NATIONAL QUALITY FORUM

Memo

TO: Palliative and End-of-Life Care Standing Committee

FR: NQF Staff

RE: Post-Comment Call to Discuss Public and Member Comments

DA: July 28th, 2016

Purpose of the Call

The Palliative and End-of-Life Care Standing Committee will meet via conference call on Wednesday, August 3rd, 2016 from 3:00-5:00 PM ET. The purpose of this call is to:

- Review and discuss comments received during the post-evaluation public and member comment period
- Provide input on proposed responses to the post-evaluation comments
- Determine whether reconsideration of any measures or other courses of action are warranted

Due to time constraints, during this call we will review comments by exception, in the case the Committee disagrees with the proposed responses.

Standing Committee Actions

1. Review this briefing memo and Draft Report
2. Review and consider the full text of all comments received and the proposed responses to the post-evaluation comments (see Comment Table)
3. Be prepared to provide feedback and input on proposed post-evaluation comment responses
4. Be prepared to re-vote on the Reliability and/or Validity subcriteria for selected measures (indicated below), and potentially to vote on those measures' Feasibility and Usability and Use. Complete measure worksheets are provided in the Appendix as needed.

Conference Call Information

Please use the following information to access the conference call line and webinar:

Speaker dial-in #: (844) 833-5571 (***Committee only***. *No conference code required*.)
Web Link: http://nqf.commpartners.com/se/Rd/Mt.aspx?621258
Registration Link: http://nqf.commpartners.com/se/Rd/Rg.aspx?621258
Public dial-in #: (844)852-2435 (*No conference code required.*)

***In order to vote, Committee members should use their individual webinar links sent via email.**

Background

For this project, the Palliative and End-of-Life Care Standing Committee evaluated 8 newly-submitted measures and 16 measures undergoing maintenance review against NQF's standard evaluation criteria. The Committee recommended 19 measures for endorsement. The Committee did not reach consensus on two measures and did not recommend two measures for endorsement. One measure was withdrawn from consideration.

Comments Received

NQF solicits comments on measures undergoing review in various ways and at various times throughout the evaluation process. First, NQF solicits comments on endorsed measures on an ongoing basis through the Quality Positioning System (QPS). Second, NQF solicits member and public comments prior to the evaluation of the measures via an online tool located on the project webpage. Third, NQF opens a 30-day comment period to both members and the public after measures have been evaluated by the full Committee and once a report of the proceedings has been drafted.

Pre-evaluation comments

The pre-evaluation comment period was open from March 28, 2016 to April 11, 2016 for all 24 measures under review. A total of 16 pre-evaluation comments were received. All of these pre-evaluation comments were provided to the Committee prior to its initial deliberations, which were held during the workgroups calls.

Post-evaluation comments

The Draft Report was released for Public and Member comment from June 20, 2016 to July 19, 2016. During this commenting period, NQF received a total of 88 comments, 52 of which were from 5 member organizations:

Consumers – 2	Professional – 0
Purchasers – 0	QMRI – 2
Health Plans – 1	Providers – 1
Supplier and Industry – 0	Public & Community Health – 0

In order to facilitate discussion, the majority of the post-evaluation comments have been categorized into major topic areas or themes. Where possible, NQF staff has proposed draft responses for the Committee to consider. Although all comments and proposed responses are subject to discussion, we will not necessarily discuss each comment and response on the post-comment call. Instead, we will spend the majority of the time considering the major topics and/or those measures with the most significant issues that arose from the comments. Note that the organization of the comments into major topic areas is not an attempt to limit Committee discussion.

We have included all of the comments that we received (both pre- and post-evaluation) in the Comment Table. This comment table contains the commenter's name, comment, associated measure, topic (if applicable), and—for the post-evaluation comments—draft responses for the Committee's consideration. Please refer to this comment table to view and consider the individual comments received and the proposed responses to each.

Committee Request for Additional Information

#0209: Comfortable Dying: Pain Brought to a Comfortable Level Within 48 Hours of Initial Assessment

During the evaluation of this patient-reported outcome-based performance measure (PRO-PM) at the in-person meeting, the Standing Committee voted against continued endorsement, primarily due to concerns about lack of risk-adjustment. Committee members were particularly concerned with potential differences in performance by region, diagnosis, and co-morbidities. For the post-comment call, the Committee requested that the developer either provide data demonstrating that risk-adjustment is not needed or provide a plan for future risk-adjustment.

NQF received 6 post-evaluation comments regarding this measure. Four of the commenters supported the decision of the Committee not to endorse the measure, with two of these agreeing that additional analyses are needed. Two commenters did not support the Committee's decision not to endorse the

measure. Both of these commenters emphasized the importance of outcome measures for pain—particularly patient-reported outcome measures—in NQF's portfolio of palliative and end-of-life care measures.

> **Developer Response**: After additional consideration, the developer, the National Hospice and Palliative Care Organization (NHPCO)), has reaffirmed its decision not to risk-adjust the measure, based on both a conceptual rationale as well as support from statistical analysis (see Appendix A). As requested by the Committee, NHPCO examined several potential risk factors at the facility level, including geographic location, service area, ownership, race and ethnicity, patient age, patient gender, patient principle diagnosis, and referral source. None of the factors examined were statistically significantly associated with the measure scores.

Given this new information, the Committee could re-vote on the measure. If a re-vote on the Validity subcriterion is desired, we will collect your votes on the call. If the measure passes the Validity subcriterion upon re-vote, we will ask you to discuss and vote on the remaining endorsement criteria (Feasibility; Usability and Use) and then vote on an overall recommendation for or against continued endorsement (see Appendix B).

> **Action Item:** After review and discussion of the comments on this measure and the additional information provided by the developer, does the Committee wish to re-vote on the Validity subcriterion for the measure?

> **Action Item:** If the Committee agrees that the measure passes the Validity subcriterion upon re-vote, it will vote on Feasibility, Usability and Use, and on overall suitability for endorsement.

> **Action Item:** If this measure is recommended for endorsement, discuss related and competing pain measures (see Appendix C).

#0211: Proportion of patients who died from cancer with more than one emergency department visit in the last 30 days of life

During the in-person meeting, the Committee expressed concerned about the lack of risk-adjustment for the measure, noting that appropriateness of ED admission may vary by factors that include patient and family characteristics, geographic region, urban versus rural environment, and availability of homecare resources. In particular, Committee members highlighted a potential unintended consequence of limiting access to care for patients in rural areas, where admission to the ED may be the only care option during an urgent situation.

Citing concerns related to the lack of risk-adjustment, the Committee agreed that the measure did not meet the Validity subcriterion as currently constructed, and instead opted to defer its endorsement decision, pending additional analysis regarding risk-adjustment.

Although initially agreeing with this stipulation, in subsequent communication with NQF, the developers withdrew this measure from consideration, stating that they would not be able to explore risk-adjustment of the measure at this time. Endorsement will be removed from this measure.

NQF received 4 comments on this measure. One commenter supported the measure without risk-adjustment, but the others supported the decision of the Committee. Two of these commenters encouraged the developers to re-submit the measure for endorsement once risk-adjustment has been addressed.

> **Action Item:** None.

Consensus Not Reached

#2651: CAHPS® Hospice Survey (experience with care): Treating family member with respect

A total of eight PRO-PMs are included under measure #2651. These measures are calculated from data obtained through the Hospice CAHPS® survey, which assesses patient and family caregiver experiences of hospice care.

The Committee had concerns with reliability scores for two of the PRO-PMs. Measure score reliability was calculated using 1) intra-class correlations (ICCs) computed from the case mix-adjusted 0-100 top-box scores and 2) estimating reliability via the Spearman-Brown prophecy formula assuming 200 surveys were completed in each agency. Across the 8 PRO-PMs, ICC values ranged from 0.008 to 0.017, and the estimated reliability from the Spearman-Brown prophecy formula ranged from 0.61 to 0.78. However, because the estimated reliability estimates were relatively lower for "*Treating family member with respect*" (ICC=0.008, reliability=0.61) and "*Getting help for symptoms*" (ICC=0.008, reliability=0.62), the Committee asked to vote on those two measures separately. The Committee did not reach consensus on the Reliability subcriterion for "*Treating family member with respect.*"

NQF received 3 post-evaluation comments regarding the 8 PRO-PMs under NQF #2651. All three commenters supported endorsement of these measures. NQF staff also presented these measures to our Person- and Family-Centered Care (PFCC) Standing Committee for their feedback, as this Committee has extensive experience in evaluating PRO-PMs from CAHPS surveys and other PRO-PM/instrument-based measures. One of the PFCC Committee members expressed concern with the low ICC values for all of the measures.

Additional Information Provided by the Developer: The developer updated the reliability estimates for all 8 PRO-PMs using data from April-September, 2015 (see Appendix D and table below; note that the reliability estimates provided in the original measure submission reflected data from April-June, 2015). The reliability estimate for "*Treating family member with respect*" increased from 0.61 to 0.68 with the additional three months of data.

	Initial submission: April - June, 2015		**Updated: April - September, 2015**	
Measures	**ICC**	**Estimated reliability for n=200**	**ICC**	**Estimated reliability for n=200**
Hospice Team Communication	0.013	0.72	0.018	0.78
Getting Timely Care	0.012	0.71	0.016	0.76
Treating Family Member with Respect	**0.008**	**0.61**	**0.011**	**0.68**
Getting Emotional and Religious Support	0.011	0.7	0.011	0.69
Getting Help for Symptoms	0.008	0.62	0.01	0.66
Getting Hospice Care Training	0.017	0.78	0.02	0.81
Rating of Hospice	0.011	0.68	0.015	0.75
Willingness to Recommend	0.017	0.78	0.021	0.81

Regarding the low ICC values, the developer cited Lyratzopoulos et al. (2011), who suggested benchmarks such that ICCs less than 0.01 are labeled "Low" and ICCs greater than 0.10 are labeled "High." Lyratzopoulos, et al. also states that the ICC can be interpreted as the reliability of the quality measure with a sample size = 1 respondent per hospice. The developers note that the ICC alone is not sufficient to determine reliability, since CAHPS Hospice Survey measures

would not be applicable when there is only 1 respondent for a hospice. They use the Spearman-Brown prophecy formula to estimate the reliability assuming 200 respondents per hospice.

Action Item: The Committee will re-vote on the Reliability subcriterion for "*Treating family member with respect*", and, if the measure passes this subcriterion, it will vote on overall suitability for endorsement.

#1639: Hospice and Palliative Care -- Dyspnea Screening

Data presented by the developer from the FY15 Hospice Item Set (HIS) indicate an average hospice facility-level performance rate of 97.3%. Additional data presented by the developer indicate slight, yet statistically significant, disparities in care between certain racial, socioeconomic, and geographic subgroups in the hospice setting. Developers did not provide clinician-level performance data for palliative care in the hospital setting. The Committee did not reach consensus on whether the measure results demonstrate opportunity for improvement, noting the high performance rate for the hospice setting but lack of information about opportunity for improvement for the clinician level of analysis in the hospital setting.

NQF received 5 post-evaluation comments on the measure, four of which were supportive of continued endorsement. One commenter expressed concern about inclusion of short-stay patients, recommending stratification of results for patients with length of stay <7 days and an exclusion for those patients who are imminently dying.

Developer response regarding concerns with specifications:

Thank you for your comment on the removal of the 7 day length of stay (LOS) exclusion. Under contract to the Centers for Medicare and Medicaid Services (CMS), the Research Triangle Institute (RTI) performed descriptive analyses on 5 quality measures (QMs) for which the University of North Carolina- Chapel Hill is the Measure Steward to examine the implications of the LOS exclusion on hospices' denominator size and QM scores. These analyses were conducted on HIS-Admission and –Discharge records for stays in October 1, 2014- September 30, 2015. Analyses encompassed 3,922 hospice organizations and approximately 1,218,786 patient stays.

This analysis showed that applying or removing the LOS exclusion generally had little impact on the distribution of hospices' QM scores. Since applying or removing the LOS exclusion does not affect scores, this means that separate analyses for patients enrolled less or more than 7 days will not yield additional quality of care information.

We agree that screening for pain, assessing pain, screening for dyspnea, treating dyspnea, and asking about patient treatment preferences may require somewhat different clinical skills for patients who are imminently dying. For example, a hospice nurse may need to use non-verbal symptom scores or ask a family surrogate about treatment preferences. However, these skills are well within the scope of hospice providers' practice, and do not preclude meeting quality measures for patients near death. In addition, the original rationale for the 7 day LOS exclusion -- allowing time for hospice providers to complete this care process -- does not appear to be necessary. Analyses show that a large portion of the care processes assessed by these QMs were performed on day 1 of admission to hospice, demonstrating a normative standard of care includes prompt attention to symptom distress.

Additional Information provided by the Developer: Preliminary data for the measure in the hospital-based palliative care setting indicate that 81.8% of patients were screened for dyspnea (see Appendix E). This result is based on data from 895 patients who were admitted to an acute care hospital for at least 1 day from January 2014 to June 2015.

Action Item: The Committee will re-vote on the Opportunity for Improvement subcriterion, and, if the measure passes this subcriterion, it will vote on overall suitability for endorsement.

Reconsideration Request

#1626: Patients Admitted to ICU who Have Care Preferences Documented

The developer has requested that this measure be reconsidered by the Committee.

During the in-person meeting, the Committee did not reach consensus on the reliability of this measure and it did not pass the measure on validity. Regarding reliability, the Committee was concerned about the ability to consistently apply the numerator specifications, particularly if there is already an advance directive available. Regarding validity, the Committee did not accept the face validity testing of the measure, noting that one of the face validity assessments was specific to cancer patients only, that none of the face validity assessments were specific to ICU patients, and that this measure was not assessed specifically but was instead discussed more generally (see Appendix F).

NQF received 6 comments on this measure. One of the commenters supported the Committee's decision not to recommend the measure for continued endorsement, while two did not support the decision, and one requested that the Committee reconsider the measure after obtaining additional information and clarification regarding the measure specification. One commenter noted the importance of emergency, critical, and advance care plans and provided specific suggestions on ensuring these are available to healthcare providers. Finally, the last commenter—the developer of the measure—formally requested a reconsideration of the measure due to inappropriate application of the evaluation criteria.

Developer Comment: We are requesting that the Committee reconsider the measure that was considered for maintenance. First, regarding the concerns about face validity, while it is true that one of the panels was cancer only, each measure was reviewed individually at each expert panel for face validity and only those that met the criteria as explained in provided documents were considered to be valid. This measure was considered to have face validity by these expert panels. I think the panel may have been confused by a paragraph in the measure testing document that talks about a higher level of evidence for validity (the process-outcome link). For this higher level of validity, there is only data in aggregate. Second, regarding the reliability of the measure we provided kappa statistics in the reliability section that showed high inter-rater reliability showing that we were able to reliably collect this data using the proposed specifications.

Action Item: Based on comments received and the information provided by the developer, would the Committee like to reconsider this measure?

Action Item: If, upon re-vote, the Committee agrees that the measure passes the Reliability and Validity subcriteria, it will vote on Feasibility, Usability and Use, and on overall suitability for endorsement (see Appendix G).

Action Item: If this measure is recommended for endorsement, discuss related and competing patient preference measures (see Appendix H).

Other Comments and their Disposition

Three major themes were identified in the remaining post-evaluation comments, as follows:

1. Support for the recommended measures
2. Consideration of Patient Choice in Measurement
3. Gaps in Palliative and End-of-Life Care Performance Measurement

Theme 1 – Support for the recommended measures

Overall, commenters supported the recommended measures. Aside from the comments already noted above, 50 additional comments expressed support for (but no additional questions or concerns regarding) the Committee's decisions to recommend 19 of the evaluated measures for endorsement. Several of the supportive comments also offered ideas for additional measure development.

Action Item: None.

Theme 2 – Consideration of Patient Choice in Measurement

Two commenters noted the importance of patient choice in measurement. One (ID #6045) noted that some patients will want to continue aggressive care near the end of life, stating that physicians should not be penalized when patients make this choice. This commenter emphasized the need for informed patient and family choice and acknowledged that 100% (or 0%) performance is not the goal for "aggressive care" measures (e.g., emergency department use). Another commenter (ID #6119) noted that patient preferences for end-of-life care often change over time, thus highlighting the importance of affording frequent opportunities for modifying their formal care and treatment preferences.

Proposed Committee Response (ID #6045): Thank you for your comment. The Committee agrees that patients and their families should be encouraged and assisted in making informed decisions regarding end-of-life care and that palliative and end-of-life care measures and related measurement programs should take patient choice into account.

Proposed Committee Response (ID #6119): Thank you for your comment. The Committee agrees that patients should be given ample opportunity to modify their care and treatment preferences over time.

Theme 3 – Gaps in Palliative and End-of-Life Care Performance Measurement

Many of the submitted comments confirmed the gaps in measurement identified in the draft report and/or identified additional gaps, including:

- Measures addressing legacy support (e.g., evidence-based dignity therapy)
- Measures focusing on creativity (e.g., art or music therapy)
- Measures that address the NQS priorities of Effective Prevention and Treatment of Illness, Best Practices for Healthy Living, and Affordable Care
- Measures that consider hospice stays of less than 30 days
- Measures that consider social determinants of care (e.g., socioeconomic, educational, spiritual, cultural, etc.), particularly as related to advance care planning
- Measures related to bereavement care
- Measures for patients with chronic or life-limiting conditions (i.e., the patient population appropriate for palliative care), including those in settings other than hospice and hospital-based palliative care (e.g., home, nursing homes, ambulatory care, etc.)
- Measure of outcomes, particularly patient-reported outcomes
- Measures of alignment between care that is provided and patients' preferences, goals, values, and wishes
- Measures related to advance care planning
- Measures that assess care longitudinally and across care settings

Proposed Committee Response: Thank you for your comment. The Committee agrees with your suggestions for future measure development and the report will be updated accordingly.

Discussion of the Measurement Framework

During the in-person meeting, the Committee offered some initial suggestions for expanding the draft measurement framework for palliative and end-of-life care (e.g., specifically including concepts related to cost, decision-making, and safety). Questions for the Committee to consider include:

- Do the "rings" of the framework make sense?
- Are the rings in the correct order?
- Should we consider different domains of care?
- Should we represent in some way the core components of quality end-of-life care, as defined by the Institute of Medicine in *Dying in America*?
- How should concepts of cost, decision-making, and safety be included in the framework?
- Should we make other modifications to the framework?

Appendix A – Response letter and table from NHPCO regarding risk-adjustment for measure #0209

National Hospice and Palliative Care Organization

MEMORANDUM

TO: Standing Committee, National Quality Forum Palliative and End-of-Life Endorsement Maintenance Project

FROM: National Hospice and Palliative Care Organization

DATE: July 22, 2016

Subject: Request to reconsider measure NQF #0209 Comfortable Dying: Pain Brought to a Comfortable Level Within 48 Hours of Initial Assessment

The National Hospice and Palliative Care Organization (NHPCO) requests that the Committee reconsider NQF #0209 for maintenance endorsement.

When it met on May 10, the Committee expressed concern about the lack of risk adjustment for this measure. NHPCO offers the following additional information to address that concern.

BACKGROUND

NQF #0209 is a patient reported outcome measure. It is the 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.

Data collection for this measure is by an initial question related to discomfort due to pain and a follow-up question related to comfort achieved. The patient is asked for a "yes" or "no" response at the initial pain assessment to the question, "Are you uncomfortable because of pain?" Two to three days later the patient is asked "Was your pain brought to a comfortable level within 2 days?"

As a patient reported outcome measure (PROM) #0209 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.

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. Change in pain intensity scores alone does not demonstrate whether comfort was achieved and does not incorporate the concept of patient self-determination. Pain is an individual experience with an individual response and the evaluation of the outcome of whether pain has been managed rests with the individual. In addition, the initial question functions as a screener for discomfort. Because the

question is not tied to a specific retain scale it identifies those patients who require intervention and has the advantage of allowing the clinician to use the most appropriate means of pain assessment for each individual patient.

Risk Adjustment

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.

Analysis

However, because of the Committee's interest in analysis of risk adjustment for the measure, we conducted an analysis of potential risk factors at the facility level. Scores used in this analysis combined data from 2012 and 2013, restricted to hospices with denominators of at least 50. 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. The specific results of these analyses are presented in attachment to this memo.

Summary

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.

Because of the supplementary information on risk analysis provided and the importance of this measure to the field of palliative and hospice care, NHPCO requests that the Committee reconsider its vote for NQF #0209.

National Hospice and Palliative Care Organization

NQF Measure #0209

SUPPLEMENTARY RISK ANALYSIS

Risk Factor	Median Measure Scores For Comparison Groups (N # of facilities in each group)	Statistical Test	Result
Ownership	Freestanding 62% (N=64) Hospital-based 62% (N=21)	Wilcoxon rank-sum (p=0.23).	No statistically significant difference
Geographic Location	Midwest 62% (N=27) Northeast 55% (N=12) South 62% (N=31) West 63% (N=24)	Kruskal-Wallis (p=0.08)	No statistically significant difference
Service Area	Primarily Urban 61% (N=20) Mixed 62% (N=51) Primarily Rural 62% (N=10)	Kruskal-Wallis (p=0.40)	No statistically significant difference
Race	Facilities above median for proportion of African American patients, 63% (N=24) Facilities below median for proportion of African American patients, 61% (N=23)	Wilcoxon rank-sum test (p=0.60).	No statistically significant difference
Ethnicity	Facilities with >5% proportion of Hispanic patients, 67% (N=10) Facilities with <1% proportion of Hispanic patients, 62% (N=13)	Wilcoxon rank-sum test (p=0.43)	No statistically significant difference
Patient Age	Facilities above median for proportion of patients ≥65 years of age, 61% (N=20) Facilities below median for proportion of patients of ≥65 years of age, 56% (N=21)	Wilcoxon rank-sum (p =0.34)	No statistically significant difference
Gender	Facilities above median for proportion of male patients, 60% (N=25) Facilities below median for proportion of male patients, 61% (N=24)	Wilcoxon rank-sum (p=0.36)	No statistically significant difference
Referral Source (proxy for acuity)	Facilities above median for proportion of patients from hospital referrals 58% (N=20) Facilities below median for proportion of patients from hospital referrals 61% (N=19)	Wilcoxon rank sum (p=0.66)	No statistically significant difference

Additional risk analysis was conducted at the facility level. Scores used in this analysis combined data from 2012 and 2013, restricted to hospices with denominators of at least 50. Because the distribution of scores showed departures from normality, nonparametric statistical tests were used (Kruskal-Wallis, Wilcoxon Rank-Sum).

Appendix B – Measure worksheet: #0209

NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 0209
Measure Title: Comfortable Dying: Pain Brought to a Comfortable Level Within 48 Hours of Initial Assessment
Measure Steward: National Hospice and Palliative Care Organization
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.
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.

Numerator Statement: Patients whose pain was brought to a comfortable level (as defined by patient) within 48 hours of initial assessment.
Denominator Statement: Patients who replied "yes" when asked if they were uncomfortable because of pain at the initial assessment.
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

Measure Type: PRO
Data Source: Patient Reported Data/Survey
Level of Analysis: Facility, Population : National

IF Endorsement Maintenance – Original Endorsement Date: Aug 10, 2009 **Most Recent Endorsement Date:** Feb 14, 2012

Maintenance of Endorsement -- Preliminary Analysis

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 outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale. In addition to the evidence required for any outcome. The evidence for a Patient-reported outcome-based performance measures (PRO-PM) should demonstrate that the target population values the measured PRO and finds it meaningful.

Evidence Summary

- The developer provides 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).

Changes to evidence from last review

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

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

Updates: The developer 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 states *"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."*

Exception to evidence – N/A

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

Questions for the Committee:

- *Does the Committee agree that hospice patients value queries about pain and pain management?*
- *Does the evidence support limiting the measure to those ages 18 and older?*
- *Is there evidence that hospice patients value this type of assessment more or less than other types of patients?*
- *The developer attests the underlying evidence for the measure has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat vote on Evidence?*

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.

- 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.

Disparities

- Disparities data were provided, although these may be patient-level, rather than facility-level, statistics.
  - An analysis of 2,329 patients in 2014 indicated that there were no statistically significant differences in the measure results by age group, sex, or race.
  - An earlier analysis of measure results according to diagnosis—cancer vs. non-cancer—indicated fairly similar results (81% vs 84.8%); however, the developer did not indicate whether or not those differences were statistically significant.

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.? Why the drop in the number of reporting facilities?*
- *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 at the facility level?*

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

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a.

*I agree that hospice patients--and their caregivers--value queries about pain and that it cannot be managed without assessment. The relationship between the measured outcome and the process is demonstrated by the rational and diagram of the pain assessment process.
* Yes, applies directly and relates to desired outcomes.
* There is evidence to support the PRO in that proper pain assessment process will support the best application of treatment to control pain. Patient report of pain relief after 48 hours provides subjective feedback of desired outcome of pain relief establishes value for the measure.
* Measure outcome applies directly to care process.

One question on the denominator/exclusions (patients that cannot self report pain)--given that there are tools for assessing pain in the cognitively impaired (PAIN AD and Abbey Pain Tool) am concerned that unless guidance is specified they would excluded. This indeed could still be considered self report.
* PASS. concerns are pts who may be sedated inappropriately and will not participate in 48 evaluation. % of non responders at 48 hs would help
* This measure determines the percentage of patients who report being uncomfortable because of pain at the initial assessment (entry into hospice) and at the follow-up assessment, report pain was brought to a comfortable level within 48 hours.

*The measure is a patient, self-report measure. One question is asked at initial assessment, "Are you uncomfortable because of pain?" and the second question is asked within 48 hours, "Was your pain brought to a comfortable level within 48 hours of the start of hospice care?".

It is a patient reported outcome (PRO).

There is a relationship between the measured outcome and at least one healthcare action. If the patient reports pain then providers need to intervene to manage pain. The evaluation of the pain management strategy then occurs within 48 hours.

If the patient has pain that has not changed within 48 hours the next steps are not clear. It does not look like the follow-up measure is used again.

The follow-up period is within 48 hours. If a patient has moderate-severe pain a follow-up period of 48 hours is long. There is no discussion of this concern.

It is unclear why the measure cannot be used for patients under the age of 18.

This measure should not be limited to hospice patients.

1b.

*Performance gap data was provided and it demonstrates the potential for improving the assessment of pain on a national level. Disparities data were provided but there were no statistically significant demographic differences nor diagnosis.
* Performance data for the measure was provided for 2102 to 2015. Presumably the initial testing of 1409 patients may not have been sufficient to establish statistical significance in PRO report of comfort. Performance scores remain consistent over time regardless of sample size. Personally I am not able to make sense of the data provided to assess whether it demonstrated a gap in care. Subgroup data for 2014 was reported indicating no significant disparity in age, gender, race, ethnicity in addition to Cancer v non-cancer pain identified.
* Data provided which demonstrates a gap warranting national performance measure. Disparities data did not indicate a gap by gender, age, or cancer vs non-cancer diagnosis.

* Yes. Yes. Age/Gender/Race of hospice population. No other characteristics/subgroups provided.
* High. Agree that drop of number of participating hospices and # pts concerning.
*No statistically significant differences in the measure results have been reported based on age group, gender, and ethnicity.

There is a gap in care that warrants a national performance measure. Data were provided that reported pain at initial hospice assessment and for a sub-set of patients continued pain at the follow-up assessment. This is a serious problem.

In February 2013, the Measure Applications Partnership supported the measure for inclusion in PQRS, finding that the measure filled an identified gap. Public comments from the Center to Advance Palliative Care and the National Coalition for Hospice and Palliative Care supported the measure.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications
Maintenance measures **– no change in emphasis – specifications should be evaluated the same as with new measures**

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

Data source(s): Self-reports of patients admitted to hospice
Specifications:

- The measure is specified at the facility level of analysis, for use in the hospice setting.
- The numerator consists of patients reporting their pain was brought to a comfortable level within 48 hours of assessment.
- The denominator consists of patients who replied "yes" when asked if they were uncomfortable because of pain at the initial assessment.
- Exclusions to the denominator include:
    - 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
- The measure score is a rate/proportion, and higher scores are better.
- A calculation algorithm is provided.
- The measure is not stratified or risk-adjusted.

Questions for the Committee :

- *Is the logic or calculation algorithm clear?*
- *The developer notes difficulties in implementation when the measure was required by CMS for the first year of the Hospice Quality Reporting Program. Could better specificity have improved implementation?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

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. NOTE: Because this is Patient-reported outcome-based performance measure (PRO-PM), reliability testing at the performance score level is required.

Summary of prior reliability testing:

- Performance measure score reliability was initially tested using 2009-2010 data from NHPCO's Patient Outcomes Survey.

Updates to Testing

- Additional performance measure score reliability testing was completed using data from the 2013 and 2014 NHPCO's Patient Outcomes Survey.

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing:

- Initial testing
  - Data from 58 hospice agencies and nearly 38,000 patients were analyzed.
  - The developer utilized the Intraclass Correlation Coefficient (ICC) to examine the agency-level between-versus-within variance of the measure numerator using 2 years of data. Use of the ICC is an appropriate method of testing reliability. Note that because only those who reported pain at initial assessment were asked whether their pain was brought to a comfortable level within 48 hours, NQF will consider this analysis of the numerator data element as essentially an analysis of the measure score.
  - The developer also conducted an analysis of variance to assess whether agency means for the measure numerator and denominator varied across quarters. However, NQF does not consider analysis of data across time to be an appropriate method of testing the reliability of the data elements.
- Updated testing
  - The developer reports using a binominal distribution model and a random selection of 50 patients to develop a guideline for the random variability of the measure. Using 2013 data of more than 16,000 patients, this guideline suggests that a hospice with 50 patients in the measure denominator would have a score of 58%, with an 80% chance of a score between 48%-68% and <1% chance of a score of <38% or >78%. ***It is unclear if or how this analysis demonstrates score-level reliability***.

- The developer also supplied data regarding changes in measure scores over time for 22 hospices with at least 50 patients in 2013-2014. However, NQF does not consider analysis of data across time to be an appropriate method of testing the reliability of the measure score.

Results of reliability testing

- **Initial testing**
  - Testing results indicated that the ICC for the between and within hospice variation was 0.71 (95% CI 0.63-0.77). ICC values can range between 0 and 1.0. ICC value of 0.71 indicates that 71% of the variance in scores are due to differences between hospice agencies. A value of 0.7 is often regarded as a minimum acceptable reliability value.

Guidance from the Reliability Algorithm
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)

Questions for the Committee:

- *Is the test sample from the initial testing adequate to generalize for widespread implementation?*
- *Does the updated testing demonstrate score-level reliability? If so, how?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Preliminary rating for reliability (based on initial testing results only):
☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence?*
- *Do you agree that the two ways the developers asked the question about "comfortable level" and "acceptable level" (see validity testing, below) are equally consistent with the evidence?*

2b2. Validity testing

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. NOTE: Because this is Patient-reported outcome-based performance measure (PRO-PM), validity testing at the performance score level is required.

For maintenance measures, summarize the validity testing from the prior review:

- Developers compared response rates from two different wordings ("comfortable" level and "acceptable" level) for the follow-up question related to pain management. Again, because only those who reported pain at initial assessment were asked the follow-up question, NQF will consider this analysis of the numerator data element as essentially an analysis of the measure score.

Describe any updates to validity testing:

- Information on updated testing was not provided, although the developers provided an additional statistic to further explain the results of the previous testing.

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

☐ **Face validity only**
☒ **Empirical validity testing of the measure score**

Validity testing method:

- **Initial Testing**
  - Testing data included 212 of 686 patients from 9 hospice agencies who reported pain on initial assessment. These patients 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. The developer notes that these two forms of the follow-up question were considered equivalent by the expert panel for the Comfortable Dying Measure.
  - Additional information from the developer will be needed to understand how this method validates the measure results.

Validity testing results:

- **Initial testing**
  - 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.
- **Updated testing**

Developers updated the testing form to report a Cohen's kappa of 0.91. This statistic appears to have been calculated using data and results of the initial validity testing, though this is not entirely clear. The kappa statistic represents the proportion of agreement that is not

explained by chance alone. According to the Landis and Koch classification, a kappa value of 0.91 indicates almost perfect agreement between the two sets of responses.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *How does this analysis validate the measure score?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developer did not provide data on exclusions (e.g., number excluded, number of exclusions by each exclusion criterion). The developer did note, however, that interpreter services can be used, although proxy answers are not acceptable.

Questions for the Committee:

- *Are any patients or patient groups inappropriately excluded from the measure? Note the CMS Rule (see Usability and Use section below), which questioned the number of patients excluded by the measure.*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

- The developers state there was not a statistically significant effect of age (>65 years old vs ≥65) or gender on the measure score in the 383 of 2,329 sampled hospice patients who qualified for the measure denominator. However, this analysis does not speak to whether there are differences in age or gender (or other characteristics) of patients between hospice agencies.

Questions for the Committee:

- *Do the results provided demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across hospice facilities?*
- *Is there any evidence that contradicts the developer's rationale and analysis underlying the decision not to risk-adjust this measure?*
- *Are there other factors besides age and gender that might have an effect on the measures score and should be considered for risk-adjustment?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- Using 2013 data submitted to NHPCO, the developers compared individual hospice agency scores to the national average score for 97 hospices with more than 50 patients in their denominator.
- Of the 97 agencies, 16 had scores that were significantly different from the national average at the $p<0.05$ level (21 were statistically significantly different if using the $p<0.1$ level).

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- There is only one set of specifications for the this measure. Comparability of data sources/methods is not applicable.

2b7. Missing Data

- The developers state that the samples used for testing had very little missing data and the missingness was not at a level to bias the measure. However, the developer did not provide data on the frequency of missing data.

Guidance from the Validity Algorithm
Specifications consistent with evidence (Box 1) →Threats to validity somewhat assessed, although questions remain, particularly around exclusions and risk-adjustment (Box 2)

Preliminary rating for validity: ☐ **High** ☐ **Moderate** ☐ **Low** ☒ **Insufficient**
Rationale: Additional information regarding frequency of exclusions and agency-level differences in potential case-mix adjusters needed. Additional score-level validity testing may also be needed.

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a, 2b.
Reliability – Specifications
* Moderate- a more representative sample would be more useful
* No concerns about reliability specifications
* Data elements and processes are clearly defined. I believe this measure can easily be consistently implemented across all patient care areas.
* Reliability testing was conducted using a measure score and demonstrate moderate certainty that the measure results are reliable.
* When is pain assessed? We have the T2 (within 48 hours) but could we fail to identify pain? How often should pain be assessed?
* The measure is specified at the facility level of analysis, for use in the hospice setting.

The numerator consists of patients reporting their pain was brought to a comfortable level within 48 hours of assessment.

The denominator consists of patients who replied "yes" when asked if they were uncomfortable because of pain at the initial assessment.

The measure score is a rate/proportion, and higher scores are better.

A calculation algorithm is provided.

The measure is not stratified or risk-adjusted.

The developer notes difficulties in implementation when the measure was required by CMS for the first year of the Hospice Quality Reporting Program.

It is unclear if the measure can be consistently implemented.

Validity – Specifications
* Pass. Personalized pain goal is quite valid
* I did not see any inconsistencies between specifications and evidence, nor where there inconsistencies with the target population values.
* No concerns about validity specifications
* Am concerned that the evidence indicates that the cognitively impaired exhibit definable signs of pain (several pain tools already cited). Given the growth of the oldest old (with 50% or great incidence of dementia) and additionally the acutely ill with delirium, should the measure include this population in the denominator and suggest validated tools?
* Empirical assessment of the potential difference between "comfortable" and "acceptable" was an important analysis of this measure score.
* This is a patient reported outcome based performance measure, thus validity testing a the performance score level is required.

Developers compared response rates from two different wordings (comfortable level and acceptable level). Because only those who reported pain at initial assessment were asked the follow-up question, NQF will consider this analysis of the numerator data element as an analysis of the measure score.

It is unclear if "acceptable" and "comfortable" are the same.

Reliability – Testing
* Pass with the comments made before
* Reliability testing was adequate.
    * Though a reliability score of 0.70 is minimally acceptable, I believe, given the one evaluative question asked of patients of their pain, that measure score of 0.71 (95% CI 0.63-0.77) where ICC values can range between 0 and 1.0, the measure can be generalized for widespread implementation.
* Yes. Yes.
* Reliability testing was conducted at both the data element and score levels. Sufficient reliability is demonstrated so that differences in performance can be identified.
* Performance measure score reliability was initially tested using 2009-2010 data from NHPCO's Patient Outcomes Survey.

Additional performance measure score reliability testing was completed using 2013-2014 data from NHPCO's Patient Outcomes Survey.

Initial testing included data from 58 hospice agencies and almost 38,000 patients. Intra-class Correlation Coefficient (ICC) was used to examine agency-level between vs. within variance of the measure numerator using 2 years of data. Because only those who reported pain at initial assessment were asked whether their pain was brought to a comfortable level within 48 hours, NQF considers this analysis of the numerator data element as an analysis of the measure score. The ICC for the between and within hospice variation was 0.71 (95% CI 0.63-0.77). ICC value of 0.71 indicates that 71% of the variance in scores are due to differences between hospice agencies. A value of .7 is regarded as a minimum acceptable reliability value. The developer conducted an analysis of variance to assess whether agency means for the measure numerator and denominator varied across quarters. NQF does not consider analysis of the data across time to be an appropriate method of testing the reliability of the data elements.

Updated testing: The developer used a binominal distribution mode and random selection of 50 patients to develop a guideline for the random variability of the measure. Using 2013 data of more than 16,000 patients, this guideline suggests that a hospice with 50 patients in the measure denominator would have a score of 58%, with an 80% chance of a score between 48-68% and <1% chance of a score of <38% or 78%. It is unclear if or how this analysis demonstrates score-level reliability. The developer also supplied data regarding changes in measure scores over time for 22 hospices with at least 50 patients in 2013-2014. NQF does not consider analysis of data across time to be an appropriate method of testing the reliability of the measure score.

Validity Testing
* Pass
* Validity testing was strong with a kappa > .90.
* I am not sure if 212 of 686 patients from 9 hospice agencies is adequate to generalize for widespread implementation. However, the sample population did provide a kappa value of 0.91 indicating acceptable validity regardless of use of the word comfortable or acceptable - which lends me to believe this is an indicator of a quality measure. One might argue that a patient may respond that their pain is at an acceptable level, but they are still not comfortable. PRO-PM was evaluated at the score level.
* Yes. As the measure developer stated, pain is an individual experience (as would be comfort and the lessening of pain). This measure relies on the person's self reported assessment and reassessment of pain. The absolute score or scale is not the defining feature but rather the presence and addressing of pain as captured by this PRO.
* The sample size does not seem adequate to generalize for widespread implementation nationally.
* Initial testing: Testing data included 212 of 686 patients from 9 hospice agencies who reported pain on initial assessment. These patients were asked if their pain was brought to a comfortable level within 48 hour and then they were asked if their pain was brought to an acceptable level with 48 hours. The developers noted that an expert panel of reviewers considered the terms equivalent. It is unclear how this method validates the results of the measure. Sixty percent of patients responded their pain was brought to an comfortable level within 48 hours and 64% responded their pain was brought to an acceptable level within 48 hours. The developers concluded that 96% of patients provided the same answer to the two wordings.

Updated testing: The developers updated the testing form and report a Cohen's kappa of 0.91. This statistic appears to have been calculated using data and results of the initial validity testing, yet this is unclear. The kappa statistic represents the proportion of agreement that is not explained by chance alone and a kappa value of 0.91 indicates almost perfect agreement between the two sets of responses.

It is unclear if the measure is generalizable for widespread implementation.

Threats to Validity

* Missing data a problem and caregiver evaluation or reporting of % missing data possible ways to address
* No issues noted.
* Per the current measure set up, missing data does not appear to constitute a threat to validity. The exclusions to this measure are reasonable given the typical hospice population and expectation that patient's self report for this measure. With the growing dementia population it would appear that cognitive impairment would lend to not so much to validity risk, but the ability to generalize to widespread hospice patient population. I would not expect any meaningful difference between hospice agencies would have an impact on quality. That may only reflect the difference in number of patients reported by each agency (min 50 vs 300). I don't see any evidence that contradicts the developer's rationale and analysis underlying the decision not to risk-adjust. I don't believe missing data would constitute a threat - for example: If the measure question was asked on admission and never followed up on, then that denominator data would/should be removed from the equation. At that point the concern would be poor number reporting.
* Meaningful differences were identified for 16 of 97 hospices with mean scores > .05 from national average.
* Data on exclusions are not provided. Persons who cannot answer the question, those under age 18 were excluded and those who cannot understand the language but there is no indication of how many. It seems important to adjust risk for race,, age, gender and diagnoses to truly understand the assessment of pain. There is no comparability of data sources or methods.. Missing data is not provided.
* The developers did not provide data on exclusions. Information was provided that interpreter services could be used to help patient communication as needed.

The developers state there was not a statistically significant effect of age or gender on the measure score for the hospice patients who qualified for the measure denominator. This analysis does not address whether there are differences in age or gender of patients between hospice agencies.

It is unclear if there are additional factors that might have an effect on the measure.

The developers compared individual hospice agency scores to the national average score for 97 hospices with more than 50 patients in their denominator. Sixteen of the 97 hospices had scores that were significantly different from the national average.

The developers state that the samples used for testing have very little missing data and the missingness was not at a level to bias the measure but they did not provide data on the frequency of missing data.

Additional data are needed to address the above issues.

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 provides a Data Submission Worksheet for hospice agency use, and also offers guidance for calculating the measure, without requiring licensing or fees.
- 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)?*
- *Do any Committee members have experience implementing the measure? Can they speak to potential difficulties?*
- *Is the data collection strategy ready to be put into operational use? Does the decreasing number of hospices reporting on the measure speak to its feasibility?*

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

Committee pre-evaluation comments

Criteria 3: Feasibility

*Data are collected by use of a Data Submission Worksheet not electronically. Variation in record systems understood, it seems that this hampers data collection. Electronic submission seems to be important. I question the reason for the decrease in hospices that utilize this measure.

*Required data elements (patient response to the initial question and the follow-up question) are not available in all settings in electronic systems (EHR).

Interesting, there has been a decrease in use of the measure. It is unclear why this has occurred.

*completely feasible
* Not enough information given about feasibility for individual hospices. Not clear if captured in electronic medical record for each patient.
* It was noted that hospices reported difficulty in implementing the measure and that the measure may have had more success had it been introduced later in the HQRP process. Hospices may not be using an EMR.
* In 2016, I can't imagine any hospice regardless of size without an EMR or other patient data collection system in place. Many CMS reporting requirements necessitate a hospice having readily retrievable patent care and medication use data. That said, it would be very easy to identify an electronic field to score the elements of this measure in an EMR or other like database. Additionally, we are only talking about 2 data points - with one question each. I believe the decreasing number of hospices reporting speaks to the individual hospices not making this measure a part of it's patient admission and reporting process.

Criterion 4: Usability and Use
Maintenance measures **– increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences**

4. Usability and 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.

Current uses of the measure [from OPUS]
Publicly reported? ☐ **Yes** ☒ **No**

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

Accountability program details:

- The developer indicates that this measure is included in the PQRS program. However, PQRS is a clinician program, so this PQRS program is using a facility-level measure to assess clinician performance.
- NHPCO provides data collection and comparative reporting (i.e., benchmarking) for those hospices that voluntarily submit data to NHPCO. The developers note that in 2014, 156 hospices provided measure data for 20,548 patients (although this does not match the data reported in section 1b).

Improvement results:

- The developer provided facility-level performance data for 2012-2015 in section 1b. These results indicate little change over time in performance, along with a decrease in the number of facilities reporting. The developer also provided patient-level national averages by quarter for 2013-2013.

Unexpected findings (positive or negative) during implementation: No unexpected findings were reported by the developer.

Potential harms: No unintended consequences were reported by the developer.

Feedback:

- CMS removed this measure from its Hospice Quality Reporting Program. From the Rule removing NQF#0209 from the HQRP: "There is a high rate of patient exclusion due to patient ineligibility for the measure and patients' denying pain at the initial assessment. This high rate of patient exclusion from the measure results in a small denominator and creates validity concerns. These concerns cannot be addressed by training or standardizing data collection."
- In February 2013, the Measure Applications Partnership supported the measure for inclusion in PQRS, finding that the measure filled an identified gap. Public comments from the Center to Advance Palliative Care and the National Coalition for Hospice and Palliative Care supported the measure.

***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: Relatively few hospices use the measure; moreover, results indicate stagnant performance over time rather than improvement over time.

Committee pre-evaluation comments

Criteria 4: Usability and Use

*The submission seems to demonstrate that decreasing participation, a high level of exclusion and little change over time in performance. Despite the compelling need to assess and intervene with hospice patients' pain, this seems to demonstrate continuing difficulty with implementing meaningful assessment strategies.

*The developers state the measure is included in the PQRS program, but this is a clinician program, so this PQRS program is using a facility level measure to assess clinician performance.

NHPCO provides data collection and comparative reporting for those hospices that voluntarily submit data to NHPCO. The developers note that in 2014, 156 hospices provided measure data for 20,548 patients. This does not match the data reported in section 1b.

CMS removed this measure from its Hospice Quality Reporting Program. CMS reports, "There is a high rate of patient exclusion due to patient ineligibility for the measure and patients' denying pain at the initial assessment. This high rate of patient exclusion from the measure results in a small denominator and creates validity concerns. These concerns cannot be addressed by training or standardizing data collection."

*low; use of this will mean hospices are measuring pain AND trying to control it fast. Both important quality measures of process and outcomes of hospice delivery
* Measure is good for the population to which it applies, but as Katherine Ast from AAHPM commented, it is a narrow population of patients to which this measure applies. It would be good if it applied more broadly to patients in other health care settings.
* It is included in PQRS which is a voluntary.

CMS removed the measure from HQRP because of numbers of patients denying pain at initial assessment. Small denominator raises validity concerns.

Again, given the cognitively impaired and the ability to use validated tools to identify their pain, would consider this an important aspect for the panel to discuss in review.
* To my knowledge, this measure is not publicly reported. Like anything in life, if there are no consequences to poor performance then, in this instance, a measure's use will wane. Many hospices utilize family surveys as feedback to their performance during care of their loved one - this feedback retrospective and often dealt with on a reactionary basis. Hospice results from this measure should be used to a. improve pain management within the hospice, and b. promote publicly through it's hospice's liaison how well they are able to manage a person's pain once admitted to the program. Consideration of this measure should be taken in to account by The Joint Commission and the Community Health Accreditation Partner (CHAP) within their survey process to establish this as a "performance" measure for those accredited agencies.

Criterion 5: Related and Competing Measures

Related 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]*
- 0676: Percentage of short-stay residents, of all ages, in a nursing facility, who have reported almost constant or frequent pain, and at least one episode of moderate to severe pain, or any severe or horrible pain, in the 5 days prior to the target assessment *[facility-level outcome measure in nursing home setting]*
- 0677: Percentage of short-stay residents, of all ages, in a nursing facility, who have reported almost constant or frequent pain, and at least one episode of moderate to severe pain, or any severe or horrible pain, in the 5 days prior to the target assessment *[facility-level outcome measure in nursing home 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.

Pre-meeting public and member comments

Name: Michele Galioto, RN, MSN
Organization: Oncology Nursing Society
Comment: Is 48-hour a standard benchmark? Might be important to reduce timeframe.

Name: Michele Galioto, RN, MSN
Organization: Oncology Nursing Society
Comment:
Overall, ONS recommends differentiating between palliative and end of life care in introductory information. ONS does not define palliative care as equal to end of life care. Hospice care is a form of palliative care but not inclusive of all palliative care. Palliative care should begin at the point of diagnosis or awareness of symptoms and continue throughout the trajectory of treatment through end of life care. See ONS position statement on palliative care for further detail.
ONS is also in favor of including recommendations for intervals of assessment as the current measures imply that one screening is sufficient. Screening at each patient encounter may be more appropriate.

Name: Katherine Ast, MSW, LCSW
Organization: American Academy of Hospice and Palliative Medicine
Comment:
Dear Ms. Johnson:

On behalf of the palliative care community, we thank the National Quality Forum for convening its Palliative and End-of-Life Care 2015-2016 Project and for the opportunity to provide preliminary feedback on the palliative and end-of-life care measures that will soon be evaluated by the project's Standing Committee.

The American Academy of Hospice and Palliative Medicine (AAHPM) is the professional organization for physicians specializing in hospice and palliative medicine, and our membership also includes nurses and other health and spiritual care providers committed to improving quality of life for seriously ill patients and their families. We support the pursuit of interdisciplinary, team-based palliative care and its emphasis on care coordination, pain and symptom management, shared decision making, and patient-centered goal-setting. The provision of palliative care has been shown to improve patient experience and satisfaction,i reduce caregiver burden,ii and increase survivaliii; it has also been shown to reduce needless hospital admissions and readmissions through effective care coordination and symptom managementiv; and through these gains in quality, it reduces costs.v

While we do not, at this time, view any of the measures under consideration as particularly controversial, their limited scope reflects the critical ongoing gaps related to palliative and end-of-life care measurement and highlights the unique challenges that have contributed to those gaps. For example, the current set of measures under consideration is largely limited to cancer or hospice settings. These measures employ a narrow denominator (e.g., *hospice* patients rather than *dying* patients). This is certainly a good start, but measuring only hospice patients in order to improve the quality of end-of-life care is like searching for a lost dollar bill only where the light is good. It will not move the needle to the extent that we need it to. The National Hospice and Palliative Care Organization (NHPCO) reports in its *2011 Facts and Figures* that only 42% of those who died in 2010 were enrolled in hospice. How do we measure the quality of end-of-life care for the majority of patients who die in hospitals, skilled nursing facilities, and homes without the benefit of hospice care? These are questions

we have not yet been able to answer. The fact that the current set of measures under consideration by the NQF only includes *one* new measure (i.e., the Hospice CAHPS) also illustrates that the standard default pathways for measure development, testing, and endorsement are not working for the patients, providers and researchers in our field. Patient and family preferences and experience of care are critical elements of quality palliative care, and ongoing funding, data analysis and personnel are required to develop these kinds of measures and keep them endorsed and in use. For example, the current NQF requirement for measure developers to test survey instrument data elements in addition to the measures themselves (double testing) poses a barrier to advancing the field. While the process of submitting the PEACE measures from the University of North Carolina has gone well because of RTI's support and the national data coming from the Hospice Item Set (HIS), the process that the NQF requires to submit measures is not feasible for the majority of the palliative care field. The absence of a national sample or 100 testing sites should not stand in the way of progress.

Another challenge our field continues to face is the perpetuation of silos in our healthcare delivery system. Since hospitals are designed to treat acute, potentially-reversible problems, they report post-discharge, patient-rated satisfaction surveys that completely miss the experience of the many patients who die during their stay. Likewise, skilled nursing facilities are viewed as places for rehabilitation, so federal reporting mandates focus only on restoration of function, even though many patients languish and die there. Since hospice is the place for dying, that is where the federal government mandates reporting of end-of-life quality measures, but again, that is not going to improve the quality of dying where most of it happens.

We have worked together with other organizations and independently to wade through numerous existing quality measures. Throughout these efforts, we have been struck by how difficult it is to design really good measures that capture the quality of palliative and end-of-life care. We are dismayed by the tendency to pursue and require "measures of convenience" in national reporting programs instead of focusing on fewer measures that really matter to patients. We continue to emphasize that more funding is needed for measure development in our field, as well as assistance from organizations like the NQF to shine a spotlight on measure gaps and encourage collaboration from various stakeholders, such as what's occurring in the NQF's measure incubator project. We encourage the NQF to help advocate for CMS to use the $75 million allocated by MACRA to invest in activities to fill critical measure gaps in our field and to collaborate with organizations such as ours that can provide appropriate clinical expertise to guide such work.

In late 2013, AAHPM and the Hospice and Palliative Nurses Association (HPNA) – in consultation with the Center to Advance Palliative Care (CAPC), NHPCO, The Joint Commission, the U.S. Department of Veterans Affairs and numerous other stakeholders – initiated the Measuring What Matters (MWM) project, which set forth to produce a consensus recommendation for a portfolio of performance measures that all hospice and palliative care programs could use for program improvement. The goal of MWM was to sort through all relevant published measures and select a concise set that would matter most for patients with palliative care needs across all settings. The belief is that voluntary adoption of these measures broadly in hospice and palliative care could lay the groundwork for benchmarking and meaningful comparison. We are now sorting through and prioritizing what will constitute Phase 2 of the project, which we hope will include more complex tasks, such as creating e-specifications and patient-reported outcome measures, field-testing altered, expanded and untested measures, and developing a common palliative care denominator.

Given the value of palliative care and our nation's rapidly aging population, there is an urgent need to focus attention on the quality and availability of palliative care services – both for acutely ill patients and older adults with life-limiting diseases. AAHPM continues to highlight the need for a common denominator that comprehensively captures the patient

population appropriate for palliative care. No measure currently used under federal quality reporting programs, or recommended for future years, focuses on this population exclusively. For example, there are currently no measures in the PQRS program that specifically address the broad category of palliative care for patients of any age, without being disease-specific. This puts palliative care providers (or really any provider who cares for seriously ill patients across settings) in the difficult position of either having to report on measures that are not clinically relevant, or being subject to CMS review and possible negative payment adjustments despite the high quality of care they provide.

For many years, experts have tried to develop a common denominator that will enable the field to target patients who are most likely to benefit from palliative care. Doing so involves striking the right balance between number and/or type of chronic conditions, extent of functional and cognitive impairments, and overarching quality of life. AAHPM is committed to the goal of transitioning from basic to more meaningful measures that focus on this broader population, important outcomes, care coordination, and patient experience. We have worked with relevant stakeholders to identify a priority list of measures and broader measure concepts that are either not quite ready for accountability purposes or are not necessarily as robust as NQF and CMS request (e.g. process vs. outcomes measures or not grounded in Grade A evidence). However, with some guidance, collaboration, and funded technical assistance, we believe these could evolve into more meaningful and useful measures and help to close the gap in measures that target the palliative care patient population specifically.

We know that NQF is increasingly emphasizing that measures developed from electronic data sources such as electronic health records (EHRs) and Qualified Clinical Data Registries (QCDRs) draw from a rich set of clinical data and can reduce data collection and reporting burden while supporting more timely performance feedback to physicians and other clinicians than is possible through traditional claims- or paper-based measures. While AAHPM agrees with this observation, our specialty has faced challenges in regards to electronic data collection and measure specifications.

The Institute of Medicine's (IOM) 2014 report titled *Dying in America*, recognized that in order to better understand and improve the care received by those at the end-of-life, we need better information about dying and about those with serious illness—not just about the demographic characteristics and health conditions of those who die, but also about their quality of life as they cope with declining health, the quality of the health care provided to them during this time, and the quality of their death. The ability to better capture this data would serve many other specialties, beyond Hospice and Palliative Medicine, and could drive patient-centered and family-oriented quality care. However, most EHRs still do not capture much of what is needed to measure palliative care quality. Processes and programs to develop standardized data elements and corresponding quality measures in partnership with large electronic medical record vendors (EPIC, Cerner) and other government agencies would spur this development.

We understand that it is not the responsibility of the NQF to solve these broader policy challenges. However, the NQF does have substantial influence over the type and scope of measures ultimately selected for both public and private payer reporting programs and seems to be playing an increasingly larger role in measure "incubation." We hope that as it continues down those paths that it keep in mind the critical need to accelerate the development and testing of *new* palliative care and end-of-life care measures that align with the goals of our organizations.

We are also working with the National Coalition for Hospice and Palliative Care (NCHPC) and other organizations in our field on issues and challenges related to measure development. Both the Hospice and Palliative Nurses Association (HPNA) and the Center to Advance Palliative Care (CAPC) endorse these comments at this time.

Thank you again for the opportunity to submit these comments. Please do not hesitate to

contact Katherine Ast, AAHPM's Director of Quality and Research (kast@aahpm.org), if we can provide any additional detail or assistance.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0209

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

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

Date of Submission: 2/29/2016

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

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

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

☐ Health outcome: Click here to name the health outcome

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

☐ Process: Click here to name the process

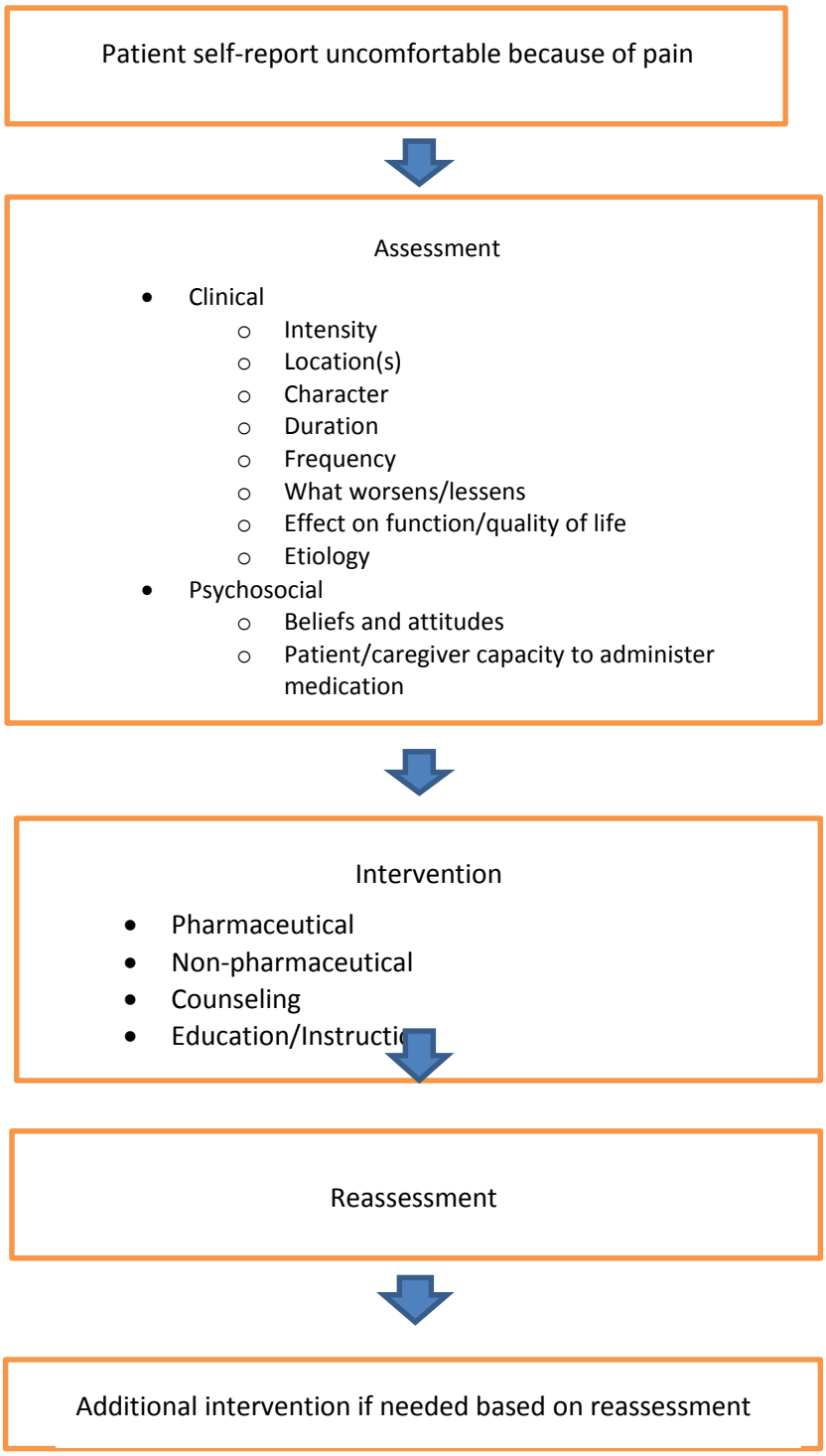
☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

Multiple care processes can influence achievement of comfort by a patient who self-reports pain.



1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

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.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section*** *1a.7*

☐ No → *report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist,* ***provide what is known from the guideline review of evidence in*** *1a.7*

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section *1a.7*

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

NATIONAL QUALITY FORUM

1. Evidence, Performance Gap, Priority – 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 subcriteria 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.docx

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., the benefits or improvements in quality envisioned by use of this measure)*

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 endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) 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 endorsement maintenance. 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 subcriterion on improvement (4b.1) 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 1b4, then provide a summary**

of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF;
  OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Patient/societal consequences of poor quality, Other
1c.2. **If Other:** Pain management is essential component of care at end-of-life

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
Inadequacies and need for improvement of pain managment for the dying have been pointed out by studies showing that 40 - 70% of Americans have substantial pain in the last days of life. Four our of 10 dying painte are in sever pain most of the time.
Poorly controlled pain diminishes patient quality of life and functional status, and causes suffering for patients and family caregivers. Pain is highly prevalent during the last week of life, so the timely evaluation and treatment of pain at the time of admission, before the patient is either unable to respond or detailed assessment becomes an additional burden is a priority.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
The Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT). The SUPPORT principle investigators. JAMA. 1995 274: 1591-98.

Hall CT, In Search of a Good Death, San Francisco Chronicle Tuesday, April 6, 1999

Fine PG. The ethical imperative to relieve pain at life´s end. J Pain Symptom Manage. 2002;23:273-277.

Conill C, Verger E, Henriquez I, Saiz N, Espier M, Lugo F, Garrigos A. Symptom prevalence in the last week of life. J Pain Symptom Manage. 1997; 14:328-331.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.**
(Describe how and from whom their input was obtained.)
The 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.

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 subcriteria 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, Musculoskeletal, Neurology, Pulmonary/Critical Care : Chronic Obstructive Pulmonary Disease (COPD), Renal

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Palliative Care and End of Life Care

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.)*
http://www.nhpco.org/patient-outcome-and-measures/comfortable-dying-measure

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.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date 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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Patients whose pain was brought to a comfortable level (as defined by patient) within 48 hours of initial assessment.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Look back period for the measure is 48hours (2 days) after patient report of being uncomfortable because of pain at initial assessment.

S.6. **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, 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.
Number of patients who replied "yes" when asked if their pain was brought to a comfortable level within 48 hours of initial assessment.

S.7. **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.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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 are able to self report pain information and replied "yes" when asked if they were uncomfortable because of pain at the initial assessment.

S.10. **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.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
None

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
N/A

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
N/A

S.16. **Type of score:**
Rate/proportion

If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; 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.
4. 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.
2. 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.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above 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. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a 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.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, 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: “Wasyour 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.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Patients who are able to answer the initial question ("Are you uncomfortable because of pain?") but who are unable to self-report at the time the follow up question ("Was your pain brought to a comfortable level within 48 hours of the start of hospice services?")will have missing data for calculation of the numerator. Responses for these patients are not imputed nor are they deleted from the denominator.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Patient Reported Data/Survey

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(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.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility, Population : National

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospice
If other:

S.28. **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.)
2a. Reliability – See attached Measure Testing Submission Form **2b. Validity – See attached Measure Testing Submission Form** 0209_Testing_attachment_v2_-635936652686462403.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 0209
Measure Title: Comfortable Dying
Date of Submission: 2/29/2016
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☒ Outcome (*including PRO-PM*)
☐ Cost/resource	☐ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

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

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

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

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

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

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing

of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

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

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.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☒ clinical database/registry	☒ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).
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.26*)**	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

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

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.

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 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.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).
Age and gender.

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 inherhently be expected to change rapidly (interventions to achieve better pain control can be and often are institituted 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.

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.

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

2b2. VALIDITY TESTING
2b2.1. 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*)

2b2.2. 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.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

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.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)
The results support validity of the measure for understanding how well hospices were able to relieve patients admitted in pain.

2b3. EXCLUSIONS ANALYSIS
NA ☐ no exclusions — *skip to section 2b4*

2b3.1. 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*)

N/A. Exclusions not examined; no patient level data available.

See 2b3.3 below.

2b3.2. 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*)

N/A. Exclusions not examined; no patient level data available.

See 2b3.3 below.

2b3.3. 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*)
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.

2b4. 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.

2b4.1. What method of controlling for differences in case mix is used?
☒ **No risk adjustment or stratification**
☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (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.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)
N/A

N/A

2b4.4a. What were the statistical results of the analyses used to select risk factors?
N/A

N/A

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)
N/A

N/A

2b4.5. 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

N/A

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. 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*)

2b4.11. 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*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE
2b5.1. 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)*

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 whos 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.

2b5.2. 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.

2b5.3. 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 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.

2b6. 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 SDS 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 SDS 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.

2b6.1. 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*)
N/A

2b6.2. 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*)
N/A

2b6.3. 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*)
N/A

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)
N/A

2b7.2. 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*)
N/A

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Samples collected for this measure had very few missing data points, and certainly not enough to bias the measure for a subpopulation of patients.

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)*
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.**
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.**
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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM 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.

Planned	Current Use (for current use provide URL)
Regulatory and Accreditation Programs Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Quality Improvement (Internal to the specific organization)	

4a.1. **For each CURRENT use, checked above, provide:**

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

d. Measure is included in PQRS. Level of utilization, if any, is unknown.
f.and g. NHPCO provides data collection and comparative reporting for voluntary submission of data by participating hospices. For 2014, 156 hospices provided aggregated measure data for 20.548 patients.

4a.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?*)

4a.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.*)

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

National mean measure scores for 2014 were Q1 66.3; Q2 61.2; Q3 67.5; Q4 63.1 with a mean for the year of 64.7. National mean measure scores for 2013 were Q1 63.1;Q2 65.4; Q3 66.7; Q4 60.8 with a mean for the year of 64.0. These scores demonstrate little difference for the two most recent years of complete data.

4b.2. **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.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
N/A

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

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 completely harmonized?

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:** Carol, Spence, cspence@nhpco.org, 703-837-3137-
Co.3 **Measure Developer if different from Measure Steward:** National Hospice and Palliative Care Organization
Co.4 **Point of Contact:** Carol, Spence, cspence@nhpco.org, 703-837-3137-

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 dervived from a commone 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?** Quarterly
Ad.5 **When is the next scheduled review/update for this measure?** 09, 2011

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:**

Appendix C – Related and Competing Measures for Pain

Pain Measures

Setting	Process				Outcome
	Screening	Assessment	Care plan	Assessment & Care Plan	
Hospice	✓ 1634	✓ 1637			✓ 0209
Hospital	✓ 1634	✓ 1637			
Home health					✓ 0177
Ambulatory (cancer pts only)	✓ 1628	✓ 0384	✓ 0383		
Ambulatory (all pts)				✓ 0420	
Nursing Facility					✓ 0676, 0677

Pain in hospice setting

	1634 - Screening	1637- Assessment	0209 – Comfortable Dying
Type	Process	Process	PRO-PM
Setting	Hospice/Hospital	Hospice/Hospital	Hospice
Numerator	Screen/severity noted using quantitative standardized tool at 1st encounter	Comprehensive assessment (severity, etiology, impact)	Pain brought to comfortable level w/n 48 hours
Denominator	In hospice or receiving specialty PC	Those with pain at 1st assessment	In pain at initial assessment
Exclusions	<1 day in hospital	<1 day in hospital	<18; cannot self-report; language

Appendix D – Updated reliability estimates for 8 PRO-PMs from Hospice CAHPS survey, April-September, 2015

CAHPS Hospice Survey – REVISED table and text regarding hospice inter-unit reliability

We updated our calculations of inter-unit (i.e., hospice-level) reliability to reflect newly available data; specifically, we combined CAHPS Hospice Survey Data Warehouse data from Quarter 3 2015 with data from Quarter 2 2015. (Quarter 3 data were not ready for analysis until after our original submission to NQF.) As in our original submission to NQF, we examined reliability for each measure using intra-class correlations (ICCs) computed from the case mix-adjusted 0-100 top-box scores. In addition, we used these ICCs to calculate the reliability that would be obtained if 200 surveys were completed per hospice using the Spearman Brown prediction formula (Allen and Yen, 1979). When entities such as hospices are being compared, composite reliability greater than 0.70 is commonly considered adequate (Hargraves, Hays & Cleary, 2003; Nunnally 1994).

These updated analyses show that five of the eight measures exhibit acceptable hospice-level reliability of 0.70 or greater at 200 completes per hospice (*Getting Hospice Care Training, Hospice Team Communication, Getting Timely Care, Getting Emotional and Religious Support, Rating of Hospice* and *Willingness to Recommend the Hospice*; Table 2a2.3b [UPDATED July 2016]*)*. At reliabilities of 0.66 to 0.69, *Getting Emotional and Religious Support*, *Getting Help for Symptoms* and *Treating Family Member with Respect*, are close to achieving the adequate reliability threshold.

Table 2a2.3b (UPDATED July 2016). Hospice-Level Reliability for CAHPS Hospice Survey Measures, Quarters 2 and 3 2015

Composite or single-item measure	Intraclass Correlation Coefficient (ICC)	Hospice reliability @ N=200 per hospice
Composite Measures		
Getting Hospice Care Training (5-items)	.020	.81
Hospice Team Communication (6-items)	.018	.78
Getting Timely Care (2-items)	.016	.76
Getting Emotional and Religious Support (3-items)	.011	.69
Getting Help for Symptoms (4-items)	.010	.66
Treating Family Member with Respect (2-items)	.011	.68
Global Measures		
Rating of Hospice	.015	.75
Willingness to Recommend	.021	.81

Appendix E – Preliminary data for NQF #1639: Dyspnea Screening

UNC
THE CECIL G. SHEPS CENTER
FOR HEALTH SERVICES RESEARCH

THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

725 MARTIN LUTHER KING JR. BLVD.
CAMPUS BOX 7590
CHAPEL HILL, NC 27599-7590

T 919.966.7100
F 919.966.8918

TO: Palliative and End-of-Life Care Project Standing Committee

FROM: Dr. Laura C. Hanson, University of North Carolina at Chapel Hill

DATE: July 15, 2016

SUBJECT: Preliminary data for NQF #1639: Dyspnea Screening from an acute care hospital

---

CONTEXT

The National Quality Forum (NQF) held an in-person meeting May 10 and 11, 2016 of the Palliative and End-of-Life Care Standing Committee to evaluate new quality measures (QM) and currently endorsed QMs undergoing maintenance review. During this meeting the committee did not reach consensus that an opportunity for improvement exists for NQF #1639 – Dyspnea Screening. We present preliminary data from hospital-based palliative care to show that a performance gap is present in a hospital acute care setting.

Additional analysis of palliative care data from multiple sites and settings is being collected in 2016-2017. This data will be reported to NQF and the Standing Committee.

METHODS

Chart abstractions were completed for 895 patients who received a specialty palliative care consultation admitted to an acute care hospital for at least 1 day from January 2014 to June 2015. Data was collected via chart abstraction on almost every patient seen by the UNC Palliative Care Consult Service during this time frame in the manner originally endorsed by NQF.

The age of the patients ranged from 23 to 103 years, with the mean age 64. Patients were predominantly Caucasian (60%), with smaller subgroups who were African American (30%) and Hispanic / Latino (3%).

FINDINGS

As this measure was designed to pair with NQF #1638: Dyspnea Treatment, we are including preliminary data for both the Dyspnea Screening and Treatment measures.

	Total
Dyspnea Screen QM (1639)	(n=895)
(Rated+UtR+Other) / n	81.8%
Dyspnea Tx QM (1638)	(n=159)
(Tx=Any Yes+Pt Refused) / (SOB ≥ 4)	79.9%
Pain Screen QM (1634)	(n=895)
(Rated+UtR+Other)/n	90.8%
Dyspnea Screen	(n=895)
Severity Rated	69.5%
Unable to Respond on Screening	12.3%
Unknown/Not Rated	18.2%
Pain Screen	(n=895)
Rated	80.0%
Unable to Respond	10.8%
Unknown/Not Rated	9.2%

Appendix F – Evaluation summary for #1626

1626 Patients Admitted to ICU who Have Care Preferences Documented

Submission **| Specifications**

Description: Percentage of vulnerable adults admitted to ICU who survive at least 48 hours who have their care preferences documented within 48 hours OR documentation as to why this was not done.

Numerator Statement: Patients in the denominator who had their care preferences documented within 48 hours of ICU admission or have documentation of why this was not done.

Denominator Statement: All vulnerable adults admitted to ICU who survive at least 48 hours after ICU admission.

Exclusions: None

Adjustment/Stratification:

Level of Analysis: Facility

Setting of Care: Hospital/Acute Care Facility

Type of Measure: Process

Data Source: Paper Medical Records

Measure Steward: RAND Corporation

STANDING COMMITTEE MEETING [05/10/2016-05/11/2016]

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

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

1a. Evidence: **H-0**; **M-24**; **L-0**; **I-0**; 1b. Performance Gap: **H-1; M-22; L-0; I-1;**

Rationale:

- For the 2012 endorsement evaluation, the developers cited two systematic reviews linking advance care planning to better patient outcomes and providing evidence that patients want to communicate their care preferences to their physicians. No updated evidence was submitted for the current evaluation. However, Committee members referenced additional guideline recommendations released since the 2012 evaluation and included in the submission for measure #1641; these recommendations support advance care planning and shared decision making.
- The Committee noted that the evidence presented does not pertain to the *documentation* of the care preferences themselves as much as to the importance of care preferences and the discussion around those.
- For the 2012 endorsement evaluation, the developers provided performance data from four individual studies with measure results ranging from 9% to 63.7%. However, these results were based on data that are more than five years old, and no updated performance data was presented for the current evaluation. However, using their own experience and judgement, Committee members agreed that there still is opportunity for improvement, and suggested there may be disparities in care for this measure.

1626 Patients Admitted to ICU who Have Care Preferences Documented

2. Scientific Acceptability of Measure Properties: The measure does not meet the Scientific Acceptability criteria

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

2a. Reliability: **H-0**; **M-13**; **L-8**; **I-3** 2b. Validity: **H-0**; **M-8**; **L-11**; **I-5**

Rationale:

- Committee members expressed concerns that limiting the denominator of the measure to 'vulnerable adults' as defined in the specifications (i.e., age 75 or older; score >2 on the Vulnerable Elder Survey-13, life expectancy <6 months, Stage IV cancer, receiving hospice care) would not capture important patient populations, including patients with acute respiratory failure.
- The developer clarified that the timing of the admission to ICU "begins" when the admission orders are written.
- Committee members asked the developers to explain the numerator requirement of having "care preferences documented within 48 hours of ICU admission", noting that the submission also indicates that "simply having an advance directive or other advance care planning document or POLST in the medical record does not satisfy this criterion". The developers clarified that the measure assesses whether a discussion regarding care preferences with either the patient or the family occurred within 48 hours of ICU admission and that discussion could be with non-ICU providers and could occur during the hospitalization but prior to the ICU admission. The developers noted that care preference information may not always be included in an advance directive and further clarified that existence of an advance directive in the record is not sufficient.
- For the 2012 endorsement evaluation, the developers provided inter-rater reliability statistics from two studies in which the kappa value for the denominator was 0.87 to 0.95 and the kappa value for the numerator was 0.86 to 0.87 and 0.86, indication acceptable agreement. The developers did not provide updated reliability testing.
- The Committee did not reach consensus on reliability of the measure due to concerns about the ability to consistently apply the numerator specifications.
- Validity testing at the time of the 2012 endorsement evaluation included three face validity assessments by three expert panels using a modified Delphi method. Developers did not update validity testing for the current evaluation.
- This measure did not pass the validity subcriterion. Committee members noted that one of the face validity assessments was specific to cancer patients only, that none of the face validity assessments were specific to ICU patients, and that this measure was not assessed specifically but was instead discussed more generally.

3. Feasibility: H-X; M-X; L-X; I-X

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Data collection strategy can be implemented)

4. Usability and Use: H-X; M-X; L-X; I-X

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

1626 Patients Admitted to ICU who Have Care Preferences Documented

5. Related and Competing Measures
• This measure is related to one measure:
  ○ 1617: Patients Treated with an Opioid who are Given a Bowel Regimen
• The definition of "vulnerable adults" is harmonized between this measure and #1617.
 	
• This measure directly competes with two measures:
  ○ 0326: Advance Care Plan. Description: Percentage of patients aged 65 years and older who have an advance care plan or surrogate decision maker documented in the medical record or documentation in the medical record that an advance care plan was discussed but the patient did not wish or was not able to name a surrogate decision maker or provide an advance care plan
  ○ 1641: Hospice and Palliative Care – Treatment Preferences. Description: Percentage of patients with chart documentation of preferences for life sustaining treatments
• Because this measure did not meet the Validity subcriterion, there was no need for a best-in-class discussion between this measure and the other competing measures.

Standing Committee Recommendation for Endorsement: **DID NOT PASS SCIENTIFIC ACCEPTABILITY** <u>Rationale</u> • The Committee did not feel that the face validity assessments that were conducted for this measure were specific enough.
6. Public and Member Comment
7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X
8. Board of Directors Vote: Y-X; N-X
9. Appeals

Appendix G – Measure worksheet: #1626

ADDITIONAL MATERIALS PROVIDED BY THE STEWARD

For all four RAND palliative care measures proposed for maintenance (1617,1624,1626,1628), validity was confirmed using RAND/UCLA Appropriateness Methodology (1). Quality indicators 1617, 1624, and 1626 were deemed valid using this methodology by both Assessing Care of Vulnerable Elders (ACOVE-3) and Assessing Symptoms Side Effects and Indicators of Supportive Treatment (ASSIST) expert panels. Quality indicator 1628 was deemed valid by the ASSIST expert panel. The methodology includes the following steps:

1. A content expert proposes quality indicators using existing guidelines, review criteria, and expert opinion.
2. Extensive systematic literature review for candidate quality indicators is completed to understand evidence for each measure and linkage to important patient outcomes. This informs a report for each quality indicator on the state of the evidence.
3. An interdisciplinary expert panel (of at least 9 members) votes on validity of each quality indicator after review of the report of supporting literature on a 1-9 scale (RANGE 1-3 considered not valid, RANGE 4-6 considered indeterminate, and 7-9 considered valid) prior to an in-person meeting.
4. During an in-person meeting of the expert panel, average pre-panel scores are reviewed and areas of disagreement are discussed in detail. Following the discussion, the members of the expert panel vote again.
5. In order to be considered a valid measure with this method, the average score for post in-person meeting ratings for a 9 person panel must be in the highest tertial (7-9) without disagreement (disagreement means that 3 panelists rate validity in the lowest tertial (1-3) and three rate in the highest tertial.
6. ACOVE Measures that had an average validity score of 7 or higher (without disagreement) after expert panel review also underwent additional review by a Clinical Committee for coherence and content validity (2).

For full details, we have attached reference number 2 (ACOVE) and reference number 6 (ASSIST).

References

1. Shekelle P. The Appropriateness Method. Medical Decision Making. March-April 2004, p.228-231.
2. Shekelle P., MacLean CH, Morton S, NS Wenger. Assessing Care of Vulnerable Elders: Methods for Developing Quality Indicators. Ann Intern Med. 2001;135:647-51.
3. Wenger NS, Roth CP, Shekelle P. and the ACOVE Investigators. Introduction to the Assessing Care of Vulnerable Elders-3 Quality Indicator Measurement Set. JAGS. 2007:55:S247-S252.

4. Wenger NS, Rosenfeld K. Quality Indicators for End-of-Live Care in Vulnerable Elders. Ann Intern Med. 2001:135:677-685.
5. Lorenz KA, Rosenfeld K, Wenger NS. Quality Indicators for Palliative and End-of-Life Care in Vulnerable Elders. JAGS. 2007;55:S318-S326.
6. Lorenz KA, Dy SM, Naeim A, Walling AM, et al. Quality Measures for Supportive Cancer Care: The Cancer Quality-ASSIST Project. Journal of Pain and Symptom Management. 2009;37:943-964.

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 1626
Measure Title: Patients Admitted to ICU who Have Care Preferences Documented
Measure Steward: RAND Corporation
Brief Description of Measure: Percentage of vulnerable adults admitted to ICU who survive at least 48 hours who have their care preferences documented within 48 hours OR documentation as to why this was not done.
Developer Rationale: The aim of this measure is to assist healthcare providers in providing care that is consistent with patient preferences.

Numerator Statement: Patients in the denominator who had their care preferences documented within 48 hours of ICU admission or have documentation of why this was not done.
Denominator Statement: All vulnerable adults admitted to ICU who survive at least 48 hours after ICU admission.
Denominator Exclusions: None

Measure Type: Process
Data Source: Paper Medical Records
Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Feb 14, 2012 **Most Recent Endorsement Date:** Feb 14, 2012

Maintenance of Endorsement -- Preliminary Analysis

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 *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

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

Evidence Summary and Summary of prior review in 2012

- The developer does not provide a rationale for the link between the process of care (documentation of care preferences within 48 hours) and patient outcomes.
- The developer cites two systematic reviews as evidence for this measure, although they do not provide summaries of the evidence from these reviews.
    - NQF staff were able to access the Lorenz, et al. (2007) article. This review includes evidence linking advanced care planning and better patient outcomes and provides evidence that patients want to communicate their care preferences with their physicians. However, this review found no empirical evidence supporting documentation of advanced care plans. Evidence concerning care preference documentation was based on evidence about patient desire not to live permanently comatose, mechanically ventilated, or tube fed, along with evidence that physicians and surrogate decision makers often do not know patients' preferences concerning life-sustaining treatment preferences.
- The developer states there is no clinical trial data demonstrating a link between this process of care and outcomes.
- During its 2012 evaluation of the measure, the Palliative/End-of-Life (EOL) Committee agreed that the evidence for the measure is solid, even though no clinical trials were cited.

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: N/A

Exception to evidence – N/A

Guidance from the Evidence Algorithm

Process measure based on systematic review(Box 3) → QQC not provided, but could use Staff information from Lorenz article and evidence presented for measure #1641 (treatment preferences measure) → Moderate

Questions for the Committee:

- *What is the relationship of this measure to patient outcomes?*
- *How strong is the evidence for this relationship?*
- *Is the evidence directly applicable to the process of care being measured?*

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

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 the 2012 submission, the developer provided data from 4 individual studies to demonstrate opportunity for improvement. Across these studies, the sample size ranged from 6 to 349 and measure results ranged from 9% to 63.7%. The developer did not provide updated performance data.

Disparities

- The developer did not provide data or cite studies examining disparities in measure performance.

Questions for the Committee:

- *The results provided are more than 5 years old. Do these results demonstrate that there is still a gap in care that warrants a national performance measure?*
- *Are you aware of evidence that disparities exist in this area of healthcare?*

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

Rationale: No information is provided to determine whether there is still opportunity for improvement in documenting care preferences for ICU patients.

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a.

* The evidence to support the measure are two systematic reviews, neither of which provides evidence about patient outcomes. The SR authors do state that based on research about what patients prefer at end of life, it would likely not be dying in the ICU with extreme life-sustaining measures. Therefore, it would seem that knowing care preferences would allow healthcare professionals to support those preferences and in doing so, provide patients with the outcome they desire. I rate the evidence as moderate.
* The measure relates to desirable outcomes but no clinical trials are cited. no clinical data identifies a link between the process of care and the outcome but 2012 committee felt that evidence for the measure is solid. No new evidence was identified by the developer. Preliminary rating for evidence is moderate
* this is a process measure that could easily be converted to an outcome measure (e.g. % pts whose death was consistent with expressed wishes) but currently has no established relationship with outcomes. The process being measured is documentation; not clear to me that the measure gets to underlying questions of determining, updating, or adhering to patient's wishes. NO QOC provided but an expert panel (2014) and ICSI guideline (2013, low quality evidence but strong recommendation) support it.

1b.
* The data provided by the developer is more than 5 years old but there are newer studies that show vulnerable populations (includes elders) do not receive quality care at EOL and that often they come to the ICU without documented care preferences. Based on current data, I think this measure could at least provide information on different subgroups who do not have documentation of care preferences within 48 hours of ICU admission. However, it has currently not done so and I am giving a rating of insufficient.
* Preliminary rating for opportunity for improvement is insufficient
* No updated data provided. A 2013 paper is cited showing n=150 VA patients of whom ~ 63% met measure, leaving 37% gap (unable to locate this paper). A JAGS paper in 2014 reported from Bronx VA that 20 of 93ts had no documentation of ACP.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications
Maintenance measures **– no change in emphasis – specifications should be evaluated the same as with new measures**

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

Data source(s): The data source specified for this measure is paper medical records.
Specifications:

- This measure is specified for the facility level of analysis for the hospital setting of care. A higher score indicates better quality.
- The numerator (patient in the denominator who had their care preferences documented within 48 hours of ICU admission) and denominator (all vulnerable adults admitted to the ICU) are collected from the medical chart.
- "Vulnerable" adults are those with any of the following characteristics:
  - 75 years of age or older
  - Score >2 on the Vulnerable Elder Survey-13
  - Life expectancy <6 months
  - Stage IV cancer
- Documentation of care preferences must occur within 48 hours of ICU admission.
- Documentation of having an advance directive, other care advanced care planning document, or POLST in the medical record is not sufficient to be counted in the numerator (i.e., the 48-hour timeframe also must be satisfied).
- There are no exclusions for the measure.
- A calculation algorithm is provided

Prior evaluation

- **T**he Committee voiced the following concerns:
  - Concern that definitions are too broad for implementation
  - Concern that many patients may not be communicative in the first 48 hours in the ICU
  - Concern that this measure is an ICU documentation issue rather than one that captures the intended process

Questions for the Committee :

- *Is this a measure of documentation of a discussion or a measure of documentation of actual preferences?*
- *Is the 48-hour survival time an appropriate eligibility criterion?*
- *This measure is specified for data collection from paper medical records. Is this reasonable given current use of EHRs in hospitals?*
- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

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, summarize the reliability testing from the prior review:

- The developer reports reliability testing was previously tested in the ACOVE3 and ASSIST studies.

Describe any updates to testing

- No updated testing was provided.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☒ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing:

- The developer cited two reliability studies, but noted the methodology of only one (data "re-abstractions"). Testing was conducted using data from 47 inpatient decedents from the ACOVE study and 22 inpatients decedents from the ASSIST study.

Results of reliability testing:

- For the ACOVE study, developers report an overall eligibility (denominator) kappa value of 0.95 and a specific care (numerator) kappa value of 0.87.
    - The kappa value represents the proportion of agreement between two raters/abstractors that is not explained by chance alone. A value of 1.0 reflects perfect agreement; a value of 0 reflects agreement that is no better than what would be expected by chance alone. A kappa of 0.95 means that the raters agreed 95% of the time over and above what would be expected by chance alone. According to the Landis and Koch classification, this represents "almost perfect" agreement. A kappa of 0.87 means that the raters agreed 87% of the time over and above what would be expected by chance alone. According to the Landis and Koch classification, this represents "almost perfect" agreement.
- For the ASSIST study, developers report an overall eligibility (denominator) kappa value of 0.87 and a specific care (numerator) kappa value of 0.86. According to the Landis and Koch classification, both of these kappa values represent "almost perfect" agreement.

Guidance from the Reliability Algorithm

Precise specifications (Box 1) → empirical reliability testing conducted with measure as specified (Box 2) → testing at score level not conducted (Box 4) → data element testing conducted with appropriate method (Box 9) → High level of agreement between raters (Box 10a) → Moderate

Questions for the Committee:

- *Are the test samples adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*
- *No updated testing information is presented. The prior testing demonstrated good reliability. Does the Committee think there is a need to re-discuss and re-vote on reliability?*

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

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ Yes ☐ Somewhat ☐ No

- The Lorenz, 2007 article specifically includes "48 hours" in its review and recommendations.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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, summarize the validity testing from the prior review:

- Three studies (Shekelle et al., 2001; Wenger et al., 2007; and Lorenz et al., 2009) involved the assessment of the ACOVE quality indicators—including this measure—by three expert panels.

Describe any updates to validity testing

- The developer attests to there been no updates to validity testing.

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

☒ **Face validity only**
☐ **Empirical validity testing of the measure score**

Validity testing method:

- The developers cited the Hagashi, 2005 and Zingmond, 20011 studies, stating the validity of the process-outcome relationship is supported by the study findings; however, these measures do not appear to be included in the measure sets discussed in these articles.
- Three expert panels assessed the set of ACOVE quality indicators, including this measure.

Validity testing results:

- The developers do not provide the results of these expert panel reviews.

Questions for the Committee:

- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*
- *No updated testing information is presented. The prior testing reflects face validity only and results were not presented. Does the Committee think there is a need to re-vote on validity, assuming threats to validity were adequately assessed?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- There are no exclusions for this measure.

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- The developer highlights the variation in performance found in the cited in research studies, but does not provide information on whether statistically significant or clinically/practically meaningful differences can be identified.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods: N/A

2b7. Missing Data

- The developer does not provide information regarding missing data (although this may not be an issue given data are collected from paper medical records).

Guidance from the Validity Algorithm
Specifications consistent with evidence (Box 1) → potential threats to validity not assessed (Box2) → Insufficient

If no concerns around missing data or meaningful differences→ empirical validity testing not conducted (Box 3) → face validity systematically assessed (Box 4) →results not provided (Box 5) →Insufficient

Preliminary rating for validity: ☐ High ☐ Moderate ☐ Low ☒ Insufficient
Rationale: The developer does not provide the results of the expert panel review and does not provide information related to missing data or meaningful differences. If this information is provided, the measure is eligible for a MODERATE rating.

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a, 2b.
Reliability – Specifications
* Due to the specificity of the measure (patients with AD do not count in the numerator and the 48 hour limit), I feel that the measure is one of documentation of a discussion rather than actual preferences. If it was the latter, then it would seem that documented care preferences prior to admission to the ICU would count. I am not a doctor and so cannot address the 48-hour survival time question. I think that limiting the measure to paper medical records is problematic and would eliminate the use of the measure in many facilities. One concern I have is the decision to use 75 years of age as a specification. I am not sure how this was decided upon or if it is an appropriate inclusion criteria. In some states, an elder is deemed vulnerable at the age of 60 (e.g. WI). I have concerns about the likelihood that the measure could be consistently implemented due to the use of paper records. Situations in the ICU can be very chaotic and there could be issues in insuring the criteria are met and that the attempt to discuss is recorded.
* Preliminary rating for reliability is moderate. The SC should discuss and revote on reliability
* No updated testing provided. Sample sizes reported were small (n=22). Studies submitted were produced by measure developer, but were judged to be "good".

Validity – Specifications
* The use of "48 hours" is included in BOTH articles that the developer referenced.
* Specifications are consistent with the evidence that the measure is one that the target population finds meaningful
* Face validity is reasonable, "48 hours" specifically mentioned in Lorenz 2007 article. 3 expert panels assessed ACOVE measures, including this one. None of these results are provided. Recommend re-assessment of validity.

Reliability – Testing
* Again, implementation will be limited due to the specification that paper medical records are used. The reliability scores are very solid so I do not feel there is a need to re-discuss and re-vote on reliability. I would rate reliability as moderate.
* Yes- reliability testing was at the moderate level with data element but not with score levels

Validity Testing
* Face validity and expert panel validity are very low tests of validity and so I have concerns, especially since the developers did not provide the results of the expert panels. I feel that based on the information available, the Committee should re-vote on validity.
* Preliminary rating is insufficient
*no

Threats to Validity
* No exclusions and no risk adjustments. In terms of difference and missing data, the developer did not provide this information. Without more information, I have concerns about the validity of this measure and would deem the rating as insufficient.
* missing data information not available but most likely does not constitute a treat to validity

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.

- Some data elements are available in electronic form. However, the documentation of care preferences often must be abstracted from the medical record.
- In the 2012 endorsement of the measure, the Palliative and End-of-Life Care Steering Committee agreed that the data for this measure easily obtainable through EMRs or medical record chart documentation.

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?*
- *How burdensome will it be for medical abstractors to obtain needed data from a paper medical record?*

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

Committee pre-evaluation comments
Criteria 3: Feasibility

* I think that the elements of this measure are routinely generated and used in the ICU. I also think that the required data elements are available in electronic form. My concern is that it would be burdensome for medical abstractors to obtain needed data from a paper medical record - the records are not as well organized and it is difficult to search and find information (not true for electronic records). I would rate feasibility as low.
* Feasibility is moderate as the documentation of care preferences often must be abstracted from the medical record, however data should be available on HER
* measure not currently implemented

Criterion 4: Usability and Use
Maintenance measures **– increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences**

4. Usability and 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.

Current uses of the measure [from OPUS]
Publicly reported? ☐ **Yes** ☒ **No**

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

Accountability program details

- The developer reports some internal quality improvement programs are in the planning phase to incorporate the measure in a future set.

Improvement results

- Longitudinal data for these measures are not yet available.

Unexpected findings (positive or negative) during implementation:

- The developer did not report any unexpected findings.

Potential harms:

- The developer did not report any unintended consequences.

Feedback :

- In February 2013, the Measure Application Partnership supported the measure for inclusion in the Physician Quality Reporting System (PQRS) to address a core concept not addressed in the programs' measure set. Public comments from the Center to Advance Palliative Care (CAPC) and the Society of Hospital Medicine (SHM) supported the inclusion of this measure in the PQRS. However, this measure is not currently included in the PQRS program.

***Questions for the Committee**:*

- *NQF guidelines indicate performance measures should be used in at least one accountability program three years after endorsement; given that this measure is not yet in use, should the measure retain endorsement without a clear path to use in accountability programs or public reporting?*
- *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: This measure does not appear to be in use and no definite plans are presented that suggest this is likely to change. There are no data on current performance or improvement in performance over time.

Committee pre-evaluation comments

Criteria 4: Usability and Use

* Since it has been three years since the measure received endorsement and it does not appear to be either in use or plans developed for its use, then I think that a clear path for use is indicated. I struggle with whether results from this measure could be used to further the goal of high-quality, efficient healthcare. The specification that documentation of an AD, POLST, etc. are not to be included in the numerator is concerning. It would seem that if the argument is that in order to have a quality dying process one's care preferences must be honored, then why is it that only those within 48 hours of ICU admission be considered? I would rate usability and use as low.
* Preliminary rating is low because the measure is not currently publicly reported and not being utilized in an accountability program although that use is supported in PQRS by CAPC and SHM
* measure not currently implemented

Criterion 5: Related and Competing Measures

Related measures

- The definition of 'vulnerable adults' is harmonized with another RAND measure:
  - 1617: Patients Treated with an Opioid who are Given a Bowel Regimen

Competing measures

- 0326: Advance Care Plan *[individual and clinician group/practice-level measure in various settings including hospital and hospice]*
- 1641: Hospice and Palliative Care – Treatment Preferences

Harmonization

- The Committee likely will be asked to select a best-in-class measure. If multiple measures are justified, recommendations for combining or harmonizing measures may be solicited.

Pre-meeting public and member comments

Name: Michele Galioto, RN, MSN
Organization: Oncology Nursing Society
Comment: There could be patients who survive to meet this criteria but who are unable to communicate their preferences and/or do not have preferences documented.

Name: Michele Galioto, RN, MSN
Organization: Oncology Nursing Society
Comment:
Overall, ONS recommends differentiating between palliative and end of life care in introductory information. ONS does not define palliative care as equal to end of life care. Hospice care is a form of palliative care but not inclusive of all palliative care. Palliative care should begin at the point of diagnosis or awareness of symptoms and continue throughout the trajectory of treatment through end of life care. See ONS position statement on palliative care for further detail.

ONS is also in favor of including recommendations for intervals of assessment as the current measures imply that one screening is sufficient. Screening at each patient encounter may be more appropriate.

Name: Katherine Ast, MSW, LCSW
Organization: American Academy of Hospice and Palliative Medicine
Comment:

On behalf of the palliative care community, we thank the National Quality Forum for convening its Palliative and End-of-Life Care 2015-2016 Project and for the opportunity to provide preliminary feedback on the palliative and end-of-life care measures that will soon be evaluated by the project's Standing Committee.

The American Academy of Hospice and Palliative Medicine (AAHPM) is the professional organization for physicians specializing in hospice and palliative medicine, and our membership also includes nurses and other health and spiritual care providers committed to improving quality of life for seriously ill patients and their families. We support the pursuit of interdisciplinary, team-based palliative care and its emphasis on care coordination, pain and symptom management, shared decision making, and patient-centered goal-setting. The provision of palliative care has been shown to improve patient experience and satisfaction, i reduce caregiver burden, ii and increase survival iii; it has also been shown to reduce needless hospital admissions and readmissions through effective care coordination and symptom management iv; and through these gains in quality, it reduces costs. v

While we do not, at this time, view any of the measures under consideration as particularly controversial, their limited scope reflects the critical ongoing gaps related to palliative and end-of-life care measurement and highlights the unique challenges that have contributed to those gaps. For example, the current set of measures under consideration is largely limited to cancer or hospice settings. These measures employ a narrow denominator (e.g., *hospice* patients rather than *dying* patients). This is certainly a good start, but measuring only hospice patients in order to improve the quality of end-of-life care is like searching for a lost dollar bill only where the light is good. It will not move the needle to the extent that we need it to. The National Hospice and Palliative Care

Organization (NHPCO) reports in its *2011 Facts and Figures* that only 42% of those who died in 2010 were enrolled in hospice. How do we measure the quality of end-of-life care for the majority of patients who die in hospitals, skilled nursing facilities, and homes without the benefit of hospice care? These are questions we have not yet been able to answer. The fact that the current set of measures under consideration by the NQF only includes *one* new measure (i.e., the Hospice CAHPS) also illustrates that the standard default pathways for measure development, testing, and endorsement are not working for the patients, providers and researchers in our field. Patient and family preferences and experience of care are critical elements of quality palliative care, and ongoing funding, data analysis and personnel are required to develop these kinds of measures and keep them endorsed and in use. For example, the current NQF requirement for measure developers to test survey instrument data elements in addition to the measures themselves (double testing) poses a barrier to advancing the field. While the process of submitting the PEACE measures from the University of North Carolina has gone well because of RTI's support and the national data coming from the Hospice Item Set (HIS), the process that the NQF requires to submit measures is not feasible for the majority of the palliative care field. The absence of a national sample or 100 testing sites should not stand in the way of progress.

Another challenge our field continues to face is the perpetuation of silos in our healthcare delivery system. Since hospitals are designed to treat acute, potentially-reversible problems, they report post-discharge, patient-rated satisfaction surveys that completely miss the experience of the many patients who die during their stay. Likewise, skilled nursing facilities are viewed as places for rehabilitation, so federal reporting mandates focus only on restoration of function, even though many patients languish and die there. Since hospice is the place for dying, that is where the federal government mandates reporting of end-of-life quality measures, but again, that is not going to improve the quality of dying where most of it happens.

We have worked together with other organizations and independently to wade through numerous existing quality measures. Throughout these efforts, we have been struck by how difficult it is to design really good measures that capture the quality of palliative and end-of-life care. We are dismayed by the tendency to pursue and require "measures of convenience" in national reporting programs instead of focusing on fewer measures that really matter to patients. We continue to emphasize that more funding is needed for measure development in our field, as well as assistance from organizations like the NQF to shine a spotlight on measure gaps and encourage collaboration from various stakeholders, such as what's occurring in the NQF's measure incubator project. We encourage the NQF to help advocate for CMS to use the $75 million allocated by MACRA to invest in activities to fill critical measure gaps in our field and to collaborate with organizations such as ours that can provide appropriate clinical expertise to guide such work.

In late 2013, AAHPM and the Hospice and Palliative Nurses Association (HPNA) – in consultation with the Center to Advance Palliative Care (CAPC), NHPCO, The Joint Commission, the U.S. Department of Veterans Affairs and numerous other stakeholders – initiated the Measuring What Matters (MWM) project, which set forth to produce a consensus recommendation for a portfolio of performance measures that all hospice and palliative care programs could use for program improvement. The goal of MWM was to sort through all relevant published measures and select a concise set that would matter most for patients with palliative care needs across all settings. The belief is that voluntary adoption of these measures broadly in hospice and palliative care could lay the groundwork for benchmarking and meaningful comparison. We are now sorting through

and prioritizing what will constitute Phase 2 of the project, which we hope will include more complex tasks, such as creating e-specifications and patient-reported outcome measures, field-testing altered, expanded and untested measures, and developing a common palliative care denominator.

Given the value of palliative care and our nation's rapidly aging population, there is an urgent need to focus attention on the quality and availability of palliative care services – both for acutely ill patients and older adults with life-limiting diseases. AAHPM continues to highlight the need for a common denominator that comprehensively captures the patient population appropriate for palliative care. No measure currently used under federal quality reporting programs, or recommended for future years, focuses on this population exclusively. For example, there are currently no measures in the PQRS program that specifically address the broad category of palliative care for patients of any age, without being disease-specific. This puts palliative care providers (or really any provider who cares for seriously ill patients across settings) in the difficult position of either having to report on measures that are not clinically relevant, or being subject to CMS review and possible negative payment adjustments despite the high quality of care they provide.

For many years, experts have tried to develop a common denominator that will enable the field to target patients who are most likely to benefit from palliative care. Doing so involves striking the right balance between number and/or type of chronic conditions, extent of functional and cognitive impairments, and overarching quality of life. AAHPM is committed to the goal of transitioning from basic to more meaningful measures that focus on this broader population, important outcomes, care coordination, and patient experience. We have worked with relevant stakeholders to identify a priority list of measures and broader measure concepts that are either not quite ready for accountability purposes or are not necessarily as robust as NQF and CMS request (e.g. process vs. outcomes measures or not grounded in Grade A evidence). However, with some guidance, collaboration, and funded technical assistance, we believe these could evolve into more meaningful and useful measures and help to close the gap in measures that target the palliative care patient population specifically.

We know that NQF is increasingly emphasizing that measures developed from electronic data sources such as electronic health records (EHRs) and Qualified Clinical Data Registries (QCDRs) draw from a rich set of clinical data and can reduce data collection and reporting burden while supporting more timely performance feedback to physicians and other clinicians than is possible through traditional claims- or paper-based measures. While AAHPM agrees with this observation, our specialty has faced challenges in regards to electronic data collection and measure specifications.

The Institute of Medicine's (IOM) 2014 report titled *Dying in America*, recognized that in order to better understand and improve the care received by those at the end-of-life, we need better information about dying and about those with serious illness—not just about the demographic characteristics and health conditions of those who die, but also about their quality of life as they cope with declining health, the quality of the health care provided to them during this time, and the quality of their death. The ability to better capture this data would serve many other specialties, beyond Hospice and Palliative Medicine, and could drive patient-centered and family-oriented quality care. However, most EHRs still do not capture much of what is needed to measure palliative care quality. Processes and programs to develop standardized data elements and corresponding quality measures in partnership with large electronic medical record vendors (EPIC, Cerner) and other government agencies would spur this development.

We understand that it is not the responsibility of the NQF to solve these broader policy challenges. However, the NQF does have substantial influence over the type and scope of measures ultimately selected for both public and private payer reporting programs and seems to be playing an increasingly larger role in measure "incubation." We hope that as it continues down those paths that it keep in mind the critical need to accelerate the development and testing of *new* palliative care and end-of-life care measures that align with the goals of our organizations.

We are also working with the National Coalition for Hospice and Palliative Care (NCHPC) and other organizations in our field on issues and challenges related to measure development. Both the Hospice and Palliative Nurses Association (HPNA) and the Center to Advance Palliative Care (CAPC) endorse these comments at this time.

Thank you again for the opportunity to submit these comments. Please do not hesitate to contact Katherine Ast, AAHPM's Director of Quality and Research (kast@aahpm.org), if we can provide any additional detail or assistance.

1. IMPACT, OPPORTUITY, EVIDENCE - IMPORTANCE TO MEASURE AND REPORT

Importance to Measure and Report is a threshold criterion that must be met in order to recommend a measure for endorsement. All three subcriteria must be met to pass this criterion. See guidance on evidence.

Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. (evaluation criteria)

1c.1 Structure-Process-Outcome Relationship *(Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process- health outcome; intermediate clinical outcome-health outcome)***:**

Process of care linked to important health outcomes

1c.2-3 Type of Evidence *(Check all that apply)***:**

Clinical Practice Guideline

Systematic review of body of evidence (other than within guideline development)

1c.4 Directness of Evidence to the Specified Measure *(State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population)***:**

There is no clinical trial directly linking the care process in this measure with outcomes. However, elicitation of preferences is one important step in the advance care planning process and in matching care with patient goals. The ACOVE expert panel, based on a clinically informed understanding of the medical literature identified this care process important for providing care to seriously ill patients receiving intensive care in the hospital.

Lorenz KA, Rosenfeld K, Wenger N. Quality indicators for palliative and end-of-life care in vulnerable elders. J Am Geriatr Soc. 2007;55 Suppl 2:S318-26.

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)***:**

1c.6 Quality of Body of Evidence (*Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b) directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events):*

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):*

1c.8 Net Benefit *(Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms)***:**

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? No

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias:

1c.11 System Used for Grading the Body of Evidence: Other

1c.12 If other, identify and describe the grading scale with definitions: RCT, non-RCT, cohort or case analysis, multiple time series, textbook, opinion, descriptive study

1c.13 Grade Assigned to the Body of Evidence:

1c.14 Summary of Controversy/Contradictory Evidence:

1c.15 Citations for Evidence other than Guidelines*(Guidelines addressed below)*:

Lorenz KA, Rosenfeld K, Wenger N. Quality indicators for palliative and end-of-life care in vulnerable elders. J Amer Geriatr Soc 2007;55:S318-S326

Walling A, Lorenz KA, Dy SM, et al. Evidence-based recommendations for information and care planning in cancer care. J Clin Oncol 2008;26(23):3896-3902

1c.16 Quote verbatim, the specific guideline recommendation *(Including guideline # and/or page #)*:

1c.17 Clinical Practice Guideline Citation:

1c.18 National Guideline Clearinghouse or other URL: None

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? No

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias:

1c.21 System Used for Grading the Strength of Guideline Recommendation: Other

1c.22 If other, identify and describe the grading scale with definitions: Not graded

1c.23 Grade Assigned to the Recommendation:

1c.24 Rationale for Using this Guideline Over Others:

Based on the NQF descriptions for rating the evidence, what was the <u>developer's assessment</u> of the quantity, quality, and consistency of the body of evidence?

1c.25 Quantity: High **1c.26** Quality: Moderate**1c.27** Consistency**:** High

NATIONAL QUALITY FORUM

Measure Information

This document contains the information submitted by measure developers/stewards, but is organized according to NQF's measure evaluation criteria and process. The item numbers refer to those in the submission form but may be in a slightly different order here. In general, the item numbers also reference the related criteria (e.g., item 1b.1 relates to subcriterion 1b).

Brief Measure Information

NQF #: 1626
De.2. **Measure Title:** Patients Admitted to ICU who Have Care Preferences Documented
Co.1.1. **Measure Steward:** RAND Corporation
De.3. **Brief Description of Measure:** Percentage of vulnerable adults admitted to ICU who survive at least 48 hours who have their care preferences documented within 48 hours OR documentation as to why this was not done.
1b.1. **Developer Rationale:** The aim of this measure is to assist healthcare providers in providing care that is consistent with patient preferences.

S.4. **Numerator Statement:** Patients in the denominator who had their care preferences documented within 48 hours of ICU admission or have documentation of why this was not done.
S.7. **Denominator Statement:** All vulnerable adults admitted to ICU who survive at least 48 hours after ICU admission.
S.10. **Denominator Exclusions:** None

De.1. **Measure Type:** Process
S.23. **Data Source:** Paper Medical Records
S.26. **Level of Analysis:** Facility

IF Endorsement Maintenance – Original Endorsement Date: Feb 14, 2012 **Most Recent Endorsement Date:** Feb 14, 2012

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?**

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
1626_Evidence_MSF5.0_Data.doc,1626_Evidence_MSF5.0_Data-635948602672238493.doc

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., the benefits or improvements in quality envisioned by use of this measure)*
The aim of this measure is to assist healthcare providers in providing care that is consistent with patient preferences.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
N, % measure performance
Assessing Symptoms Side Effects and Indicators of Supportive Treatment (ASSIST): Walling 2013, Inpatients in a national VA sample, N=150, 63.7%

N, % measure performance
Assessing Care of Vulnerable Elders (ACOVE3)(Walling 2010): Inpatient decedents, N=369, 46%

Assessing Symptoms Side Effects and Indicators of Supportive Treatment (ASSIST) (Dy 2010): Inpatient decedents, N=22, 9%

ACOVE (Wenger 2003): Vulnerable elders, N=6, 17%

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.
Walling AM, et al. The Quality of Supportive Cancer Care in the Veterans Affairs Health System and Targets for Improvement. JAMA IM. 2013:173:2071-2079.

Dy SM, Asch SM, Lorenz KA, et al. Quality of end-of-life care for patients with advanced cancer in an academic medical center. J Palliat Med 2011;14(4)"451-457

Walling AM, Asch AM, Lorenz KA, et al. The quality of care provided to hospitalized patients at the end of life. Arch Intern Med 2010;170(12):1057-1063

Wenger NS, Solomon DH, Roth CP, et al. The quality of medical care provided to vulnerable community-dwelling older patients. Ann Intern Med 2003;139(():740-E759

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.
No known information yet available on disparities in care.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF;
  OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Patient/societal consequences of poor quality
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
Many patients would prefer to die rather than live permanently comatose, mechanically ventilated, or tube fed (Pearlman 1993; Wenger 1998), yet physicians and surrogate decision makers often do not know patients´ preferences concerning life-sustaining treatment (Wenger 1998; Guidelines 1987; AMA 1994, Wenger 2000; Kish 2000). Patients entering ICUs are likely to receive invasive care, making the elicitation and documentation of preferences necessary to guide these potentially burdensome treatments. (Lorenz 2007) Care in United States hospitals tends to be aggressive. Even patients with lung and colorectal cancer enrolled in hospice receive aggressive care when brought to the hospital. (Cintron 2003) In a study of Medicare claims that evaluated patients who died within one year of a diagnosis of lung, breast, colorectal or other gastrointestinal cancer, patients receiving chemotherapy within two weeks of death increased from 13.8% in 1993 to 18.5% in 1996, and patients had more hospitalizations, ER visits, and ICU stays during the latter time period. (Earle 2004) Another retrospective study of 335 breast cancer patients who died in the 1990s found that within approximately two months prior to death, 64% continued to receive endocrine therapy and 20% continued to receive chemotherapy. (Asola 2006)

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
AMA Council on Ethical and Judicial Affairs. Code of Medical Ethics: Current Opinions with Annotations. Chicago: American Medical Association, 1994

Asola R, Huhtala H, Holli K. Intensity of diagnostic and treatment activities during the end of life of patients with advanced breast cancer. Breast Cancer Res Treat 2006;100(1):77-82

Cintron A, Hamel MB, Davis RB, et al. Hospitalization of hospice patients with cancer. J Palliat Med 2003;6(%):757-768

Earle CC, Neville BA, Landrum MB, et al. Trends in the aggressiveness of cancer care near the end of life. J Clin Oncol 2004;22(2):315-321

Guidelines on the Termination of Life-Sustaining Treatment and the Care of the Dying. Briarcliff Manor, NY:Hasting Center, 1987

Kish SK, Martin CG, Price KJ. Advance directives in critically ill cancer patients. Crit Care Nurs Clin North Am 2000;12(#):373-383

Lorenz KA, Rosenfeld K, Wenger, N. Quality indicators for palliative and end-of-life care in vulnerable elders. J Am Geriatr Soc 2007;55:S318-S326

Pearlman RA, Cain KC, Patrick DL, et al. Insights pertaining to patient assessments of states worse than death. J

Clin Ethics 1993;4:33-41

Wenger NS, Phillips RS, Teno JM, et al. Physician understanding of patient resuscitation preferences: insights and clinical implications. J Am Geriatr Soc 2000;48(5 Suppl):S44-S51

Wenger NS, Kanouse DE, Lie HH, et al. Preferences for aggressiveness of care among HIV-infected persons and use of advance directives. J Gen Intern Med 1998;13(Suppl 1):93

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

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 subcriteria 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, Pulmonary/Critical Care : Critical Care

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Palliative Care and End of Life Care

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.)*

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.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.

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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Patients in the denominator who had their care preferences documented within 48 hours of ICU admission or have documentation of why this was not done.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
48 hours starting from time of ICU admission

S.6. **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, 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.
Edits indicated by [brackets]
Patients whose medical record includes documentation of care preferences within 48 hours of admission to ICU. Care preferences may include any of the following:
- Code status, preferences for general aggressiveness of care, mechanical ventilation, hemodialysis, transfusion, or permanent feeding tube, OR
- Documentation that a care preference discussion was attempted and/or reason why it was not done

[Simply having an advance directive or other advance care planning document or POLST in the medical record does not satisfy this criterion. However, a notation in the record during the allotted time period referring to preferences or decisions within such a document satisfies this requirement.]

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
All vulnerable adults admitted to ICU who survive at least 48 hours after ICU admission.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk : Individuals with multiple chronic conditions, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
All vulnerable adults admitted to ICU who survive at least 48 hours after ICU admission.
"Vulnerable" is defined as any of the following:
- >74 years of age
- Vulnerable Elder Survey-13 (VES-13) score >2 (Saliba 2001)
- Poor prognosis/terminal illness defined as life expectancy of <6 months
- Stage IV cancer

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
None

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
1. Identify all vulnerable adults admitted to ICU who survive at least 48 hours after ICU admission
2. Examine the medical record for evidence of a statement of patient care preferences OR attempt to elicit these or other reason why this was not done within 48 hours of ICU admission.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
No sampling.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
No survey

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Medical record abstraction tool

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **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.)*

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
1626_MeasureTesting_MSF5.0_Data.doc

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 1626	**NQF Project:** Palliative Care and End-of-Life Care

2. RELIABILITY & 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. **(**evaluation criteria**)** Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate field. Supplemental materials may be referenced or attached in item 2.1. See guidance on measure testing.
2a2. **Reliability Testing.** (*Reliability testing was conducted with appropriate method, scope, and adequate demonstration of reliability.*)
2a2.1 **Data/Sample** *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:** See 2a2.3. 2a2.2 **Analytic Method** *(Describe method of reliability testing & rationale)***:** See 2a2.3. 2a2.3 **Testing Results** *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:** ACOVE3 (Walling 2010) inpatient decedents (n=369) 47 re-abstraction records: Eligibility kappa=0.95; specified care kappa=0.87 ASSIST (Dy 2010, 2011) inpatient decedents (n=22): Overall eligibility kappa=0.87; overall specified care kappa=0.86

Dy SM, Asch AM, Lorenz KA, et al. Quality of end-of-life care for patients with advanced cancer in an academic medical center. J Pall Med 2011;14(4):451-459

Dy SM, Lorenz KA, O´Neill S, et al. Cancer quality-ASSIST supportive oncology quality indicator set. Feasibility, reliability, and validity. Cancer 2010;116:3267-3275

Walling AM, Asch SM, Lorenz KA, et al. The quality of care provided to hospitalized patients at the end of life. Arch Intern Med 2010;170(12):1057-1063

2b. VALIDITY. Validity, Testing, <u>including all Threats to Validity</u>: H☐ M☐ L☐ I ☐

2b1.1 **Describe how the measure specifications** *(measure focus, target population, and exclusions)* **are consistent with the evidence cited in support of the measure focus (***criterion 1c)* **and identify any differences from the evidence:**

See 2b2.2

2b2. Validity Testing. (*Validity testing was conducted with appropriate method, scope, and adequate demonstration of validity.)*

2b2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**

See 2b2.2

2b2.2 Analytic Method *(Describe method of validity testing and rationale; if face validity, describe systematic assessment)***:**

Although validity has not been tested empirically for this measure alone, the process - outcome link of the set of quality measures including this measure has been tested. Process of care measured using the ACOVE quality indicator set is related to two important outcomes in vulnerable elders and persons 75 years and older: mortality and functional status. In 372 vulnerable elders there was a graded positive relationship between quality score and 3-year survival. After adjustment for sex, health status, and health service use, quality score was not associated with mortality for the first 500 days, but a higher quality score was associated with lower mortality after 500 days (hazard ratio, 0.64 [95% CI, 0.49 to 0.84] for a 10% higher quality score).(Higashi 2005) Using an administrative data implementation of a subset of these measures, 21,310 older persons from 19 California counties had their quality of care measured and outcomes followed over the next year. After accounting for number of measures triggered, baseline function and other covariates, better quality was associated with better function at follow-up. Ten percent better quality was associated at follow-up with 0.21 lower ADL need score [95% confidence interval (CI), 0.25-0.17], 0.022 lower IADL need score (95% CI, 0.032-0.013), and lower odds of death (0.91; 95% CI, 0.89 to 0.93).(Zingmond 2011) Validity of the process-outcome link was explicitly evaluated by the ACOVE, ACOVE3, and ASSIST expert panels that reviewed the relevant literature and used a modified Delphi panel of voting on the validity of the measure. (Shekelle 2001; Wenger 2007; Lorenz 2009) Although validity has not been tested empirically for this measure alone, the process-outcome link of the set of quality measures including this measure has been tested. Process of care measured using the ACOVE quality indicator set is linked to patient function and survival. (Higashi 2007)

Higashi T, Shekelle PG, Adams J, et al. Quality of care is associated with survival in vulnerable older patients. Ann Intern Med 2005;143:274-281

Lorenz KA, Dy SM, Naeim A, et al. Quality measures for supportive cancer care: the cancer quality-ASSIST project. J Pain Symptom Manage 2009;37(6):943-964

Shekelle PG, MacLean CH, Morton SC, et al. Assessing care of vulnerable elders: Methods for developing quality indicators. Ann Intern Med 2001;135:647-652

Wenger NW, Roth CP, Shekelle P, et al. Introduction to the assessing care of vulnerable elders-3 quality indicator measurement set. J Am Geriatr Soc 2007;55:S247-S252

Zingmond DS, Ettner SL, Wilber KH, et al. Association of claims-based quality of care measures with outcomes among community-dwelling vulnerable elders. Med Care 2011;49:553-559

2b2.3 Testing Results *(Statistical results, assessment of adequacy in the context of norms for the test conducted; if face validity, describe results of systematic assessment)***:**

Face validity was tested in the panels described in 2b2.2 above as well as the strength of the process–outcome link.

POTENTIAL THREATS TO VALIDITY. (*All potential threats to validity were appropriately tested with adequate results.*)

2b3. Measure Exclusions. (*Exclusions were supported by the clinical evidence in 1c or appropriately tested with results demonstrating the need to specify them.*)

2b3.1 Data/Sample for analysis of exclusions *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**

None

2b3.2 Analytic Method (*Describe type of analysis and rationale for examining exclusions, including exclusion related to patient preference)***:**

2b3.3 Results *(Provide statistical results for analysis of exclusions, e.g., frequency, variability, sensitivity analyses):*

2b4. Risk Adjustment Strategy. (*For outcome measures, adjustment for differences in case mix (severity) across measured entities was appropriately tested with adequate results.*)

2b4.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**

None

2b4.2 Analytic Method (*Describe methods and rationale for development and testing of risk model or risk stratification including selection of factors/variables)***:**

2b4.3 Testing Results *(Statistical risk model: Provide quantitative assessment of relative contribution of model risk factors; risk model performance metrics including cross-validation discrimination and calibration statistics, calibration curve and risk decile plot, and assessment of adequacy in the context of norms for risk models. Risk stratification: Provide quantitative assessment of relationship of risk factors to the outcome and differences in outcomes among the strata)***:**

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment:

2b5. Identification of Meaningful Differences in Performance. (*The performance measure scores were appropriately analyzed and discriminated meaningful differences in quality.*)

2b5.1 **Data/Sample** *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**

The awareness of patient preferences is vital to facilitate matching end-of-life care with that which the patient would want. Failure to attempt to elicit patient preferences, if unknown, when a patient is in ICU is significant. As noted in 1b2., performance was low for this measure (9-46%).

2b5.2 **Analytic Method** *(Describe methods and rationale to identify statistically significant and practically/meaningfully differences in performance)***:**

2b5.3 **Results** *(Provide measure performance results/scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:

2b6. Comparability of Multiple Data Sources/Methods. (*If specified for more than one data source, the various approaches result in comparable scores.*)

2b6.1 **Data/Sample** *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**

None

2b6.2 **Analytic Method** *(Describe methods and rationale for testing comparability of scores produced by the different data sources specified in the measure)***:**

2b6.3 **Testing Results** *(Provide statistical results, e.g., correlation statistics, comparison of rankings; assessment of adequacy in the context of norms for the test conducted)***:**

2c. Disparities in Care: H☐ M☐ L☐ I ☐ NA☐ (*If applicable, the measure specifications allow identification of disparities.*)

2c.1 **If measure is stratified for disparities, provide stratified results** *(Scores by stratified categories/cohorts)***:**
N/A

2c.2 **If disparities have been reported/identified (e.g., in 1b), but measure is not specified to detect disparities, please explain:**

2.1-2.3 Supplemental Testing Methodology Information:
Steering Committee: Overall, was the criterion, *Scientific Acceptability of Measure Properties*, met? **(*Reliability and Validity must be rated moderate or high*) Yes☐ No☐** **Provide rationale based on specific subcriteria:**
If the Committee votes No, STOP

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.**

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

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)*

No 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.**

While some EHRs could provide information about the presence of an advance directive in the record, most preference information and discussions by their nature, do not lend themselves to electronic data capture. This is true for other aspects of geriatric care as well. (MacLean 2006) However, the data elements are discrete and could be delineated in an EHR.

MacLean GH, Louie R, Shekelle PG, et al. Comparison of administrative data and medical records to measure quality of medical care provided to vulnerable older patients. Med Care 2006;44(2):141-148

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.**

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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

As described avoe, we have found that this measure can be reliably abstracted by different groups.

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)*.

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.

Planned	Current Use (for current use provide URL)
Quality Improvement (Internal to the specific organization)	

4a.1. **For each CURRENT use, checked above, provide:**

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

4a.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?*)

Still in planning phase--programs that are planning to use these measures are in their early phases

This measure is referenced in the VA´s draft handbook for planned Life Sustaining Treatment policy in section related to ICU admissions. It is not being used to mandate a time frame for conversations, but the metric is referenced as a quality standard that clinicians should consider adopting.

4a.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.*)

This measure is referenced in the VA´s draft handbook for planned Life Sustaining Treatment policy in section related to ICU admissions. It is not being used to mandate a time frame for conversations, but the metric is referenced as a quality standard that clinicians should consider adopting.This measure is referenced in the VA´s draft handbook for planned Life Sustaining Treatment policy in section related to ICU admissions. It is not being used to mandate a time frame for conversations, but the metric is referenced as a quality standard that clinicians should consider adopting.

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

Not available

4b.2. **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.**
As organizations strive to improve advance care planning to match patient preferences with care received, measures such as this will be needed to measure their success.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
Documentation of patient preferences or an attempt to elicit them is not a care process that is likely to produce unintended consequences. In the validity testing noted, we are not aware of unintended consequences.

I repeat information here regarding what is known about the link between process and outcomes:

The validity of ACOVE measures is based in the development methodology, which includes linking of measures to the medical literature and then use of an expert panel process to develop valid measures of quality of care. The link between process and outcome is extremely difficult to carry out at the individual measure level, (Parast L, et al. Challenges in assessing the process-outcome link in practice. J Gen Intern Med. 2015;30:359-64) however, the full set of ACOVE measures has been tested in aggregate to evaluate the link between process and outcome of care. This link has been demonstrated among a population of Medicare Advantage patients in which better process of care was linked to lower mortality (Higashi T, et al. Quality of care is associated with survival in vulnerable older patients. Ann Intern Med. 2005;143:274-81) and in a nursing home population in which better process of care was linked to less functional decline (Zingmond DS, et al. Association of claims-based quality of care measures with outcomes among community-dwelling vulnerable elders. Med Care. 2011;49:553-9). We are not aware of any explicit studies of the validity of this measure other than the development mechanism and reliability data.

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

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 completely harmonized?
Yes

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.)
This measure was part of the National Palliative Care Research Center (NPCRC) Key Palliative Measures Bundle during the original submission. At that time, a NPCRC cover letter and table of bundle measures for description of the selection and harmonization of the Key Palliative Measures Bundle was provided.

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.
Attachment:

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** RAND Corporation
Co.2 **Point of Contact:** Carol, Roth, roth@rand.org, 310-393-0411-6425
Co.3 **Measure Developer if different from Measure Steward:** RAND Corporation
Co.4 **Point of Contact:** Neil, Wenger, nwenger@mednet.ucla.edu, 310-794-2288-

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.
ACOVE-3 project expert panel members, ACOVE-3 Clinical Committee members, ASSIST project expert panel members and Advisory Board as listed below.

ACOVE-3 project (Panel 2) expert panel members:

Helena Chang, MD
UCLA School of Medicine, Los Angeles, CA

Nick Fitterman, MD
Northshore Medical Group, Huntington, NY

Jean S. Kutner, MD, MSPH
University of Colorado Health Sciences Center, Aurora, CO

Patrick J. Loehrer, Sr., MD
Indiana University School of Medicine, Indianapolis, IN

Thomas Mattimore, MD
University of California at Los Angeles, Los Angeles, CA

Hyman B. Muss, MD
Vermont Cancer Center at University of Vermont, Burlington, VT

James L. Naughton, MD
Alliance Medical Group, Pinole, CA

Cheryl Phillips, MD
Sutter Medical Group, Sacramento, CA

Doron Schneider, MD
Muller Center for Senior Health, Abington Memorial Hospital, Abington, PA

Michael Stamos, MD
University of California, Irvine, CA

Ronald D. Stock, MD
Center for Senior Health, Eugene, OR

May Lin Tao, MD, MSPH
John Wayne Cancer Institute, Saint John´s Health Center, Santa Monica, CA and Valley Radiotherapy Associates Medical Group, El Segundo, CA

Role of ACOVE Expert Panel: Expanded and updated the Assessing Care of Vulnerable Elders (ACOVE) quality indicators via literature review, face-to-face discussion, and 2 rounds of anonymous ratings to evaluate whether the QIs were valid measures of quality of care using a process that is an explicit combination of scientific evidence and professional consensus.

ACOVE-3 CLINICAL COMMITTEE MEMBERS:

Alpesh N. Amin, MD - Hospitalist
University of California, Irvine Medical Center, Irvine, CA

Richard W. Besdine, MD - Geriatrician and Clinical Committee Chair
Brown University Center for Gerontology and Health Care Research, Providence, RI

Dan G. Blazer, MD - Geriatric Psychiatrist
Duke University Medical Center, Durham, NC

Harvey J. Cohen, MD - Geriatric Oncologist
Duke University Medical Center, Durham, NC

Terry Fulmer, PhD, RN, FAAN - Nurse
New York University, New York, NY

Patricia A. Ganz, MD - Oncologist
UCLA Schools of Medicine & Public Health, Jonsson Comprehensive Cancer Center, Los Angeles, CA

Mark A. Grunwald, MD - Family Practitioner
Gunderson Lutheran Clinic, Prairie du Chien, WI

William J. Hall, MD, MACP - Geriatrician
Highland Hospital, Rochester, NY

Ira R. Katz, MD, PhD - Psychiatrist
University of Pennsylvania, Philadelphia, PA

Paul R. Katz, MD - Geriatrician
Monroe Community Hospital, Rochester, NY

Dalane W. Kitzman, MD - Geriatric Cardiologist
Wake Forest University School of Medicine, Winston-Salem, NC

Rosanne M. Leipzig, MD, PhD - Geriatrician
Mount Sinai School of Medicine, New York, NY

Ronnie A. Rosenthal, MD - Surgeon
Yale University School of Medicine, New Haven, CT

Role of ACOVE-3 Clinical Committee: Evaluated the coherence of the complete set of QIs that the experts rated as valid as well as determined exclusions for advanced dementia and poor prognosis.

ASSIST project expert panel members:
Kurt Kroenke, MD
Indiana University Cancer Center, Indianapolis, Indiana

Terry Altilio, LCSW
Beth Israel Medical Center, New York, New York

Lodovico Balducci, MD
H. Lee Moffitt Cancer Center & Research Institute, Tampa, Florida

Jeannine M. Brant PhD(c),
St. Vincent Healthcare, Billings, Montana

Eduardo Bruera, MD
UT M. D. Anderson Cancer Center, Houston, Texas

Peter Eisenberg, MD
California Cancer Care, Greenbrae, California

Pr Stein Kaasa
St. Olavs University Hospital HF, Trondheim, Norway

Sean Morrison, MD
Mt. Sinai Medical School, New York, New York

Mary Simmonds, MD
Family practice, New Cumberland, Pennsylvania

Role of ASSIST Expert Panel: Helped to develop and refine the quality indicators for the Addressing Symptoms Side effects and Indicators for Supportive Treatment (ASSIST) project via literature review, face-to-face discussion, and 2 rounds of anonymous ratings to evaluate whether the QIs were valid measures of quality of care using a process that is an explicit combination of scientific evidence and professional consensus.

ASSIST Project Advisory Board:

Neil S. Wenger, MD, MPH
UCLA Division of Gen Internal Med and Health Svcs Research, Los Angeles, CA

Steven B. Clauser, PhD
Chief, Outcomes Research Branch, Applied Research Program, Div of Cancer Control and Pop. Sciences, National Cancer Institute, Bethesda, MD

David Currow, MD
CEO, Cancer Australia, Flinders University, South Australia

Molla S. Donaldson, Dr.PH, MS
Adjunct Professor, Dept. of Medicine, George Washington University School of Medicine and Health Sciences and Principal, QuantaNet, Chevy Chase, MD

Betty Ferrell, PhD, RN, FAAN
City of Hope National Medical Center, Duarte, CA

Michael T. Halpern, MD, PhD
Strategic Director, Health Svcs Research, American Cancer Society, Atlanta, GA

Laura C. Hanson, MD, MPH
Division of Geriatric Medicine, University of North Carolina School of Medicine, Chapel Hill, NC

Catherine D. Harvey, Dr.PH, RN, AOCN
Principal, The Oncology Group, LLC, Raleigh, NC

Jorn Herrstedt, MD

Copenhagen University Hospital Department of Oncology, Herlev, Denmark

Paul Hesketh, MD
Chief, Division of Hematology/Oncology, Caritas St. Elizabeth´s Medical Center, Boston, MA

Catherine H. MacLean, MD, PhD
Medical Director, Programs for Clinical Excellence Health Solutions, Wellpoint, Inc., Thousand Oaks, CA

Thomas J. Smith, MD
Division of Hematology/Oncology and Palliative Care, Virginia Commonwealth University, Massey Cancer Center, Richmond, VA

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2001
Ad.3 **Month and Year of most recent revision:** 07, 2010
Ad.4 **What is your frequency for review/update of this measure?** Every 3 years
Ad.5 **When is the next scheduled review/update for this measure?**

Ad.6 **Copyright statement:**
Ad.7 **Disclaimers:**

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

Appendix H - Related and Competing Measures for Preferences

Preferences

	1626 – Care preferences	1641- Treatment preferences	0326 – Advance Care Plan
Type	Process	Process	Process
Setting	Hospital	Hospice/Hospital	All settings
Level of analysis	Facility	Clinician group/practice Facility	Clinician – individual & group/practice
Numerator	Documentation of care preferences w/n 48 hrs or documentation of why not done	Documentation of preferences for life sustaining treatments	Documented ACP or surrogate or documentation of discussion
Denominator	Vulnerable adults in ICU who survive ≥48 hours	In hospice or receiving specialty PC	All patients ≥ 65
Exclusions	N/A	<1 day in hospital or hospice	N/A
Data source	Paper medical records	EHR; electronic clinical data (HIS)	Claims, electronic clinical data

Appendix I – Measurement Framework for Palliative and End-of-Life Care