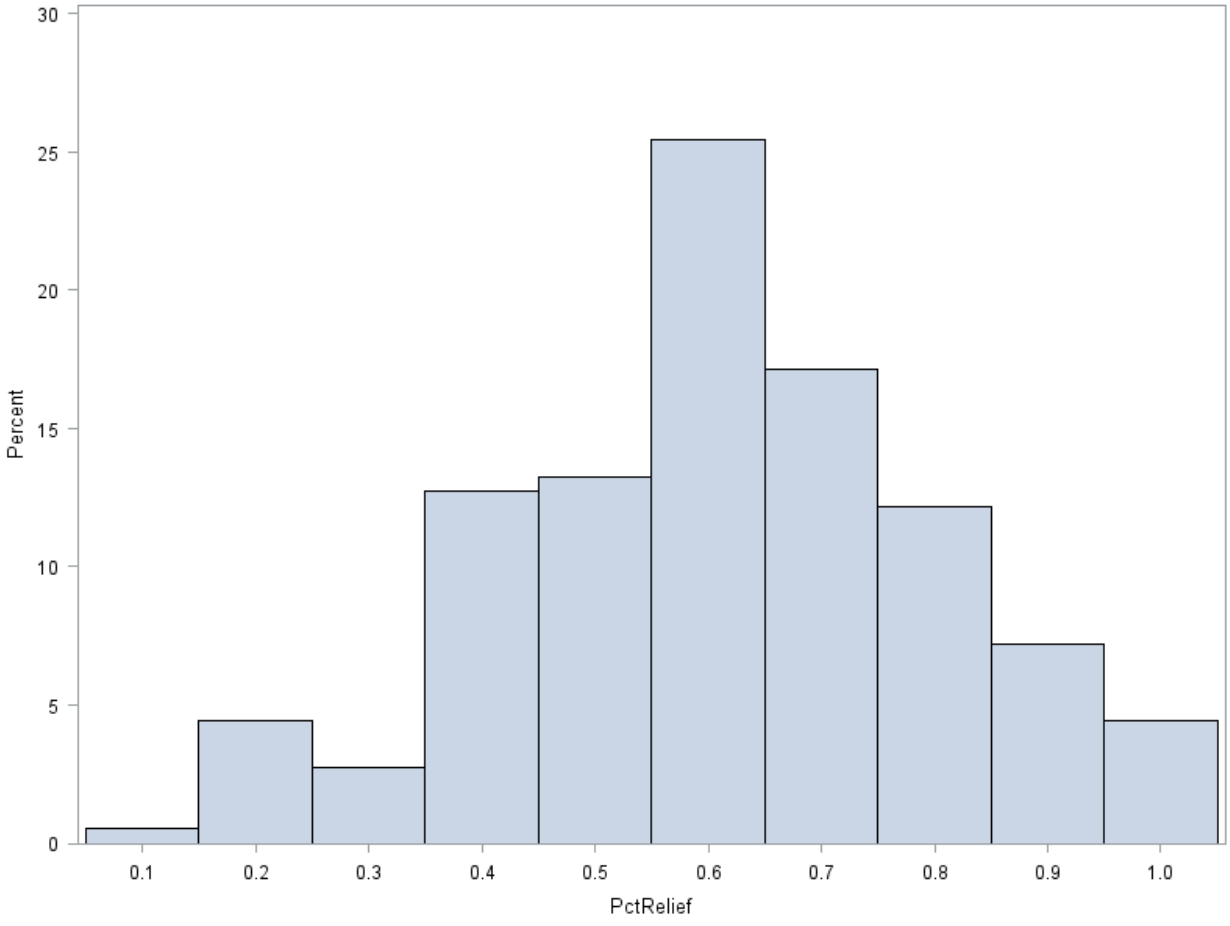
Hospices evaluate their individual results for subpar performance by comparing their percent of patients whose pain was brought to a comfortable level within 48 hours of admission with the national mean and quartile scores. A score below the national average, or even below the 75th percentile, generally indicates significant room for improved pain management care.

It is accepted in the field that there is a clinical significance to a change that results in a long-term decrease from 60% to 40% for the fraction of patients who are in pain on admission and are made comfortable by 48 hours. It is likewise accepted that an increase from 60% to 80% is clinically important.

It is less obvious whether such differences from the national average exist, and can be shown statistically significant. In order to address that question, we used data from 97 hospices that had denominators greater than 50 patients, for the year 2013.

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*)

The seven year Comfortable Dying Measure data collection by NHPCO represents a sample of more than 625 hospice providers, reporting on over 470,000 hospice patients. Data were collected annually from 2004 through 2007 and then quarterly from 2008 through the present. The seven-year national mean score of 72.2% (SD = 4.2% 95% CI = 68.4% to 76.1%) indicates that more than a quarter of hospice patients do not receive sufficient interventions to bring their pain to a comfortable level within 48 hours after admission to hospice. The yearly national averages have stayed within a relatively narrow range of scores (minimum = 65.3%, maximum 77.4%) indicating a consistent measure performance over time.



More recent results obtained from the quarterly submissions of hospices during 2010, show a wide range of individual hospice performance within the quarter. The 2010 mean national percent of patients whose pain was brought to a comfortable level within 48 hours of admission was 72.6% (95% CI 69.1% - 76.2%). The 75th percentile of hospice's performance each quarter for 2010 was 94.7%, 98%, 100%, and 96.2% while the 25th percentile was 50%, 61.5%, 62.5%, and 55.6% respectively.

Using an exact test based on the binomial distribution, facility comfortable dying scores were compared with the average for the group. 16 of those hospices were significantly different from average at the p<.05 level, and 21 of them were significantly different from national average at the p<.1 level.

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 distribution of FEHC Composite Scores across providers reveals a large spread between the top performing hospice and the lowest performing hospice (100% and 13% respectively). The first, second, and third quartiles were 49%, 62%, and 73% respectively. The histogram shows a good approximation of a normal distribution of Comfortable Dying scores. The analysis of variance and subsequent Tukey's test revealed statistically significant differences in the Comfortable Dying scores across all performance categories.

These results indicate that Comfortable Dying scores are well distributed with a substantial range of scores. Statistically significant differences between hospices are measurable by the Composite Score, and those differences are also meaningful.

The result shows that the comfortable dying measure allows identification of some hospices that have scores enough worse than or better than the national average that the differences cannot be explained by simple random variation in the characteristics of patient populations.

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*)

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*)

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*)

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 non-responders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

Since this measure is implemented as part of the normal process of care, truly missing data is not a significant issue. However, a certain proportion of patients are not able to respond to the follow-up question either because they have unavailable, incapable of responding, or are deceased. To analyze the effects of these individuals who are lost-to-follow-up on the measure, we performed a frequency analysis on patient demographic characteristics to determine if there are systematic difference in loss-to-follow-up.

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*)

Demographic Category	Included: Frequency (Expected freq) Column %	Lost-to-Follow-up: Frequency (Expected freq) Column %
Ethnicity (P= 0.276)	*	*
Hispanic (1)	3	2
Hispanic (2)	(3.9)	(1.1)
Hispanic (3)	1.3%	3.1%
Non-Hispanic (1)	223	62
Non-Hispanic (2)	(222.1)	(62.9)
Non-Hispanic (3)	98.7%	96.9%
Race (P=0.430)	*	*
All Other Races (1)	1	1
All Other Races (2)	(1.6)	(0.4)
All Other Races (3)	0.5%	1.6%
Black (1)	6	3
Black (2)	(7.0)	(1.9)
Black (3)	2.7%	4.8%
White (1)	216	58
White (2)	(214.4)	(59.6)
White (3)	96.9%	93.6%
Gender (P = 0.063)	*	*
Male (1)	99	20
Male (2)	(92.4)	(26.6)
Male (3)	43.8%	30.8%
Female (1)	127	45
Female (2)	(133.6)	(38.4)
Female (3)	56.2%	69.2%
Age (P=0.510)	*	*
80 years or older (1)	83	29

Demographic Category	Included: Frequency (Expected freq) Column %	Lost-to-Follow-up: Frequency (Expected freq) Column %
80 years or older (2)	(87.0)	(25.0)
80 years or older (3)	36.7%	44.6%
65 - 79 years (1)	77	20
65 - 79 years (2)	(75.3)	(21.7)
65 - 79 years (3)	34.1%	30.8%
20 - 64 years (1)	66	16
20 - 64 years (2)	(63.7)	(18.3)
20 - 64 years (3)	29.2%	24.6%
19 years or younger (1)	0	0
19 years or younger (2)	(0)	(0)
19 years or younger (3)	0.0%	0.0%

*cell intentionally left blank

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 non-responders) 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*)

There were no significant differences found in the proportion of individuals lost to follow-up based on the demographic characteristics of age, gender, race, or ethnicity. As such, there is no evidence that all individuals are equally as likely to be available to answer the follow-up question. And, likewise, there is no evidence of a systematic difference in LTF based on these categories.

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 by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition

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.***

Some data elements are in defined fields in electronic sources

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

Not all providers may collect the measure data as part of the patient electronic record. Those providers who do not can keep separate paper records of the measure question responses for individual patients. Data are aggregated for submission to NHPCO which is done online. NHPCO provides a downloadable Data Submission Worksheet for providers to print out and complete before entering data online on the NHPCO website.

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

NHPCO maintains ongoing support(in the form of written materials and one-on-one guidance)for hospice providers who use the measure. Hospices vary in size and resources, and data collection strategies employed tend to vary with the individual characteristics of the hospices. We regularly plan and implement modifications to support materials to improve clarity and assist hospice with implementation of the measure.

when the measure was required by CMS for the first year of the Hospice Quality Reporting Program, many hospices reported difficulties with measure implementation. For example, understanding that the measure questions were separate from pain assessment proved problematic. However, hospices were not accustomed to implementing a quality measure with specification that could not be modified and also, unless a hospice was already using 0209, had no experience with a PRO measure. Had 0209 been implemented later in the HQRP program and/or given more time along with education and support, hospices would likely have had more success with implementation.

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

There are no costs or other requirements imposed by NHPCO associated with use of this measure. There is open access from the NHPCO website for all materials provided for support of measure implementation.

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)
Regulatory and Accreditation Programs	Quality Improvement (Internal to the specific organization) NHPCO Patient Outcomes and Measures https://www.nhpco.org/pom/

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

NHPCO is the sponsor of this voluntary measure which is part of our quality program. The purpose of the measure is to ensure timely recognition of discomfort due to pain and prompt and effective intervention. Currently, there are no other quality measures reported during hospice care. We no longer collect data on this measure but we still encourage use of the measure. We are engaging our members in a new quality program in 2021, so we may revisit the opportunity to collect data on this measure. This measure is included in MIPS. The level of utilization is unknown. NHPCO provides a manual for data collection but there is no comparative reporting since 2015. For 2014, 156 hospices provided aggregated measure data for 20,548 patients.

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?*)

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.*)

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.

NHPCO maintained ongoing support(in the form of written materials and one-on-one guidance)for hospice providers who use the measure. Hospices vary in size and resources, and data collection strategies employed tend to vary with the individual characteristics of the hospices. We regularly planned and implemented modifications to support materials to improve clarity and assist

hospice with implementation of the measure.

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.

NHPCO provided quarterly national report to members through 2017. NHPCO has retired data submission and reporting; however, we continue to encourage hospice programs to use the measure as part of their QAPI program. The data collection and measurement tools continue to be available on our website.

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.

Feedback was obtained through survey and webinar meetings with participants.

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

When the measure was required by CMS for the first year of the Hospice Quality Reporting Program, many hospices reported

difficulties with measure implementation. For example, understanding that the measure questions were separate from pain

assessment proved problematic. However, hospices were not accustomed to implementing a quality measure with specification

that could not be modified and also, unless a hospice was already using 0209, had no experience with a PRO measure. Had 0209

been implemented later in the HQRP program and/or given more time along with education and support, hospices would likely have

had more success with implementation.

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

N/A

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.

NHPCO continues to encourage use of the measure as part of QAPI program even though the measure is no longer required for public reporting.

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.

Because the results reflect national level means, improvements by individual providers are not reflected. Also, those hospice patients who are able to self-report at the time the first measure question is asked may not be able to self-report at the 48-72 hour period when the follow up question is asked. These patients remain in the denominator. Some hospice may have many such patients, which will depress their measure scores. Keeping patients in the denominator is included in the measure specifications to encourage hospices to make a strong effort to contact patients to ask the follow up question. A patient population that not a rapidly functionally

declining as many hospice patients would be able to respond to both the initial and the follow up questions. This is likely to be true for the patients who are receiving palliative care who are more functional than the 1/3 of hospice patients who die within 7 days of admission to hospice services.

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.

N/A

4b2.2. Please explain 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.

No

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

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

5a. Harmonization 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?

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

5b. Competing Measures

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

OR

Multiple measures are justified.

5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

N/A

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.

Available at measure-specific web page URL identified in S.1 Attachment:

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): National Hospice and Palliative Care Organization

Co.2 Point of Contact: Lori, Bishop, lbishop@nhpco.org, 571-397-2687-

Co.3 Measure Developer if different from Measure Steward: National Hospice and Palliative Care Organization

Co.4 Point of Contact: Lori, Bishop, lbishop@nhpco.org, 571-397-2687-

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.

Members of Outcomes Forum - a group of experts that was convened and worked together over a three year period (1998 through 2000) to develop and test measures derived from a common conceptual framework as delineated in the NHPCO publication: A Pathway for Patients and Families Facing Terminal Illness. Members included:

Carla Alexander, Ina Boyd, Deborah Childs, Stephen Clauser, Chis Cody, Stephen Connor, Gail Cooney, Jeanne Dennis, Kathy Egan, Perry Fine, Melinda Garverick, Barbara Head, Marcia Lattanzi-Licht, Judi Lund-Person, Dale Lupu, Susan Mann, Melanie Merriman, Naomi Naierman, Betty Oldanie, Peggy Parks, True Ryndes, Shareefa Sabur, Sherri Solomon, Janet Snapp, Sharon Sprenger, Carol Spence, Joan Teno, Patti Thielmann.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2000

Ad.3 Month and Year of most recent revision: 06, 2011

Ad.4 What is your frequency for review/update of this measure? Annually

Ad.5 When is the next scheduled review/update for this measure? 09, 2020

Ad.6 Copyright statement: Copyright holder of the Comfortable Dying Measure is NHPCO which makes the measure available for use free of charge with the provision it is not modified or sold.

Ad.7 Disclaimers:

Ad.8 Additional Information/Comments: