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

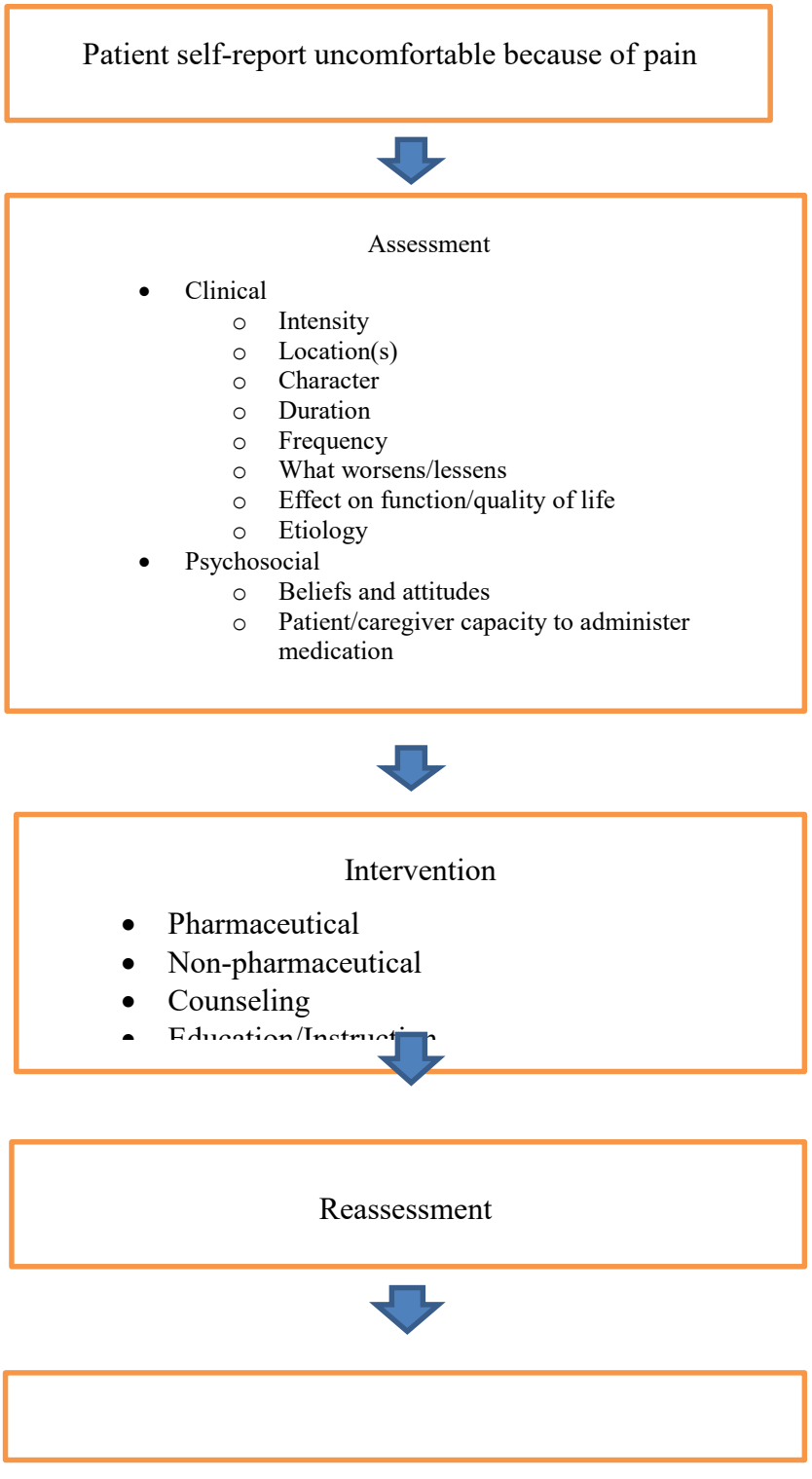
☐ Structure: Click here to name the structure

☐ 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.



Additional intervention if needed based on

Patient self-report that comfort achieved

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) <u>across studies</u> 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.

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 <u>performance measure score</u> 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:**

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 #: 2651
Measure Title: CAHPS® Hospice Survey (experience with care)
Measure Steward: Centers for Medicare and Medicaid Services
Brief Description of Measure: The measures submitted here are derived from the CAHPS® Hospice Survey, which is a 47-item standardized questionnaire and data collection methodology. The survey is intended to measure the experiences of hospice patients and their primary caregivers.
The measures proposed here include the following six multi-item measures.

- Hospice Team Communication
- Getting Timely Care
- Treating Family Member with Respect
- Getting Emotional and Religious Support
- Getting Help for Symptoms
- Getting Hospice Training

In addition, there are two other measures, also called "global ratings."

- Rating of the hospice care
- Willingness to recommend the hospice

Below we list each multi-item measure and its constituent items, along with the two ratings questions. Then we briefly provide some general background information about CAHPS surveys.

List of CAHPS Hospice Survey Measures

Multi-Item Measures
Hospice Team Communication (Composed of 6 items)
+ While your family member was in hospice care, how often did the hospice team keep you informed about when they would arrive to care for your family member?
+ While your family member was in hospice care, how often did the hospice team explain things in a way that was easy to understand?
+ How often did the hospice team listen carefully to you when you talked with them about problems with your family member's hospice care?
+ While your family member was in hospice care, how often did the hospice team keep you informed about your family member's condition?
+ While your family member was in hospice care, how often did the hospice team listen carefully to you?
+ While your family member was in hospice care, how often did anyone from the hospice team give you confusing or contradictory information about your family member's condition or care?
Getting Timely Care (Composed of 2 items)
+ While your family member was in hospice care, when you or your family member asked for help from the hospice team, how often did you get help as soon as you needed it?
+ How often did you get the help you needed from the hospice team during evenings, weekends, or holidays?
Treating Family Member with Respect (Composed of 2 items)
+ While your family member was in hospice care, how often did the hospice team treat your family member with dignity and respect?
+ While your family member was in hospice care, how often did you feel that the hospice team really cared about your family member?

Providing Emotional Support (Composed of 3 items)

+ While your family member was in hospice care, how much emotional support did you get from the hospice team?
+ In the weeks after your family member died, how much emotional support did you get from the hospice team?
+ Support for religious or spiritual beliefs includes talking, praying, quiet time, or other ways of meeting your religious or spiritual needs. While your family member was in hospice care, how much support for your religious and spiritual beliefs did you get from the hospice team?

Getting Help for Symptoms (Composed of 4 items)

+ Did your family member get as much help with pain as he or she needed?
+ How often did your family member get the help he or she needed for trouble breathing?
+ How often did your family member get the help he or she needed for trouble with constipation?
+ How often did your family member receive the help he or she needed from the hospice team for feelings of anxiety or sadness?

Getting Hospice Care Training (Composed of 5 items)

+ Did the hospice team give you enough training about what side effects to watch for from pain medicine?
+ Did the hospice team give you the training you needed about if and when to give more pain medicine to your family member?
+ Did the hospice team give you the training you needed about how to help your family member if he or she had trouble breathing?
+ Did the hospice team give you the training you needed about what to do if your family member became restless or agitated?
+ Side effects of pain medicine include things like sleepiness. Did any member of the hospice team discuss side effects of pain medicine with your or your family member?

Rating Measures:

In addition to the multi-item measures, there are two "global" ratings measures. These single-item measures indicate on the one hand the need for quality improvement and on the other hand provide families and patients looking for care with evaluations of the care provided by the hospice. The items are rating of hospice care and willingness to recommend the hospice.

+ Rating of Hospice Care: Using any number from 0 to 10, where 0 is the worst hospice care possible and 10 is the best hospice care possible, what number would you use to rate your family member's hospice care?
+ Willingness to Recommend Hospice: Would you recommend this hospice to your friends and family?

The CAHPS Hospice Survey is a standardized survey instrument designed to collect reports and ratings of experiences with hospice care. The survey is completed by the primary caregiver of the patient who died while receiving hospice care (hereafter, "decedent"). The primary caregiver is intended to be the family member or friend most knowledgeable about the decedent's hospice care, and is identified through hospice administrative records. Data collection for sampled decedents/caregivers is initiated two months following the month of the decedent's death.

The CAHPS Hospice Survey is part of the CAHPS family of experience of care surveys and is available in the public domain at https://cahps.ahrq.gov/surveys-guidance/hospice/index.html. CMS initiated national implementation of the CAHPS Hospice Survey in 2015. Hospices meeting CMS eligibility criteria were required to administer the survey for a "dry run" for at least one month of sample from the first quarter of 2015. Beginning with the second quarter of 2015, hospices are required to participate on an ongoing monthly basis in order to receive their full Annual Payment Update from CMS. Information regarding survey content and national implementation requirements, including the latest versions of the survey instrument and standardized protocols for data collection and submission, are available at: http://www.hospicecahpssurvey.org/.

A list of the CAHPS Hospice Survey measures, including the components of the multi-item measures can be found in Appendix A.

Developer Rationale: Each year 2.4 million Americans die. Too often, this dying experience is marred by untreated pain or other symptoms, lack of shared decision making, and insufficient emotional support (Teno Clarridge et al. 2004; Teno, Freedman et al. 2015).The Medicare hospice care benefit offers palliative care for individuals with a prognosis of living 6 months or less if their terminal illness runs its normal course. The number of Medicare beneficiaries receiving hospice services has grown from 513,000 in FY 2000 to over 1.3 million in FY 2013. Similarly, Medicare hospice expenditures have risen from $2.9 billion in FY 2000 to an estimated $15.1 billion in FY 2013. The CMS Office of the Actuary projects that hospice expenditures are expected to continue to increase, by approximately 8 percent annually, reflecting an increase in the number of Medicare beneficiaries, more beneficiary awareness of the Medicare Hospice Benefit for end-of-life care, and a growing preference for care provided in home and community-based settings. (More details may be found in the Federal Register. (Please see: 80 FR 47142.)

Previous research has shown striking variation across hospices with regard to their provision of key hospice care processes. For

example, visits by professional hospice staff in the last two days of life vary considerably across hospices, with 3.2% of hospices at the 25th percentile providing such visits and 19.5% at the 75th percentile providing them. A 2013 Office of Inspector General report found that 27% of the hospice providers did not provide any general inpatient (GIP) level of care (intensive inpatient care that is meant to be short-term; Office of the Inspector General 2013). The provision of GIP care varies by region of the country with 77% of hospice providers in the South providing GIP level of care compared to 91% of providers in New England (Plotzke, Christian et al., 2014). Previous analyses of the Family Evaluation of Hospice Care (FEHC) voluntary repository maintained by the National Hospice and Palliative Care Organization have found substantial variation in several of the key domains measured by the survey (e.g., 12.6% of respondents reporting unmet needs for emotional and spiritual support for hospices in the 25th percentile of hospices compared to 21.4% of respondents in the 75th percentile; Connor, Teno et al. 2005). Evidence of variation in hospice care processes and patient and family experiences of hospice care point to the need for performance measures that can be systematically implemented to monitor hospice performance and make fair comparisons across hospices.

To that end, the Affordable Care Act mandated that the US Department of Health and Human Services (DHHS) implement a quality reporting program for hospices. The CAHPS Hospice Survey is a component of that program. The conceptual model for the development of the survey was grounded in review of existing surveys (Lendon, Ahluwalia et al., 2015), a previous review of the existing guidelines regarding quality of end of life (Teno, Casey et al., 2001), National Qualify Forum Preferred Practices in Palliative Care, and the work of the National Consensus Project for Quality Palliative Care (http://www.nationalconsensusproject.org/Guidelines_Download2.aspx).

The CAHPS Hospice Survey is part of a family of CAHPS surveys that focuses on patient or caregiver experiences with health care. The CAHPS Hospice Survey presents primary caregivers (as identified by hospices) with a set of standardized questions about the hospice care they and their family member or friend received. The survey and associated methodology is intended to create reliable and valid data about hospice experiences that can be published on the web to help families choose hospices for their family members or friends, and to create incentives for quality improvement among hospice providers.

Citations:

Connor SR, Teno J, Spence C, Smith N. Family evaluation of hospice care: results from voluntary submission of data via website. J Pain Symptom Manage 2005;30:9-17.
Lendon JP, Ahluwalia SC, Walling AM, et al. Measuring Experience With End-of-Life Care: A Systematic Literature Review. J Pain Symptom Manage 2015;49:904-15.e1-3.
Department of Health and Human Services Office of the Inspector General. Medicare Hospice: Use of General Inpatient Care. Washington, DC. 2013.
Plotzke M, Christian TJ, Pozniak A, et al. Medicare Hospice Payment Reform: Analyses to Support Payment Reform. 2014 May 1.
Rhodes RL, Mitchell SL, Miller SC, Connor SR, Teno JM. Bereaved family members´ evaluation of hospice care: what factors influence overall satisfaction with services? J Pain Symptom Manage 2008;35:365-71.
Teno JM, Casey VA, Welch L, Edgman-Levitan S. Patient-Focused, Family-Centered End-of-Life Medical Care: Views of the Guidelines and Bereaved Family Members. J Pain Symptom Manage-Special Section on Measuring Quality of Care at Life´s End II 2001;22:738-51.
Teno JM, Clarridge BR, Casey V, et al. Family perspectives on end-of-life care at the last place of care. JAMA 2004;291:88-93.
Teno JM, Freedman VA, Kasper JD, Gozalo P, Mor V. Is Care for the Dying Improving in the United States? J Palliat Med 2015;18:662-6.

Numerator Statement: CMS calculates CAHPS Hospice Survey measures using top-box scoring. The top-box score refers to the percentage of caregiver respondents that give the most positive response. Details regarding the definition of most positive response are noted in Section S.6 below.

Denominator Statement: The measure's denominator is the number of survey respondents who answered the item. The target population for the survey is primary caregivers of hospice decedents. The survey uses screener questions to identify respondents eligible to respond to subsequent items. Therefore, denominators will vary by survey item (and corresponding multi-item measures, if applicable) according to the eligibility of respondents for each item.

Denominator Exclusions:

Cases are excluded from the survey target population if:

- The hospice patient is still alive
- The decedent's age at death was less than 18
- The decedent died within 48 hours of his/her last admission to hospice care
- The decedent had no caregiver of record

- The decedent had a caregiver of record, but the caregiver does not have a U.S. or U.S. Territory home address
- The decedent had no caregiver other than a nonfamilial legal guardian
- The decedent or caregiver requested that they not be contacted (i.e., by signing a no publicity request while under the care of hospice or otherwise directly requesting not to be contacted)
- The caregiver is institutionalized, has mental/physical incapacity, has a language barrier, or is deceased
- The caregiver reports on the survey that he or she "never" oversaw or took part in decedent's hospice care

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

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

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

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. 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. Summary of evidence:

This submission contains information for 8 Patient-Reported Outcome based Performance Measures (PRO-PMs) that are calculated from data aggregated from responses to the Hospice CAHPS survey. These 8 PRO-PMs include:

1) Average proportion of respondents who gave the most positive responses on the ***hospice team communications*** items
2) Average proportion of respondents who gave the most positive responses on the ***timely care*** items
3) Average proportion of respondents who gave the most positive responses on the ***respect*** items
4) Average proportion of respondents who gave the most positive responses on the ***emotional and religious support*** items
5) Average proportion of respondents who gave the most positive responses on the ***symptom*** items
6) Average proportion of respondents who gave the most positive responses on the ***training*** items
7) Average proportion of respondents who gave the most positive responses on the ***rating of care*** item
8) Average proportion of respondents who gave the most positive response on the ***willingness to recommend the hospice*** item

Evidence Summary

- The developer identifies several structures and processes of care that can impact the first six measures. Although not stated explicitly, these activities likely also would affect overall ratings of the care provided and willingness to recommend the hospice.
- To assess topics of interest important to caregivers of hospice patients, the developer conducted focus group and individual interviews with family members of hospice decedents. Results from these activities suggested that caregivers find communication, information, and respect to be important facets of high-quality hospice care.

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

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*
- *Does the Committee agree that hospice patients value queries about the various domains included in the Hospice CAHPS survey?*

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.

- Current performance data are calculated from Quarter 2, 2015 data submitted to the Hospice CAHPS Data Warehouse by 2,512 hospice agencies serving at least 50 patients annually (approximately 70% of the survey-eligible hospice agencies). Performance data on the individual items used for the various measures are included in the supplementary materials.

Measures	Mean	Standard Deviation	25th Percentile	50th Percentile	75th Percentile
Hospice Team Communication	82.0	9.0	77.0	82.8	87.7
Getting Timely Care	77.4	12.1	70.7	78.0	85.3
Treating Family Member with Respect	91.4	7.0	88.0	92.1	96.3
Getting Emotional and Religious Support	91.8	6.5	88.8	92.9	96.4
Getting Help for Symptoms	75.7	11.9	69.3	76.6	83.2
Getting Hospice Care Training	72.7	12.8	65.5	73.2	81.0
Rating of Hospice	84.0	11.0	78.0	85.1	91.4
Willingness to Recommend	84.9	11.4	79.3	86.2	92.7

Disparities

- The developer summarizes Q1-Q2 2015 data, noting potential racial and ethnic disparities in the experience of care measures, although the "direction" of disparities varies across the measures (e.g., caregivers of minority decedents report better experience for some measures, worse for others, and similar for others, as compared to caregivers of white decedents).
- The developer also notes disparities in experiences of care across racial/ethnic groups, based on data from the Family Evaluation of Hospice Care survey.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Are you aware of evidence that other disparities exist in these areas of healthcare?*

Preliminary ratings for opportunity for improvement:

1 Hospice team communications:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**
2) Getting Timely Care:	☒ **High**	☐ **Moderate**	☐ **Low**	☐ **Insufficient**
3) Treating Family Member with Respect:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**
4) Getting Emotional and Religious Support:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**
5) Getting Help for Symptoms:	☒ **High**	☐ **Moderate**	☐ **Low**	☐ **Insufficient**
6) Getting Hospice Care Training:	☒ **High**	☐ **Moderate**	☐ **Low**	☐ **Insufficient**
7) Rating of Hospice:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**

8) Willingness to Recommend: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

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

1a

*This 47-item process measure aims to evaluate the experiences of hospice patients and their primary caregivers. The measure applies directly to the experiences as gathered through individual interviews and focus groups. The relationship between the outcome and at least one healthcare action is identified and supported.
*Pass
* Yes, the relationship is identified and the rationale clearly stated and supported. The measures are direct indicators of perceived quality of care.
* Evidence applies directly and data relate to desired outcome.

1b

*yes, and there is clear variation as well as disparity data
*Performance data was provided and it demonstrates moderate to high ratings demonstrating opportunities for improvement. Disparities are noted--care experiences vary by racial group both positively and negatively.
* Yes, performance data was provided though limited to the second quarter of data collection due to the short time frame of utilization of these measures. The data is large and seems comprehensive in terms of the scope and number of hospices that submitted data. Disparities regarding race and/or ethnicity admittedly require further ongoing analysis.
* Gap demonstrated areas for improvement. Limited disparities data provided.

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 caregivers of patients who died while under hospice care

Specifications:

- The measure is specified at the facility level of analysis, for use in the hospice setting. A higher score indicates better quality.
- Six of the PRO-PMs are based on multiple questions (or items) in the survey; two of the PRO-PMs (the "global" measures) are based on one question (or item) in the survey.
- Measure results are calculated based on "top-box scoring". For example, for items with choices of "Never/Sometimes/Usually/Always", the top box score is the number of respondents who answer "Always." (NOTE: this method is consistent across all CAHPS-derived PRO-PMs.)
- The target population for the Hospice CAHPS survey is the primary caregiver of hospice decedents. However, because not all questions are applicable to each decedent, each item has a denominator (i.e., the number of respondents who answered that item).
- Exclusions to the denominator include:
    - The hospice patient is still alive
    - The decedent's age at death was less than 18
    - The decedent died within 48 hours of his/her last admission to hospice care
    - The decedent had no caregiver of record
    - The decedent had a caregiver of record, but the caregiver does not have a U.S. or U.S. Territory home address

- The decedent had no caregiver other than a non-familial legal guardian
- The decedent or caregiver requested that they not be contacted
- The caregiver is institutionalized, has mental/physical incapacity, has a language barrier, or is deceased
- The caregiver reports on the survey that he or she "never" oversaw or took part in decedent's hospice care

- Multi-item measure scores for each respondent are calculated as the average proportion of respondents that gave responses in the most positive category(ies) across the items in the multi-item measure. A calculation algorithm is provided, as is an example of how to calculate the measure.
- The measure is adjusted for mode of administration (mail only, telephone only, or mixed (mail plus telephone follow up))and case mix. The parameters for case-mix adjustment are provided as part of the submission materials, although the parameters for the mode adjustment are not.
- Sampling is allowed, depending on the number of eligible decedents in the measurement year. The sampling methods are well-defined.
- Data collection is initiated two months following the month of the decedent's death.
- Hospices with fewer than 50 decedents per year are not required to field the survey.
- The survey must be conducted by a survey vendor unless the hospice has fewer than 50 decedents per year.

Questions for the Committee :

- *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.

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:

- Data used in various analyses included information on caregiver experience of care for those who died between January – June (Q1-Q2) while receiving hospice care. Data for an agency was included in the analysis if it had at least 5 responses in the quarter. This included data on:
  - 2,267 agencies with at least 10 respondents in Q1-Q2 2015; characteristics of these agencies and the patients represented are provided
  - 2,512 agencies with at least 5 respondents in Q2 2015; these were the data used for reliability testing
- Data element reliability was calculated by examining 1) the internal consistency of the multi-item measures using Cronbach's alpha and 2) the item-total correlation using Pearson's correlation
- 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.
- All of these are appropriate methods for testing reliability.

Results of reliability testing:

	Data element		Score-level	
Measures	**Cronbach's alpha**	**Item-total correlations (range)**	**ICC**	**Estimated reliability for n=200**

Hospice Team Communication	0.84	0.32-0.82	.013	0.72
Getting Timely Care	0.60	0.44	.012	0.71
Treating Family Member with Respect	0.69	0.59	.008	0.61
Getting Emotional and Religious Support	0.66	0.46-0.54	.011	0.70
Getting Help for Symptoms	0.74	0.52-0.62	.008	0.62
Getting Hospice Care Training	0.86	0.65-0.76	.017	0.78
Rating of Hospice	n/a	n/a	.011	0.68
Willingness to Recommend	n/a	n/a	.017	0.78

- For Cronbach's alpha, .70 or higher is a widely-accepted rule of thumb for a set of items to be considered a scale
- For Pearson correlations, Cohen guidelines for the correlation effect size in the social sciences are: 0.10-0.23—small; 0.24-0.36—medium; 0.37 or larger—large
- For estimated reliability based on ICCs, .70 often is regarded as a minimum acceptable value
- Based on the above rules of thumb for Cronbach's alpha and item-total correlations, the Hospice CAHPS instrument is reliable (although the items on timely care have relatively lower internal consistency), as are the PRO-PMs that are calculated from the Hospice CAHPS survey data.

Guidance from the Reliability Algorithm:
Precise specifications (Box 1) → Empirical testing conducted with measure as specified (Box 2) → Score-level testing conducted (Box 4) → Method of testing appropriate (Box 5) → Moderate certainty that the scores are reliable for 6 measures; lower certainty for 2 measures, although reliability will likely be higher if number of respondents is higher (than 200).

Questions for the Committee:
- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Preliminary ratings for reliability:

1 Hospice team communications:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**
2) Getting Timely Care:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**
3) Treating Family Member with Respect:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**
4) Getting Emotional and Religious Support:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**
5) Getting Help for Symptoms:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**
6) Getting Hospice Care Training:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**
7) Rating of Hospice:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**
8) Willingness to Recommend:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**

2b. Validity

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?*

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.

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:

- Developers conducted construct validation analyses of the measure scores in two ways, both of which are appropriate methods of score-level validation:
  - First, they examining the relationship of the agency-level results from the 6 multi-items measures to the agency-level results of the global rating measure and to the willingness to recommend measure via ~~a~~ linear regression analysis. The developers hypothesized that the measures would be related but did not suggest a direction (although it is assumed positive) nor a magnitude. They used Q2 2015 for these analyses.
  - Second, they examined Pearson correlations between the agency-level multi-item measures to assess the magnitude of association between them. They noted that moderate inter-correlations were expected because the PRO-PMs assess experience of care, although very low inter-correlations could be interpreted as demonstration of divergent validity.

Validity testing results:

- Method 1:

Association with global rating measure

	β (Standard Error)
Hospice Team Communication	0.853 (0.005)*
Treating Family Member with Respect	0.812 (0.006)*
Getting Emotional and Religious Support	0.746 (0.007)*
Getting Help for Symptoms	0.609 (0.006)*
Getting Timely Care	0.517 (0.004)*
Getting Hospice Care Training	0.472 (0.004)*

*p<.001

Association with willingness to recommend measure

	β (Standard Error)
Hospice Team Communication	0.781 (0.005)*
Treating Family Member with Respect	0.776 (0.006)*
Getting Emotional and Religious Support	0.707 (0.008)*
Getting Help for Symptoms	0.539 (0.006)*
Getting Timely Care	0.464 (0.004)*
Getting Hospice Care Training	0.432 (0.004)*

*p<.001

  - These results indicate that each of the 6 multi-item PRO-PMs are positively and statistically significantly associated with the global hospice rating PRO-PM and the willingness to recommend PRO-PM. Results confirmed the developer's hypotheses.
  - ~~No testing for the global Willingness to Recommend PRO-PM was conducted.~~

- Method 2

	Hospice Team	**Treating**	**Getting**	**Getting**	**Getting**	**Getting**

	Communication	Family Member with Respect	Hospice Care Training	Timely Care	Emotional and Religious Support	Help for Symptoms
Hospice Team Communication	1					
Treating Family Member with Respect	.68	1				
Getting Hospice Care Training	.55	.39	1			
Getting Timely Care	.61	.52	.36	1		
Getting Emotional and Religious Support	.47	.45	.31	.35	1	
Getting Help for Symptoms	.56	.48	.47	.46	.37	1

Note: All two-tailed p-values < .001.

- The correlations found in these analyses are, for the most part, large, according to Cohen's guidelines (cited above), although the developer interprets them as moderate and infers that the measures reflect "*unique but related constructs*".
- ~~No testing for the global Willingness to Recommend PRO-PM was conducted.~~

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *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:

- In this section, the developer indicates that there are no exclusions for these 8 PRO-PMs; however, this does not match what is stated in item S.10.

Questions for the Committee:

- *Are the exclusions noted in section S.10 truly exclusions to the measure denominator, or are they persons who are not eligible for the survey?*
- *Are any patients or patient groups inappropriately excluded from the measure or from target population?*
- *Are the exclusions consistent with the evidence?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

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

Conceptual rationale for SDS factors included ? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☒ **Yes** ☐ **No**

Risk adjustment summary:

- Each of the 8 PRO-PM measures are case-mix adjusted using a linear regression model.
- The same 9 factors are included in the case-mix adjustment approach for each PRO-PM. These include:
  - response percentile
  - decedent age group
  - payer
  - primary diagnosis
  - length of final hospice episode
  - respondent age group
  - respondent education
  - decedent's relationship to respondent

    - a variable indicating survey language and respondent's home language.
- The final case-mix factor parameters for the 8 PRO-PMs that were calculated from Q2 2015 data are included in the accompanying excel spreadsheet.
- The developers intend to re-calibrate the case-mix adjustment models on a routine basis. Specifically, they state that "*In keeping with longstanding practice in CAHPS initiatives, the parameterization of the case-mix adjustment model will be re-evaluated periodically, and the coefficients of the model will be reestimated regularly, as new data become available*".
- Each of the 8 PRO-PM measures also are adjusted for the mode of survey administration (this is done prior to the case-mix adjustment). See section 2b6. Parameters for this adjustment have not been provided in the submission materials.

Conceptual analysis of the need for SDS adjustment:

- The developer cites previous CAHPS and other survey research that has identified respondent characteristics (e.g., older, less education, better overall heath, better mental health) that are related to survey responses.

Empirical analysis of SDS factors:

- The developers state that they *" identified patient and caregiver characteristics as candidates for case-mix adjustment if they were available in hospice administrative or survey response data and were not within the hospice's contro*l".
- The developers identify several analyses they conducted in their case-mix model-building approach. These analysis and their decision rules applied to both SDS and non-SDS factors. It is not clear whether other candidate risk-factors were examined but ultimately not included in the case-mix adjustment. Some results (but not all) of these analyses are provided in the submission form, but only for the Rating of Hospice PRO-PM, based on Q1 and Q2 2015 data.
- Several SDS factors are included in the case-mix adjustment models.

Risk Model Diagnostics:

- The developers do not provide typical model discrimination statistics (e.g., R^2 values) but instead, correlated hospice rankings that result from calculating the measures with and without case-mix adjustment. They interpreted these results as indicating that between 3%-7% of hospice pairs would switch in terms of relative rankings across the various measures if the case-mix adjustment was not applied. These results demonstrate that the case-mix adjustment does have some effect; moreover, the developer also notes that these effects are in line with what is seen for other CAHPS-based PRO-PMs.
- Beyond the analysis showing some effect of case-mix adjustment, however, the developer did not provide additional information regarding the goodness-of-fit of the case-mix models.

Questions for the Committee:

- *Is an appropriate risk-adjustment strategy included for the 8 PRO-PMs?*
- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*
- *Two of the case-mix adjustment factors are not present at start of care (response percentile and length of final hospice episode). Is this okay?*
- *Do you agree with the developer's decision, based on their analysis, to include SDS factors in their risk-adjustment model?*

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

- For each PRO-PM. the developer calculated the number and percentage of hospices that were significantly above or below the national hospice average. Because only Q2 2015 data were available for analysis, they multiplied the number of respondents by 4 to approximate the number expected in a full measurement year.

They used a two-side test at an α=.05 level of significance.

Hospice Multi-Item Measure/Items	Count of Hospices Significantly Above	Count of Hospices Significantly Below	% Statistically Different from Hospice National Average
Hospice Team Communication	781	449	49%
Getting Timely Care	712	443	46%
Treating Your Family Member with Respect	864	290	46%
Getting Help with Symptoms	716	490	48%
Getting Emotional and Religious Support	850	330	47%
Getting Hospice Care Training	738	448	47%
Rating of Hospice	717	399	44%
Willingness to Recommend	754	413	46%

Question for the Committee:

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

2b6. Comparability of data sources/methods:

- Using information from 59 large hospice programs for 17,121 decedents from Q1-Q2 2015, the developers randomly assigned one-third of the cases to one of three modes of survey administration (mail-only, telephone only, and mixed) and fielded the Hospice CAHPS survey and then evaluated the effects of mode of administration on PRO-PM results.
- Based on the results of this analysis, each of the 8 PRO-PM measures also are adjusted for the mode of survey administration (this is done prior to the case-mix adjustment).

2b7. Missing Data

- Survey response rates, based on Q1-Q2 2015 data:
  - Overall response rate: 32.22%
  - Mail response rate: 31.69%
  - Telephone response rate: 33.32%
  - Mixed mode response rate: 43.69%
  - The developers note that the overall response rate is comparable to that of other CAHPS surveys
- The developers also present item non-response rates, including those missing due to appropriate skip, inappropriate skip. The developers interpret these results as indicating that <6% of respondents inappropriately skipped questions on the survey. They note their believe that "*it is unlikely that CAHPS Hospice Survey item results are biased due to systematic skipping of items by respondents*".
- In section S.14, the developer notes that if values for factors included in the case-mix adjustment are missing, the missing value, the missing value "*should be imputed using the hospice-specific mean for that variable*".

Guidance from the Validity Algorithm

Specifications consistent with evidence (Box 1) →Threats to validity mostly assessed, although information on frequency of exclusions needed (Box 2) → Empirical testing conducted for the measure as specified (Box 3) → Testing at the score-level conducted (Box 6) →High certainly that the scores are valid indicators of quality

Preliminary ratings for validity:

1 Hospice team communications: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

2) Getting Timely Care: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

3) Treating Family Member with Respect:	☒ **High**	☐ **Moderate**	☐ **Low**	☐ **Insufficient**
4) Getting Emotional and Religious Support:	☒ **High**	☐ **Moderate**	☐ **Low**	☐ **Insufficient**
5) Getting Help for Symptoms:	☒ **High**	☐ **Moderate**	☐ **Low**	☐ **Insufficient**
6) Getting Hospice Care Training:	☒ **High**	☐ **Moderate**	☐ **Low**	☐ **Insufficient**
7) Rating of Hospice:	☒ **High**	☐ **Moderate**	☐ **Low**	☐ **Insufficient**
8) Willingness to Recommend:	~~☒~~☐ **High**	☐ **Moderate**	☐ **Low**	☐~~☒~~ **Insufficient**

~~Rationale for Willingness to Recommend PRO-PM rating: No testing was conducted for this measure.~~

Committee pre-evaluation comments

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

2a & 2b.

Reliability Specifications

*yes well –described

* Both measure scores and data elements were tested for reliability and indicated demonstrating moderate certainty for reliability for 6 measures and less certainty for two others.
* Elements are well defined. Exclusion are identified. Calculation of top box scoring and logic/calculations are clear.
* no concerns about reliability specifications

Validity Specifications

*Yes

*Both measure scores and data elements were tested for reliability and indicated demonstrating moderate certainty for reliability for 6 measures and less certainty for two others.

* Moderate intercorrelations reflect the expected results for assessment of experience of care. Because of the subjectivity of a patient experience survey I would like to hear more about this analysis.
* no inconsistencies detected

Reliability Testing

*Yes

*Reliability was tested with an adequate scope to generalize and conducted at both the data level and score level.

* Scope adequate for widespread implementation.
* Reliability testing with an adequate number of patients. Cronbach's alpha and intercorrelation scores were in general adequate to demonstrate reliability.

Validity Testing

*Yes

*The 6 multi-item PRO-PMs demonstrate positive and statistically significant associations with the global hospice rating. but the global Willingness to Recommend was not tested.

* Scope of survey was extensive and thorough. Intercorrelations are considered "generally moderate" Exclusions are noted. The case-mix adjustment is detailed as are response percentiles. I agree that these measures are an indicator of quality.
* Sufficient testing and validity. See 2a2 above.

Threats to Validity

*missing data the main limitation. Alternative would be live assessment while pt/ caregiver are actually receiving care

*The exclusions listed in S.10 seem more to be people for whom the survey would not accurately be able to assess their experience of hospice because of factors that would make cloud the picture of what quality is for them. The Risk Adjustment methods seem appropriate and are in line with longstanding practice. I agree with the inclusion of SDS factors in the risk adjustment model to assess for differential experience.

* No from the report the missing data does not appear to constitute a threat to validity.
* I am not sure how the following statement in the measure evaluation summary is known to be accurate: "To ensure fair comparisons across hospices, CAHPS Hospice Survey scores must be adjusted for mode of survey administration, which affects scores but is not related to quality of hospice care. "

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.

- Data for these 8 PRO-PMs are collected via a survey that can be administered via mail, telephone, or mixed. Currently the responses of these surveys are not entered by the caregiver directly into an electronic system.
- The Hospice CAHPS survey is a survey that is fielded to the caregiver following the death of the hospice patient, and thus the data elements are not routinely generated and used during care delivery.
- The Hospice CAHPS survey underwent a successful 'dry run' in early 2015, and has been in continuous use with all eligible hospices since April 2015.
- The developer indicates that no fees or licensing costs are associated with use of the measures.
- No information is offered in the submission on expected costs to facilities to analyze survey results and contract with a vendor to distribute surveys and collect responses. Supporting Statement A for the CAHPS Hospice Survey, published by CMS on 10-29-14 as part of the OMB clearance process to implement the CAHPS Hospice Survey, suggests that conducting the survey will cost each hospice $399.04 per annum for developing the monthly data file for CMS (effectively 16 hours of a Medical Reviewer's time), and $3,300 to contract with a certified survey vendor.
- The response rate for the survey is fairly low, although it is comparable to other CAHPS surveys.

Question for the Committee:

- *Is there any evidence that the 2015 "dry run" or the past year of full implementation have led to unexpectedly burdensome or onerous administration requirements?*

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

Committee pre-evaluation comments

Criteria 3: Feasibility

*My only concern is that the data for the 8 PRO-MPS are collected by three different means: mail, phone or mixed. This method raises the question of data equivalency--would the answers to the question be the same regardless of route of administration? Or, is answering a question when asked by another person over the phone different than checking a box on a survey?
*high
* I wonder about the costs to the hospice organization vs actual costs for hospice organizations to develop, administer and analyze the survey results as compared to Supporting Statement A. The estimates seem very low.
* Approximately 33% response rate which varies on several factors including mode of survey administration, site of death and level of education of caregiver.

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

- Submission of the Hospice CAHPS data is required in the CMS Hospice Quality Reporting program.
- The results of the 8 PRO-PMs calculated from data from the Hospice CAHPS survey will be publicly reported in 2017, after a minimum of four quarters of data have been collected. The results will be reported in a

"Compare"-type website.

Improvement results: No trend data are available.

Unexpected findings (positive or negative) during implementation: No unexpected findings are reported to date.

Potential harms: No unintended consequences have been reported to date.

- **Feedback :**The CAHPS Hospice Survey has not been considered by the Measure Applications Partnership. A related measure, Family Evaluation of Hospice Care (NQF #0208), was recommended for inclusion in the Hospice Quality Reporting Program.

***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 measures outweigh any potential unintended consequences?*

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

Committee pre-evaluation comments

Criteria 4: Usability and Use

*The measure results will be widely available on a "Compare" type website. The performance results can further the goal of high-quality, efficient healthcare by highlighting caregivers' recalled perceptions of communication, timeliness of care, respect, emotional and religious support, symptom management and training in care--as fundamental needs for attention when a loved one is dying. The only possible unintended consequence is that a bereaved caregiver revisits potentially difficult experiences (voluntarily) when taking this survey. In my opinion, the benefits--learning about their experiences and the opportunity to improve care--far outweigh potential unintended consequences.
*high
* This data will be publicly reported in the near future. Because this measure will soon be used to determine some level of reimbursement for hospices could there be unintended consequences in the future for some organizations in rural, inner-city or other areas where the underserved often reside?
* Not publicly reported yet. Lots of data but not sure how it will be meaningfully used to improve quality of hospices.

Criterion 5: Related and Competing Measures

Competing measures

- 0208: Family Evaluation of Hospice Care
- 1623: Bereaved Family Survey
  - The result of the Family Evaluation of Hospice Care (FEHC) measure (#0208) is a single score that indicates a hospice agency's overall performance on symptom management, communication, provision of information, emotional support, and care coordination. However, because use the Hospice CAHPS survey is required as part of the CMS Hospice Quality Reporting Program for non-exempt agencies (i.e., those with at least 50 deaths in a year), the measure steward, NHPCO, has stated that "Hospices that utilize FEHC must be exempt from CMS-required CAHPS Hospice Survey" and "NHPCO is no longer accepting FEHC survey submissions through the DART system.".
  - The result of the Bereaved Family Survey measure (#1623) is a single score that indicates the family's perceptions of the quality of care that veterans received from the VA during the last month of life; aspects of care included in the measure are communication, emotional and spiritual support, pain management, and personal care needs.

Harmonization: N/A

Pre-meeting public and member comments

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*)**:** Click here to enter NQF number

Measure Title: CAHPS Hospice Survey

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Hospice Team Communication, Getting Timely Care, Treating Family Member with Respect, Getting Emotional and Religious Support, Getting Help for Symptoms, Getting Hospice Care Training

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): Click here to name the PRO

  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.

The CAHPS Hospice Survey is grounded in existing guidelines and conceptual models of the quality of hospice care, including NHPCO standards of practice for hospice programs (http://www.nhpco.org/standards), review of existing guidelines (Teno et al. 2001) , a conceptual model of end of life care developed by Stewart and colleagues (1999), the National Quality Forum Preferred Practices of Palliative and Hospice Care (2006), the advice of a Technical Expert Panel, and input from qualitative inquiry with caregivers of hospice decedents. Based on the NQF preferred practices, we provide an overview of the conceptual model for high quality hospice care and a detailed table that links structure, processes, and the proposed outcomes measures.

The survey assesses key processes of care that are critical to high quality hospice care. Specifically, the survey measures the degree to which the hospice program:

1) Keeps the hospice primary caregiver informed, listen to their concerns, and explain things in a way that they can understand (Hospice Team Communication)
2) Provides timely help when asked, even on weekends and holidays (Getting Timely Care)
3) Treats the patient with respect (Treating Family Member with Respect)
4) Provides the desired emotional and religious support prior to and after the patient's death (Getting Emotional and Spiritual Support)
5) Provides the desired help for treatment of symptoms (Getting Help for Symptoms)
6) Attends to the needs of the hospice primary caregiver for information and training to safely care for the patient at home (Getting Hospice Care Training)

Citations:

Teno JM, Casey VA, Welch L, Edgman-Levitan S. Patient-Focused, Family-Centered End-of-Life Medical Care: Views of the Guidelines and Bereaved Family Members. J Pain Symptom Manage-Special Section on Measuring Quality of Care at Life's End II 2001;22:738-51.

Stewart AL, Teno J, Patrick DL, Lynn J. The concept of quality of life of dying persons in the context of health care. J Pain Symptom Manage 1999;17:93-108.

National Quality Forum. National Framework and Preferred Practices for Palliative and Hospice Care. 2006.

Scores for all the all the multi-item measures can be used to improve the quality of care and, through public reporting improve, patients' and caregivers' experiences with hospice care.

Goal	Multi-Item Measure(s)	Relevant NQF Preferred Practices	Structure	Process	Outcomes	Example from the CAHPS Hospice Survey
Keeps the hospice primary caregiver informed, listen to their concerns, and explain things in a way that they can understand	Hospice Team Communication	NQF Preferred Practice 6	Qualified interdisciplinary staff with appropriate certification, excellent knowledge, and assessment skills Clear policy and procedures for documentation education of the hospice primary care giver regarding the plan of care and medicine administration	Interdisciplinary assessment and care incorporates needs of the dying patient and family Care plan is documented and updated	Hospice primary care giver felt their concerns and needs were addressed Hospice primary care giver felt that the hospice team listened to them Hospice primary caregiver felt that things were explained to them in a way they could understand	While your family member was in hospice care, how often did the hospice team explain things in a way that was easy to understand?
Provides timely help when asked, even on weekends and holidays	Getting Timely Care	NQF preferred Practice 2	Hospice with 24X7 call center Adequate staffing to ensure that visits are made when needed 24X7	Visits are provided in a timely manner	Hospice primary caregiver received the needed support in their role as a caregiver in a timely manner	While your family member was in hospice care, when you or your family member asked for help from the hospice team, how often did you get help as soon as you needed it?
Treats the hospice patient with respect	Treating Family Member with Respect	NQF Preferred Practice 19	Appropriately trained staff in communication that is sensitive to individuals' culture and values	Sensitively communicates with dying patient and family Staff treats the patient with dignity	Patient and family is treated with dignity and respect	While your family member was in hospice care, how often did the hospice team treat your family member with dignity and respect?

Goal	Multi-Item Measure(s)	Relevant NQF Preferred Practices	Structure	Process	Outcomes	Example from the CAHPS Hospice Survey
Provides the desired help for treatment of symptoms	Getting Help for Symptoms	NQF Preferred Practice 2, 12, 13, 14, 15, 16, 17, 29, and 31	Policy and procedures are in place for screening of symptoms, appropriate assessment, care planning, and monitoring of treatment plan Qualified interdisciplinary team, including physician, nurse, social worker, pharmacist, spiritual counselor and others, that is available 24 hours a day, 7 days a week	Patient screened for needs Among those with a concern, an in-depth assessment is done utilizing standardized measurement tools An individualized plan of care developed Plan is monitored for whether it is achieving patient and family goals	Patient receives desired level of palliation of symptoms	Did your family member get as much help with pain as he or she needed?

Goal	Multi-Item Measure(s)	Relevant NQF Preferred Practices	Structure	Process	Outcomes	Example from the CAHPS Hospice Survey
Attend to the needs of caregivers for emotional and spiritual support prior to and after the death of the patient	Getting Emotional and Religious Support	NQF preferred practices 20,21,22, and 23	Policy and procedures are in place for screening of the need for emotional and spiritual support, appropriate assessment, care planning, and monitoring of treatment plan Staff trained in sensitively communicating with dying patient and family regarding Religious and spiritual needs Qualified interdisciplinary team, including physician, nurse, social worker, pharmacist, spiritual counselor and others, that is available 24 hours a day, 7 days a week Hospice foster relat	Patient screened for needs Among those with a concern, an in-depth assessment is done An individualized plan of care Plan is monitored for whether it is achieving patient and family goals Relationship and involvement fostered with patients' clergy and religious advisors	Family receives desired emotional support prior to and after the death of their loved one Dying patient and family receive the desired religious support	While your family member was in hospice care, how much emotional support did you get from the hospice team? Support for religious or spiritual beliefs including talking, praying, quiet time, or other ways of meeting your religious or spiritual needs. While your family member was in hospice care, how much support for your religious and spiritual beliefs did you get from the hospice team?
Attends to the needs of the hospice primary caregiver for information and training to safely care for the patient at home	Getting Hospice Care Training	NQF Preferred Practice 11	Staff with appropriate education and communication skills in supporting family primary care giver in their role	Hospice staff provides information and training in a way that the hospice primary caregiver can understand about providing care for the dying hospice patient and when to contact the hospice for additional help	Hospice primary caregiver receives the needed training in their caregiving role	Did the hospice team give you the training you needed about if and when to give more pain medicine to your family member?

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*).

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 each 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

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 #: 2651
De.2. **Measure Title:** CAHPS® Hospice Survey (experience with care)
Co.1.1. **Measure Steward:** Centers for Medicare and Medicaid Services
De.3. **Brief Description of Measure:** The measures submitted here are derived from the CAHPS® Hospice Survey, which is a 47-item standardized questionnaire and data collection methodology. The survey is intended to measure the experiences of hospice patients and their primary caregivers.
The measures proposed here include the following six multi-item measures.

- Hospice Team Communication
- Getting Timely Care
- Treating Family Member with Respect
- Getting Emotional and Religious Support
- Getting Help for Symptoms
- Getting Hospice Training

In addition, there are two other measures, also called "global ratings."

- Rating of the hospice care
- Willingness to recommend the hospice

Below we list each multi-item measure and its constituent items, along with the two ratings questions. Then we briefly provide some general background information about CAHPS surveys.

List of CAHPS Hospice Survey Measures

Multi-Item Measures
Hospice Team Communication (Composed of 6 items)
+ While your family member was in hospice care, how often did the hospice team keep you informed about when they would arrive to care for your family member?
+ While your family member was in hospice care, how often did the hospice team explain things in a way that was easy to understand?
+ How often did the hospice team listen carefully to you when you talked with them about problems with your family member's hospice care?
+ While your family member was in hospice care, how often did the hospice team keep you informed about your family member's condition?
+ While your family member was in hospice care, how often did the hospice team listen carefully to you?
+ While your family member was in hospice care, how often did anyone from the hospice team give you confusing or contradictory information about your family member's condition or care?
Getting Timely Care (Composed of 2 items)
+ While your family member was in hospice care, when you or your family member asked for help from the hospice team, how often did you get help as soon as you needed it?
+ How often did you get the help you needed from the hospice team during evenings, weekends, or holidays?
Treating Family Member with Respect (Composed of 2 items)
+ While your family member was in hospice care, how often did the hospice team treat your family member with dignity and respect?

+ While your family member was in hospice care, how often did you feel that the hospice team really cared about your family member?

Providing Emotional Support (Composed of 3 items)

+ While your family member was in hospice care, how much emotional support did you get from the hospice team?
+ In the weeks after your family member died, how much emotional support did you get from the hospice team?
+ Support for religious or spiritual beliefs includes talking, praying, quiet time, or other ways of meeting your religious or spiritual needs. While your family member was in hospice care, how much support for your religious and spiritual beliefs did you get from the hospice team?

Getting Help for Symptoms (Composed of 4 items)

+ Did your family member get as much help with pain as he or she needed?
+ How often did your family member get the help he or she needed for trouble breathing?
+ How often did your family member get the help he or she needed for trouble with constipation?
+ How often did your family member receive the help he or she needed from the hospice team for feelings of anxiety or sadness?

Getting Hospice Care Training (Composed of 5 items)

+ Did the hospice team give you enough training about what side effects to watch for from pain medicine?
+ Did the hospice team give you the training you needed about if and when to give more pain medicine to your family member?
+ Did the hospice team give you the training you needed about how to help your family member if he or she had trouble breathing?
+ Did the hospice team give you the training you needed about what to do if your family member became restless or agitated?
+ Side effects of pain medicine include things like sleepiness. Did any member of the hospice team discuss side effects of pain medicine with your or your family member?

Rating Measures:

In addition to the multi-item measures, there are two "global" ratings measures. These single-item measures indicate on the one hand the need for quality improvement and on the other hand provide families and patients looking for care with evaluations of the care provided by the hospice. The items are rating of hospice care and willingness to recommend the hospice.

+ Rating of Hospice Care: Using any number from 0 to 10, where 0 is the worst hospice care possible and 10 is the best hospice care possible, what number would you use to rate your family member's hospice care?
+ Willingness to Recommend Hospice: Would you recommend this hospice to your friends and family?

The CAHPS Hospice Survey is a standardized survey instrument designed to collect reports and ratings of experiences with hospice care. The survey is completed by the primary caregiver of the patient who died while receiving hospice care (hereafter, "decedent"). The primary caregiver is intended to be the family member or friend most knowledgeable about the decedent's hospice care, and is identified through hospice administrative records. Data collection for sampled decedents/caregivers is initiated two months following the month of the decedent's death.

The CAHPS Hospice Survey is part of the CAHPS family of experience of care surveys and is available in the public domain at https://cahps.ahrq.gov/surveys-guidance/hospice/index.html. CMS initiated national implementation of the CAHPS Hospice Survey in 2015. Hospices meeting CMS eligibility criteria were required to administer the survey for a "dry run" for at least one month of sample from the first quarter of 2015. Beginning with the second quarter of 2015, hospices are required to participate on an ongoing monthly basis in order to receive their full Annual Payment Update from CMS. Information regarding survey content and national implementation requirements, including the latest versions of the survey instrument and standardized protocols for data collection and submission, are available at: http://www.hospicecahpssurvey.org/.

A list of the CAHPS Hospice Survey measures, including the components of the multi-item measures can be found in Appendix A.

1b.1. **Developer Rationale:** Each year 2.4 million Americans die. Too often, this dying experience is marred by untreated pain or other symptoms, lack of shared decision making, and insufficient emotional support (Teno Clarridge et al. 2004; Teno, Freedman et al. 2015).The Medicare hospice care benefit offers palliative care for individuals with a prognosis of living 6 months or less if their terminal illness runs its normal course. The number of Medicare beneficiaries receiving hospice services has grown from 513,000 in FY 2000 to over 1.3 million in FY 2013. Similarly, Medicare hospice expenditures have risen from $2.9 billion in FY 2000 to an estimated $15.1 billion in FY 2013. The CMS Office of the Actuary projects that hospice expenditures are expected to continue to increase, by approximately 8 percent annually, reflecting an increase in the number of Medicare beneficiaries, more beneficiary awareness of the Medicare Hospice Benefit for end-of-life care, and a growing preference for care provided in home and community-based settings. (More details may be found in the Federal Register. (Please see: 80 FR 47142.)

Previous research has shown striking variation across hospices with regard to their provision of key hospice care processes. For example, visits by professional hospice staff in the last two days of life vary considerably across hospices, with 3.2% of hospices at the 25th percentile providing such visits and 19.5% at the 75th percentile providing them. A 2013 Office of Inspector General report found that 27% of the hospice providers did not provide any general inpatient (GIP) level of care (intensive inpatient care that is meant to be short-term; Office of the Inspector General 2013). The provision of GIP care varies by region of the country with 77% of hospice providers in the South providing GIP level of care compared to 91% of providers in New England (Plotzke, Christian et al., 2014). Previous analyses of the Family Evaluation of Hospice Care (FEHC) voluntary repository maintained by the National Hospice and Palliative Care Organization have found substantial variation in several of the key domains measured by the survey (e.g., 12.6% of respondents reporting unmet needs for emotional and spiritual support for hospices in the 25th percentile of hospices compared to 21.4% of respondents in the 75th percentile; Connor, Teno et al. 2005). Evidence of variation in hospice care processes and patient and family experiences of hospice care point to the need for performance measures that can be systematically implemented to monitor hospice performance and make fair comparisons across hospices.

To that end, the Affordable Care Act mandated that the US Department of Health and Human Services (DHHS) implement a quality reporting program for hospices. The CAHPS Hospice Survey is a component of that program. The conceptual model for the development of the survey was grounded in review of existing surveys (Lendon, Ahluwalia et al., 2015), a previous review of the existing guidelines regarding quality of end of life (Teno, Casey et al., 2001), National Qualify Forum Preferred Practices in Palliative Care, and the work of the National Consensus Project for Quality Palliative Care (http://www.nationalconsensusproject.org/Guidelines_Download2.aspx).

The CAHPS Hospice Survey is part of a family of CAHPS surveys that focuses on patient or caregiver experiences with health care. The CAHPS Hospice Survey presents primary caregivers (as identified by hospices) with a set of standardized questions about the hospice care they and their family member or friend received. The survey and associated methodology is intended to create reliable and valid data about hospice experiences that can be published on the web to help families choose hospices for their family members or friends, and to create incentives for quality improvement among hospice providers.

Citations:

Connor SR, Teno J, Spence C, Smith N. Family evaluation of hospice care: results from voluntary submission of data via website. J Pain Symptom Manage 2005;30:9-17.
Lendon JP, Ahluwalia SC, Walling AM, et al. Measuring Experience With End-of-Life Care: A Systematic Literature Review. J Pain Symptom Manage 2015;49:904-15.e1-3.
Department of Health and Human Services Office of the Inspector General. Medicare Hospice: Use of General Inpatient Care. Washington, DC. 2013.
Plotzke M, Christian TJ, Pozniak A, et al. Medicare Hospice Payment Reform: Analyses to Support Payment Reform. 2014 May 1.
Rhodes RL, Mitchell SL, Miller SC, Connor SR, Teno JM. Bereaved family members´ evaluation of hospice care: what factors influence overall satisfaction with services? J Pain Symptom Manage 2008;35:365-71.
Teno JM, Casey VA, Welch L, Edgman-Levitan S. Patient-Focused, Family-Centered End-of-Life Medical Care: Views of the Guidelines and Bereaved Family Members. J Pain Symptom Manage-Special Section on Measuring Quality of Care at Life´s End II 2001;22:738-51.
Teno JM, Clarridge BR, Casey V, et al. Family perspectives on end-of-life care at the last place of care. JAMA 2004;291:88-93.
Teno JM, Freedman VA, Kasper JD, Gozalo P, Mor V. Is Care for the Dying Improving in the United States? J Palliat Med 2015;18:662-6.

S.4. **Numerator Statement:** CMS calculates CAHPS Hospice Survey measures using top-box scoring. The top-box score refers to the percentage of caregiver respondents that give the most positive response. Details regarding the definition of most positive response are noted in Section S.6 below.

S.7. **Denominator Statement:** The measure's denominator is the number of survey respondents who answered the item. The target population for the survey is primary caregivers of hospice decedents. The survey uses screener questions to identify respondents eligible to respond to subsequent items. Therefore, denominators will vary by survey item (and corresponding multi-item measures, if applicable) according to the eligibility of respondents for each item.

S.10. **Denominator Exclusions:** Cases are excluded from the survey target population if:

- The hospice patient is still alive
- The decedent's age at death was less than 18
- The decedent died within 48 hours of his/her last admission to hospice care

- The decedent had no caregiver of record
- The decedent had a caregiver of record, but the caregiver does not have a U.S. or U.S. Territory home address
- The decedent had no caregiver other than a nonfamilial legal guardian
- The decedent or caregiver requested that they not be contacted (i.e., by signing a no publicity request while under the care of hospice or otherwise directly requesting not to be contacted)
- The caregiver is institutionalized, has mental/physical incapacity, has a language barrier, or is deceased
- The caregiver reports on the survey that he or she "never" oversaw or took part in decedent's hospice care

De.1. **Measure Type:** PRO
S.23. **Data Source:** Patient Reported Data/Survey
S.26. **Level of Analysis:** Facility

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

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

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

De.4. **IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** N/A

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
Template_MeasSubm_Evidence_2_26-635921001262567153.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)*
Each year 2.4 million Americans die. Too often, this dying experience is marred by untreated pain or other symptoms, lack of shared decision making, and insufficient emotional support (Teno Clarridge et al. 2004; Teno, Freedman et al. 2015).The Medicare hospice care benefit offers palliative care for individuals with a prognosis of living 6 months or less if their terminal illness runs its normal course. The number of Medicare beneficiaries receiving hospice services has grown from 513,000 in FY 2000 to over 1.3 million in FY 2013. Similarly, Medicare hospice expenditures have risen from $2.9 billion in FY 2000 to an estimated $15.1 billion in FY 2013. The CMS Office of the Actuary projects that hospice expenditures are expected to continue to increase, by approximately 8 percent annually, reflecting an increase in the number of Medicare beneficiaries, more beneficiary awareness of the Medicare Hospice Benefit for end-of-life care, and a growing preference for care provided in home and community-based settings. (More details may be found in the Federal Register. (Please see: 80 FR 47142.)

Previous research has shown striking variation across hospices with regard to their provision of key hospice care processes. For example, visits by professional hospice staff in the last two days of life vary considerably across hospices, with 3.2% of hospices at the 25th percentile providing such visits and 19.5% at the 75th percentile providing them. A 2013 Office of Inspector General report found that 27% of the hospice providers did not provide any general inpatient (GIP) level of care (intensive inpatient care that is meant to be short-term; Office of the Inspector General 2013). The provision of GIP care varies by region of the country with 77% of hospice providers in the South providing GIP level of care compared to 91% of providers in New England (Plotzke, Christian et al., 2014). Previous analyses of the Family Evaluation of Hospice Care (FEHC) voluntary repository maintained by the National Hospice and Palliative Care Organization have found substantial variation in several of the key domains measured by the survey (e.g., 12.6% of respondents reporting unmet needs for emotional and spiritual support for hospices in the 25th percentile of hospices compared to 21.4% of respondents in the 75th percentile; Connor, Teno et al. 2005). Evidence of variation in hospice care processes and patient and family experiences of hospice care point to the need for performance measures that can be systematically implemented to

monitor hospice performance and make fair comparisons across hospices.

To that end, the Affordable Care Act mandated that the US Department of Health and Human Services (DHHS) implement a quality reporting program for hospices. The CAHPS Hospice Survey is a component of that program. The conceptual model for the development of the survey was grounded in review of existing surveys (Lendon, Ahluwalia et al., 2015), a previous review of the existing guidelines regarding quality of end of life (Teno, Casey et al., 2001), National Qualify Forum Preferred Practices in Palliative Care, and the work of the National Consensus Project for Quality Palliative Care (http://www.nationalconsensusproject.org/Guidelines_Download2.aspx).

The CAHPS Hospice Survey is part of a family of CAHPS surveys that focuses on patient or caregiver experiences with health care. The CAHPS Hospice Survey presents primary caregivers (as identified by hospices) with a set of standardized questions about the hospice care they and their family member or friend received. The survey and associated methodology is intended to create reliable and valid data about hospice experiences that can be published on the web to help families choose hospices for their family members or friends, and to create incentives for quality improvement among hospice providers.

Citations:

Connor SR, Teno J, Spence C, Smith N. Family evaluation of hospice care: results from voluntary submission of data via website. J Pain Symptom Manage 2005;30:9-17.
Lendon JP, Ahluwalia SC, Walling AM, et al. Measuring Experience With End-of-Life Care: A Systematic Literature Review. J Pain Symptom Manage 2015;49:904-15.e1-3.
Department of Health and Human Services Office of the Inspector General. Medicare Hospice: Use of General Inpatient Care. Washington, DC. 2013.
Plotzke M, Christian TJ, Pozniak A, et al. Medicare Hospice Payment Reform: Analyses to Support Payment Reform. 2014 May 1.
Rhodes RL, Mitchell SL, Miller SC, Connor SR, Teno JM. Bereaved family members´ evaluation of hospice care: what factors influence overall satisfaction with services? J Pain Symptom Manage 2008;35:365-71.
Teno JM, Casey VA, Welch L, Edgman-Levitan S. Patient-Focused, Family-Centered End-of-Life Medical Care: Views of the Guidelines and Bereaved Family Members. J Pain Symptom Manage-Special Section on Measuring Quality of Care at Life´s End II 2001;22:738-51.
Teno JM, Clarridge BR, Casey V, et al. Family perspectives on end-of-life care at the last place of care. JAMA 2004;291:88-93.
Teno JM, Freedman VA, Kasper JD, Gozalo P, Mor V. Is Care for the Dying Improving in the United States? J Palliat Med 2015;18:662-6.

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.*

The performance data reported here refer to the 2,512 hospices that submitted data to the Hospice CAHPS Data Warehouse for Quarter 2 of 2015. There are 4,259 hospices in the U.S. according to CMS's latest count. We believe that approximately 16% of them (681) serve fewer than 50 patients annually, which would make them exempt from the Hospice CAHPS survey. This means that approximately 70% of the 3,578 survey-eligible hospices in the U.S. are accounted for in the Quarter 2 data. Since this is a new survey, we expect that the participation percentage will increase over time.

Table 1b.2a displays the mean, standard deviation, median, and interquartile range for each of the composite and global CAHPS Hospice Survey measures. Scores are for the 2,512 hospices that submitted data in Quarter 2 2015, reflecting the complete universe of hospices eligible to participate in the CAHPS Hospice Survey during that quarter (i.e., all hospices nationwide that were established prior to 2015 and had at least 50 survey-eligible decedents/caregivers in calendar year 2014).
Measures with the highest top-box scores were Getting Emotional and Religious Support (top-box mean=91.8; interquartile range=88.8, 96.4) and Treating Family Member with Respect (top-box mean=91.4; interquartile range=88.0, 96.3). Measures with the lowest top-box scores were Getting Hospice Care Training (top-box mean=72.7; interquartile range=65.5, 81.0) and Getting Help for Symptoms (top-box mean=75.7; interquartile range=69.3, 83.2).

Table 1b.2a. Distribution of Top-Box Scores for CAHPS Hospice Survey Measures, Quarter 2 2015

	Mean	Standard Deviation	25th Percentile	50th Percentile	75th Percentile
Composite Measures					
Hospice Team Communication	82.0	9.0	77.0	82.8	87.7
Getting Timely Care	77.4	12.1	70.7	78.0	85.3
Treating Family Member with Respect			91.4	7.0	88.0	92.1	96.3

Getting Emotional and Religious Support		91.8	6.5	88.8	92.9	96.4
Getting Help for Symptoms	75.7	11.9	69.3	76.6	83.2	
Getting Hospice Care Training	72.7	12.8	65.5	73.2	81.0	
Global Measures						
Rating of Hospice 84.0	11.0	78.0	85.1	91.4		
Willingness to Recommend	84.9	11.4	79.3	86.2	92.7	

A description of current performance for all items included in the composite measures, calculated using Quarter 2 2015 data from the CAHPS Hospice Survey Data Warehouse and including all of the requested statistics, is attached in the Excel file, "CAHPS Hospice Survey Main Submission Form Supplementary Tables 2016_3_14.xlsx" in the worksheet entitled "Table 1b.2b."

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.

Patient/caregiver reported experience of care is an increasingly measured construct used to assess health care service quality with the goal of providing consumer information and improving service quality through provider competitiveness (Goldstein, Cleary, Langwell, Zaslavsky, & Heller, 2005). Table 1b.2 in Appendix A documents variation among the 2,512 hospice providers that submitted data in Quarter 2 2015. For example, the interquartile range for the multi-item measure Hospice Team Communication was 77.0 to 87.2. This variation is comparable to that reported based on analyses of the FEHC repository published in 2005 (Connor, Teno et al. 2005).

Citations:

Connor SR, Teno J, Spence C, Smith N. Family evaluation of hospice care: results from voluntary submission of data via website. J Pain Symptom Manage 2005;30:9-17.

Goldstein, E., Cleary, P. D., Langwell, K. M., Zaslavsky, A. M., & Heller, A. (2005). Medicare CAHPS: A tool for performance improvement. Health Care Financing Review, 22, 101–107.

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.*

Comparison of case-mix and mode-adjusted CAHPS Hospice Survey scores from Quarter 1 and Quarter 2 2015 indicate that compared to caregivers of white decedents within a given hospice, caregivers of black decedents report worse experiences for three measures, similar experiences for three measures, and better experiences for two measures. Compared to caregivers of white decedents within a given hospice, caregivers of Hispanic decedents report worse experiences for one measure, similar experiences for three measures, and better experiences for four measures. Importantly, this analysis does not take into account the propensity of black or Hispanic decedents to receive care from hospices that deliver worse care on average, a phenomenon that has previously been hypothesized (Rhodes, Teno and Connor, 2007). Ongoing analyses will address the degree to which there may be unequal access to high quality hospices among racial/ethnic groups.

Citation:

Rhodes RL, Teno JM, Connor SR. African American bereaved family members´ perceptions of the quality of hospice care: lessened disparities, but opportunities to improve remain. J Pain Symptom Manage 2007;34:472-9.

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.

CMS plans to examine disparities using CAHPS Hospice Survey data, but because we have not completed four quarters of data collection, this is planned for future research. Previous analyses of data from the FEHC identified disparities in experiences of care across racial/ethnic groups; however, these disparities were smaller than those observed for those dying in sites of care other than hospice (Rhodes, Teno & Connor 2007; Welch, Teno & Mor 2005).

There is an extensive literature that discusses disparities in hospice care. For several recent examples, please see Payne 2016, Institute of Medicine 2015, and Ong et al 2015.

Citations:

Payne (2016) Racially Associated Disparities in Hospice and Palliative Care Access: Acknowledging the Facts While Addressing the

Opportunities to Improve. Journal of Palliative Medicine v. 19. n 2.

IOM (Institute of Medicine). 2015. Dying in America: Improving quality and honoring individual preferences near the end of life. Washington, DC: The National Academies Press.
Ong Jeremy, Brennsteiner Alex, Chow Elizabeth, and Hebert Randy S. (2015) Correlates of Family Satisfaction with Hospice Care: General Inpatient Hospice Care versus Routine Home Hospice Care. Journal of Palliative Medicine. December 2015, 19(1): 97-100.

Rhodes RL, Teno JM, Connor SR. African American bereaved family members´ perceptions of the quality of hospice care: lessened disparities, but opportunities to improve remain. J Pain Symptom Manage 2007;34:472-9.
Welch LC, Teno JM, Mor V. End-of-life care in black and white: race matters for medical care of dying patients and their families. J Am Geriatr Soc 2005;53:1145-53.

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

Affects large numbers, Patient/societal consequences of poor quality, Other

1c.2. If Other: Required under the Affordable Care Act

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.

A large body of evidence, including a recent Institute of Medicine report, describes inadequate palliation of pain and other symptoms, unmet needs of families for emotional support, and greater need for bereavement services among families of hospice decedents (Hall Schroder & Weaver 2002; Hickman, Tilden & Tolle 2001; Institute of Medicine 2015; Kayser-Jones 2002; Pan, Morrison et al. 2001; Puntillo, Arai et al. 2010; Teno, Clarridge et al. 2004; Teno, Freedman et al. 2015; Teno, Kabumoto, et al. 2000).

Hospice patients' and families' assessments of hospice care are a concern to the U.S. Department of Health and Human Services for several reasons: (1) to address CMS' aims of better care, smarter spending, and healthier people and communities; (2) reporting of hospice quality to CMS is required under the Affordable Care Act; (3) the large expense the hospice benefit represents to Medicare, and (4) the large number of Medicare beneficiaries receiving the benefit.

CMS is focused on providing a higher quality of health care by making health care more person-centered, reliable, accessible, and safe. In addition CMS seeks to reduce the cost of quality health care for individuals, families, and CMS is working to improve the health of Americans by supporting proven interventions to address behavioral, social, and environmental determinants of health, and deliver higher-quality care. (CMS 2016)

The Affordable Care Act requires that hospices submit quality data to the Department and mandates that hospices that fail to report their data are subject to having market basket update reduced by 2 percentage points. (Federal Register 2015)

More than 85% of hospice patients were paid for by Medicare in 2014, while about seven percent were paid for by private insurance and five percent were paid for by Medicaid (NHPCO, 2015, 10). According to the Medicare Payment Advisory Commission's March 2015 Report to Congress, Medicare hospice expenditures in 2013 totaled about $15.1 billion. More than 1.3 million Medicare beneficiaries received hospice services in that year. (Medpac, 2015)

1c.4. Citations for data demonstrating high priority provided in 1a.3

CMS (2016) CMS Quality Strategy. Retrieved from https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/Downloads/CMS-Quality-Strategy.pdf. February 23, 2016.
Federal Register (2015) Medicare Program; FY 2016 Hospice Wage Index and Payment Rate Update and Hospice Quality Reporting Requirements; Final Rule. 80 FR 47147
Puntillo KA, Arai S, Cohen NH, et al. Symptoms experienced by intensive care unit patients at high risk of dying. Critical care medicine 2010;38:2155-60.
Breitbart W, Alici Y. Agitation and delirium at the end of life: "We couldn´t manage him". JAMA 2008;300:2898-910, E1.

Hall P, Schroder C, Weaver L. The last 48 hours of life in long-term care: a focused chart audit. J Am Geriatr Soc 2002;50:501-6.
Hickman SE, Tilden VP, Tolle SW. Family reports of dying patients´ distress: the adaptation of a research tool to assess global symptom distress in the last week of life. J-Pain-Symptom-Manage 2001;22:565-74.
IOM (Institute of Medicine). 2015. Dying in America: Improving quality and honoring individual preferences near the end of life. Washington, DC: The National Academies Press.
Kayser-Jones J. The experience of dying: an ethnographic nursing home study. Gerontologist 2002;42 Spec No 3:11-9.
Medpac (2015) Report to Congress Medicare Payment Policy, March 2015 p 285. Retrieved from http://www.medpac.gov/-documents-/reports, February 23, 2016.
NHPCO (2015). NHPCO's Facts & Figures, Hospice Care in America. Retrieved from http://www.nhpco.org/sites/default/files/public/Statistics_Research/2015_Facts_Figures.pdf, February 23, 2016
Pan CX, Morrison RS, Meier DE, et al. How prevalent are hospital-based palliative care programs? Status report and future directions. J Palliat Med 2001;4:315-24.
Teno JM, Clarridge BR, Casey V, et al. Family perspectives on end-of-life care at the last place of care. JAMA 2004;291:88-93.
Teno JM, Freedman VA, Kasper JD, Gozalo P, Mor V. Is Care for the Dying Improving in the United States? J Palliat Med 2015;18:662-6.
Teno JM, Kabumoto G, Wetle T, Roy J, Mor V. Daily pain that was excruciating at some time in the previous week: prevalence, characteristics, and outcomes in nursing home residents. J Am Geriatr Soc 2004;52:762-7.

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.)*

To develop the CAHPS Hospice survey, CMS called for topics and questions in a federal register notice (78 FR 5458), conducted a systematic review of the literature on existing tools to measure experiences with end of life care, and solicited input from key stakeholders and national experts via a Technical Expert Panel meeting, to develop a comprehensive list of topic areas for potential inclusion in the survey. In addition, CMS conducted focus groups and individual interviews in English and Spanish with the caregivers (i.e., family members) of hospice decedents. The focus groups and individual interviews were intended to identify content areas of importance for the survey. The results of these interviews highlighted the importance of communication by the hospice team, respect for the patient, and the provision of information.

Based on all of these inputs, a set of questions were developed and tested in cognitive interviews. CMS tested these questions and refined the list of items that were included in a field test. The field test involved three similar versions of the survey which were differentiated by hospice care setting: 1. home, 2. nursing home and 3. in-patient setting (hospital or stand alone hospice unit)

After the field test, the questionnaires were re-evaluated statistically and with additional cognitive testing. The decision was made to create a single questionnaire for all settings that includes some setting-specific items.

A description of the survey development process, including results of the field test, is available at: http://www.hospicecahpssurvey.org/Documents/Hospice_Field_Test_Report_2014.pdf.

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):*

De.6. **Cross Cutting Areas** *(check all the areas that apply):*

Access, Palliative Care and End of Life Care, Patient and Family Engagement

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.)*

www.hospicecahpssurvey.org

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)*
Attachment **Attachment:** CAHPS_Hospice_Survey_Main_Submission_Form_Supplementary_Tables_2016_3_14-635936455961497856.xlsx

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

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.
CMS calculates CAHPS Hospice Survey measures using top-box scoring. The top-box score refers to the percentage of caregiver respondents that give the most positive response. Details regarding the definition of most positive response are noted in Section S.6 below.

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.)*
Caregiver respondents describe experiences with the hospice named on the survey cover while the decedent was in hospice care.

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.
For each survey item, the top box numerator is the number of respondents who selected the most positive response category(ies), as follows:

For items using a "Never/Sometimes/Usually/Always" response scale, the top box numerator is the number of respondents who answer "Always."

For items using a "Yes, definitely/Yes, somewhat/No" response scale, the top box numerator is the number of respondents who answer "Yes, definitely."

For items using a "Too Little/Right Amount/Too Much" response scale, the top box numerator is the number of respondents who answer "Right Amount."

The top box numerator for the Rating of Hospice item is the number of respondents who answer 9 or 10 for the item (on a scale of 0 to 10, where 10 is the "Best Hospice Care Possible").

The top box numerator for the Willingness to Recommend item is the number of respondents who answer "Definitely Yes" (on a scale of "Definitely No/Probably No/Probably Yes/Definitely Yes").

Calculation of hospice-level multi-item measures
0. Score each item using top box method, possible values of 0 or 100
1. Calculate mode adjusted scores for each item for each respondent
2. Calculate case-mix adjusted scores for each item for each hospice
3. Take the unweighted means of the mode- and case-mix-adjusted hospice-level items to form multi-item measures

Example: hospice-level multi-item measure for 'Getting Timely Care':

0. Score each item using top box method, possible values of 0 or 100
Both items in "Getting Care Quickly" have four response options: Never, Sometimes, Usually, Always. Recode each item as 100 for "Always" and 0 for "Never", "Sometimes", or "Usually".

Item #1. While your family member was in hospice care, when you or your family member asked for help from the hospice team, how often did you get help as soon as you needed it?

Item #2. How often did you get the help you needed from the hospice team during evenings, weekends, or holidays?

1. Calculate mode-adjusted scores for each item for each respondent

2. Calculate case-mix adjusted scores for each item for each hospice
Each item is case mix adjusted separately; this step produces case-mix adjusted item-level scores for each hospice.

3. Take the unweighted means of the case-mix adjusted hospice-level items to form multi-item measures.
If the case-mix adjusted scores for a hospice are 95 for item #1 and 90 for item #2, then the hospice-level 'Getting Timely Care' would be calculated as (Item1 + Item2) / 2 = (95 + 90) / 2 = 92.5.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The measure's denominator is the number of survey respondents who answered the item. The target population for the survey is primary caregivers of hospice decedents. The survey uses screener questions to identify respondents eligible to respond to subsequent items. Therefore, denominators will vary by survey item (and corresponding multi-item measures, if applicable) according to the eligibility of respondents for each item.

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

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)*
For each item in a multi-item measure, as well as for the ratings measures,the top box denominator is the number of respondents per hospice who answered the item. For each multi-item measure score, the denominator is the number of respondents that answers at least one item within the multi-item measure. Multi-item measure scores are the average proportion of respondents that gave responses in the most positive category(ies) across the items in the multi-item measure (as discussed in S.6).

Survey population: Primary caregivers of patients who died while receiving care from a given hospice in a given month.

Denominator for Multi-Item Measures: The number of respondents who answer at least one item within the multi-item measure.

Denominator for Rating Measures: The number of respondents who answered the item.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
The exclusions noted in here are those who are ineligible to participate in the survey. The one exception is caregivers who report on the survey that they "never" oversaw or took part in the decedent's care; these respondents are instructed to complete the "About You" and "About Your Family Member" sections of the survey only.

Cases are excluded from the survey target population if:

- The hospice patient is still alive
- The decedent's age at death was less than 18
- The decedent died within 48 hours of his/her last admission to hospice care
- The decedent had no caregiver of record
- The decedent had a caregiver of record, but the caregiver does not have a U.S. or U.S. Territory home address
- The decedent had no caregiver other than a nonfamilial legal guardian
- The decedent or caregiver requested that they not be contacted (i.e., by signing a no publicity request while under the care

of hospice or otherwise directly requesting not to be contacted)

- The caregiver is institutionalized, has mental/physical incapacity, has a language barrier, or is deceased
- The caregiver reports on the survey that he or she "never" oversaw or took part in decedent's hospice care

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)*

Please see S.10.The CAHPS Hospice Survey Quality Assurance Guidelines (available at: (http://www.hospicecahpssurvey.org/Content/QualityAssurance.aspx) contain detailed information regarding how to code decedent/caregiver cases, and how to code appropriately and inappropriately skipped items, as well as items with multiple responses.

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)*

CAHPS Hospice Survey measure scores are used for reporting at the hospice-level (i.e., not stratified by region or other characteristics).

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)

Other

If other: Case Mix Adjustment

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)*

Case-mix adjustment is conducted via linear regression. The following items are included in the case-mix adjustment model:

Items from survey responses:

What is your age?
1=18 to 24 years
2=25 to 34 years
3=35 to 44 years
4=45 to 54 years
5=55 to 64 years
6=65 to 74 years
7=75 to 84 years
8=85 years or older

What is the highest grade or level of school that you have completed?
1=8th grade or less
2=Some high school, but did not graduate
3=High school graduate or GED
4=Some college or 2-year degree
5=4-year college graduate
6=More than 4-year college degree

How are you related to the person listed on the survey cover letter?
1=My spouse or partner
2=My parent
3=My mother-in-law or father-in-law
4=My grandparent
5=My aunt or uncle
6=My sister or brother
7=My child
8=My friend
9=Other

What language do you mainly speak at home?

1=English
2=Spanish
3=Chinese
4=Some other language

Response percentile
Calculated by ranking lag time (days between date of death and date of survey response) among respondents for each hospice in each month, then dividing by total sample size. This is included as a continuous predictor in case mix.

From hospice administrative data:

Decedent age at death is calculated from date of birth and date of death. Categories used in case mix:
18-54 years
55-64 years
65-69 years
70-74 years
75-79 years
80-84 years
85-89 years
90+ years

Up to three payers for hospice care are collected. Categories used in case mix:
Medicare only
Medicaid only/Medicaid and private
Medicare and Medicaid
Private only
Medicare and private
Uninsured/no payer
Other

Primary diagnosis is collected as ICD-9 or ICD-10 code. Categories used in case mix:
Dementia
Parkinson's and other degenerative diseases
Solid tumor cancers
Acute Lymphomas
Chronic Lymphomas
Acute Lymphoid or Myeloid or Monocytic Leukemia
Chronic Lymphoid or Myeloid or Monocytic Leukemia
All other blood cancers
Other cancers
Non-cancer end organ
ESRD
Pneumonias and other infectious lung diseases
Cerebrovascular accident/stroke
Other

Length of final episode of hospice care is calculated from date of final admission and date of death. Categories used in case mix:
2–5 days
6–12 days
13–29 days
30–80 days
81 days or more

Language of survey administration (English, Spanish, or Chinese) is combined with respondent's home language in case mix to form these categories:
Spanish survey language
Home language is Spanish but survey language was not Spanish

Chinese survey language
Home language is Chinese but survey language was not Chinese
All others

The case-mix adjustment uses a regression methodology, also called covariance adjustment. If data are missing for an adjuster variable, the missing value should be imputed using the hospice-specific mean for that variable.

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.
Provided in response box S.15a

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
Please see S.14 for a list of case-mix adjustment variables and their recommended parameterization.

S.16. **Type of score:**
Other (specify):
If other: 1. Top-box score 2. Case-mix adjusted score

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.)*
Top Box Score Calculation:

1) Identify target respondent population (i.e., primary caregivers of hospice patients who died while receiving hospice care from a given hospice in a given month)
2) Identify any exclusions from the respondent population (as described above in S.10)
3) Score each item using top box method, possible values of 0 or 100
4) Calculate mode adjusted top box scores for each item.
5) Calculate case-mix adjusted top box scores for each item for each hospice; case-mix adjustment is a linear regression based approach that adjusts for all variables listed in S.14. Specifically, a regression model predicting item scores is fit using the case-mix adjustor variables and fixed effects for hospices. Adjusted hospice means are then calculated (e.g., using LSMEANS in SAS).
6) Top-box scores are averaged across the items within each multi-item measure, weighting each item equally. If data are missing for a respondent for an item(s) within a multi-item measure, the respondent's answers to other items within the measure are still used in the calculation of multi-item measure scores. (Please see S.22 below for more details).

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.
The CAHPS Hospice Survey Quality Assurance Guidelines (available at http://www.hospicecahpssurvey.org/Content/QualityAssurance.aspx) specify guidance for sampling. Specifically, hospices with 50 to 699 survey-eligible decedents/caregivers in the prior year are required to survey all cases (conduct a census). Hospices with 700 or more survey-eligible decedents/caregivers in the prior year are required to survey a minimum sample of 700 using an equiprobable approach (simple random sampling) and may conduct a census, if desired. If a hospice chooses to survey more than the required sample of 700 decedents/caregivers, all data collected must be submitted to the CAHPS Hospice Survey Data Warehouse.

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.
The CAHPS Hospice Survey Quality Assurance Guidelines (available at http://www.hospicecahpssurvey.org/Content/QualityAssurance.aspx) provide detailed guidance regarding data collection procedures using Mail Only, Telephone Only, or Mixed Mode (mail plus telephone follow up) administration. Briefly:

- Users need to choose a data collection protocol that maximizes the survey response rate at an acceptable cost. Some sponsors, as well as researchers conducting field tests, have found that Mixed Mode (mail with telephone follow-up) yields the highest response rates: results from CAHPS mode experiments and national implementations indicate that Mixed Mode can add 10 to 15 percentage points to the response rate.

This section provides an overview of the protocol for collecting responses for each of the three CMS-approved modes of data collection: Mail Only, Telephone Only, and Mixed Mode.

Mail Only

- Data collection for sampled decedents/caregivers must be initiated two months following the month of patient death. Survey vendors must send sampled caregivers a first questionnaire with a cover letter within the first seven days of the field period. A second questionnaire with a follow- up cover letter must be sent to all sampled caregivers who did not respond to the first questionnaire, approximately 21 calendar days after the first questionnaire mailing. Data collection must be closed out for a sampled caregiver by six weeks (42 calendar days) following the mailing of the first questionnaire.
- Mailings must include a personalized cover letter, a questionnaire, and a business reply envelope. The cover letters may be sent in both English and one of the official translations. The Mail Only mode of survey administration may be conducted in English, Spanish, Chinese, Russian, and Portuguese.
- Cover letters sent to respondents must be personalized with the decedent's name, the caregiver's name and the hospice's name. The letter must also provide a toll-free number for respondents to call if they have questions. The cover of the questionnaire must include a label indicating the name of the hospice, and if applicable, may include the specific hospice inpatient unit, acute care hospital or nursing home facility in which their family member or friend resided.
- To increase the likelihood that the respondent is the person within the sampled caregiver's household who is most knowledgeable about the decedent's hospice care, language must be included in the questionnaire, and optionally in the cover letter, clearly stating that the survey should be given to the person in the household who knows the most about the hospice care received by the decedent.

Telephone Only

- Data collection may be completed by telephone only. Outbound calling must be scheduled in a manner to ensure all cases have a first attempt within seven days of the start of the field period. Data collection must be closed out for a sampled caregiver by six weeks (42 calendar days) following the first call attempt. If it is known that the caregiver may be available in the latter part of the 42 calendar day data collection time period (e.g., caregiver is on vacation the first two or three weeks of the 42 calendar day field period and there would be an opportunity to reach the caregiver closer to the end of the field period), then survey vendors must use the entire field time period to schedule telephone calls.
- Survey vendors must attempt to reach each and every caregiver in the sample. Telephone call attempts are to be made between the hours of 9 AM and 9 PM respondent time. Repeated attempts must be made until the caregiver is contacted, found ineligible or five attempts have been made. After five attempts to contact the caregiver have been made, no further attempts are to be made.
- Telephone data collection is permitted in both English and Spanish. English must be the default language in the continental U.S. and Spanish must be the default language in Puerto Rico. Survey vendors are provided standardized telephone scripts in both English and Spanish for CAHPS Hospice Survey administration.
- If, during a telephone attempt, the sampled caregiver indicates that someone within the household is more knowledgeable about the hospice care that the decedent received, the more knowledgeable person may be a proxy respondent. If a sampled caregiver indicates that he or she never oversaw, was not involved in, or is not knowledgeable about the hospice care provided to the decedent, interviewers may ask if someone else in the household is knowledgeable about the decedent's hospice care. If such a person exists, he or she may be a proxy respondent. Interviewers must not accept individuals outside of the sampled caregiver's household as proxy respondents.
- Consistent monitoring of interviewers' work is essential to achieve standardized and accurate results. Properly trained and supervised interviewers ensure that standardized, non-directive interviews are conducted. Interviewers conducting the telephone survey must be trained prior to interviewing. Survey vendors must monitor at least 10 percent of all CAHPS Hospice Survey

interviews, dispositions and call attempts in their entirety (both English and Spanish) through silent monitoring of interviewers using the electronic telephone interviewing system software or an alternative system.

Mixed Mode

- Data collection for sampled decedents/caregivers must be initiated two months following the month of patient death within the first seven days of the field period. Survey vendors must send sampled caregivers a questionnaire with a cover letter within the first seven days of the field period, then beginning approximately 21 calendar days after mailing the questionnaire conduct a maximum of five telephone attempts to non-respondents. Data collection must be closed out for a sampled caregiver by six weeks (42 calendar days) following the mailing of the questionnaire.
- Reversing the protocol (telephone attempts followed by mail attempt) is not allowed.
- If a caregiver completes the CAHPS Hospice Survey via the telephone and a questionnaire is subsequently returned by the caregiver, the survey vendor must use the telephone CAHPS Hospice Survey responses since they were completed first. Alternately, if a questionnaire is returned after calling has begun, no further telephone attempts should be made, and the survey vendor must use the responses in the returned mail survey.
- All of the guidelines for both the Mail Only and the Telephone Only survey administration apply to the Mixed Mode survey administration.

The CAHPS Hospice Survey Quality Assurance Guidelines offer the following suggestions for maximizing response rates:

- For Mail Only and Mixed Mode survey administration, vendors must perform address updates for missing or incorrect information, including working with client hospices to obtain the most current caregiver contact information, using the National Change of Address and the United States Postal Service CASS Certified Zip+4 software, and using other means such as commercial software and internet search options to locate current addresses.
- For Telephone Only and Mixed Mode survey administration, vendors must use commercial software or other means to update telephone numbers provided by the hospice for sampled caregivers. This includes running update program software against the sample file just before or after uploading data to survey management systems, utilizing commercial software, Internet directories and/or directory assistance, and contacting the hospice to request updated telephone numbers.
- For Mail Only and Mixed Mode survey administration, send all mailings with first class postage or indicia to ensure delivery in a timely manner and to maximize response rates, as first class mail is more likely to be opened.
- For Telephone Only and Mixed Mode survey administration, make telephone attempts at various times of the day, on different days of the week, and in different weeks to maximize the probability that the survey vendor will contact the caregiver.
- Survey vendors should make every reasonable effort to achieve optimal telephone response rates, such as thoroughly familiarizing interviewers with the study purpose, carefully supervising interviewers, retraining those interviewers having difficulty enlisting cooperation, and re- contacting reluctant respondents at different times until the data collection protocol is completed.

S.22. Missing data (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Top-box scores for individual items are computed individually for each hospice and then averaged across items to calculate the final multi-item measure score. All responses to all items within multi-item measures are used to calculate measure scores; no values are imputed or deleted when calculating hospice-level multi-item measure scores.

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.
CAHPS Hospice Survey

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

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.)*
Please see Section S.6 for a description of how multi-item measure scores are calculated for the CAHPS Hospice Survey.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
CAHPS_Hospice_Survey_Measure_Testing_Form_REV_3-15-16.docx

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

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: CAHPS Hospice Survey, Version 2.0
Date of Submission: 2/29/2016
Type of Measure: PRO-PM

☐ 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.

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 that influence the measured outcome (but not factors related to disparities in care or the quality of care) 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. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.

16. 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: CAHPS Hospice Survey Data Warehouse	☒ other: CAHPS Hospice Survey Data Warehouse

*Metrics presented throughout are derived from analysis of the CAHPS Hospice Survey data collected during national implementation of that survey in 2015.

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*).

Existing data from the CAHPS Hospice Survey Data Warehouse were analyzed. The Warehouse contains survey response data for all hospices eligible to participate in national implementation of the CAHPS Hospice Survey. Eligible hospices authorize a CMS-approved survey vendor to administer the CAHPS Hospice Survey and submit data to the Warehouse in accordance with the standard protocols and deadlines outlined in the *CAHPS Hospice Survey Quality Assurance Guidelines* (available at http://www.hospicecahpssurvey.org/Content/QualityAssurance.aspx). Data submissions to the Warehouse are used to determine hospices' compliance with CMS requirements for national implementation of the CAHPS Hospice Survey, and will be used by CMS to calculate hospices' official scores for the purpose of public reporting once public reporting is introduced (timeline to be determined).

1.3. What are the dates of the data used in testing? Quarter 1 and Quarter 2 2015 (data collected in April through September 2015 regarding care experiences of patients who died while receiving hospice care in January through June 2015).

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: hospice program/agency	☒ other: hospice program/agency

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*)

The measured entity is a hospice program or agency. Hospice programs or agencies (hereafter, "hospices") are differentiated from one another by their CMS Certification Numbers (CCNs). Hospice-level survey results are calculated across the primary caregiver respondents within a hospice.

Data for these analyses were submitted to the CAHPS Hospice Survey Warehouse on behalf of hospices by CMS-approved survey vendors for Quarter 1 and Quarter 2, 2015 (data collected in April through September 2015 regarding care experiences of patients who died while receiving hospice care in January through June 2015). Throughout this submission, analyses were restricted using a cumulative rule of at least five respondents per quarter. Analyses using both quarters of data include 2,267 hospices with at least ten respondents combined across the Quarter 1 and Quarter 2, 2015. Hospices with fewer than 10 completed surveys were excluded. Analyses using solely data from Quarter 2 data include the 2,512 hospices with at least five respondents in Quarter 2. The threshold of five respondents per quarter was chosen because such a threshold would predict that these hospices would have roughly 20 respondents with a full year of data. The highest appropriate skip rate among evaluative items on the CAHPS Hospice Survey was 60 percent, and thus 20 respondents per year would be expected to lead to 11 responses per evaluative item, the CMS threshold for public reporting.

Table 1.5 below shows the distribution of characteristics for hospices that had 10 or more respondents in Quarter 1 and Quarter 2 combined; characteristics were derived from analyses of the September 2015 Medicare Provider of Services File and Medicare hospice claims data from 2013, the most recent data available at the time of the analyses. Thirty-five percent of hospices were non-profit and 17.0 percent were located in rural areas. Hospices with fewer than 50 survey-eligible decedents/caregivers per year could apply for exemption from participation in national implementation; therefore, hospices that were this small as of 2013 compose a small proportion (4.9%) of hospices in the national implementation data set. Eighty-two percent of hospices submitting data in Quarter 1 or Quarter 2 had between 50 and 699 survey-eligible decedents/caregivers (as of 2013), and if they maintained that size in 2014, would have been required to conduct a census of caregivers. The remaining 13.2 percent of hospices had 700 or more survey-eligible decedents/caregivers (as of 2013), and if they maintained that size in 2014, could have elected to conduct a simple random sample of 700 or more to comply with CAHPS Hospice Survey national implementation requirements.

Table 1.5. Characteristics of Hospices Participating in National Implementation of the CAHPS Hospice Survey, N=2,267 with 10 or more respondents in Quarter 1 and Quarter 2 2015

Hospice Characteristic	Hospices Submitting Q1 or Q2 CAHPS Hospice Survey Data n (%)
Region	
Northeast	346 (15.3%)
South	836 (36.9%)
Midwest	632 (27.9%)
West	437 (19.3%)
U.S. Territories (PR, GU, & VI)	14 (0.6%)
Urban/Rural	
Urban	1,880 (83.0%)
Rural	385 (17.0%)
Size (# survey-eligible decedents/caregivers, Medicare hospice claims, 2013)	
Fewer than 50	109 (4.9%)
50 to 699	1,833 (81.9%)
700 or more	295 (13.2%)
Ownership	
Non-profit	792 (35.0%)
For-profit	1,151 (50.8%)
Government	322 (14.2%)
Rate of live discharge, 2013 Medicare claims data	
Less than 10%	388 (17.3%)
10% to less than 20%	1,287 (57.5%)
20% to less than 30%	418 (18.7%)
30% to less than 40%	110 (4.9%)
40% or higher	34 (1.5%)
Mean length of stay in days, 2013 Medicare claims data	
Less than 20	31 (1.4%)
20-39	791 (35.4%)
40-59	1,115 (49.8%)
60-79	279 (12.5%)
80+	21 (0.9%)

Source: CMS September 2015 Provider of Services file and 2013 hospice claims.

NOTE: Totals may not add to 2,267 due to missing data.

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*)

The CAHPS Hospice Survey is completed by the primary caregiver (i.e., family member or friend) of the patient who died while receiving hospice care ("decedent").

The CAHPS Hospice Survey is designed to be completed by the primary caregiver. The primary caregiver is the family member or friend most knowledgeable regarding the hospice care received by the patient who died while receiving hospice care ("decedent"). Throughout this application, the term "caregiver" refers to this primary caregiver. Primary caregivers are identified from hospice administrative records; one primary caregiver is selected to respond to the survey for a given decedent. Decedents/caregivers are eligible for inclusion in the CAHPS Hospice Survey if:

- Decedents were age 18 or over
- Decedents died at least 48 hours following last admission to hospice care
- Decedents had a caregiver of record
- Decedents had a caregiver with a U.S. or U.S. Territory home address
- Decedents had a caregiver other than a nonfamilial legal guardian.

Decedents or caregivers who requested that they not be contacted (those who signed no publicity requests while under the care of hospice or otherwise directly requested not to be contacted) are also excluded.

As noted in Question 1.5, hospices with 50 to 699 survey-eligible decedents/caregivers in the prior year are required to survey all cases (conduct a census). Hospices with 700 or more survey-eligible decedents/caregivers in the prior year are required to survey a minimum sample of 700 using an equiprobable approach (simple random sampling) and may conduct a census, if desired. If a hospice chooses to survey more than the required sample of 700 decedents/caregivers, all data collected must be submitted to the CAHPS Hospice Survey Data Warehouse.

The *CAHPS Hospice Survey Quality Assurance Guidelines* provide additional information on sampling, and are available at: http://www.hospicecahpssurvey.org/Content/QualityAssurance.aspx.

Table 1.6 shows descriptive characteristics of the 116,141 caregiver respondents for whom survey responses were submitted to the CAHPS Hospice Survey Data Warehouse in Quarter 1 and Quarter 2 2015 from the hospices described in Table 1.5. The mean age of decedents was 81.2 (Table 1.6); 4.2 percent were black, and 4.1 percent were Hispanic. For more than half of decedents (57.9%), the last setting of hospice care was a home or assisted living facility; last location was a nursing home for one in five decedents, a hospice freestanding IPU for 16.1 percent, and an acute care hospital for 5.1 percent. One-quarter of decedents had a length of stay in hospice of less than one week, while one in ten decedents had a stay of six months or more.

Table 1.6. Characteristics of Decedents and Caregiver Respondents in National Implementation of the CAHPS Hospice Survey, Quarter 1 and Quarter 2 2015

Characteristic	n (%)
Decedent Characteristics	
Gender	
Male (%)	53,471 (46.3%)
Age (mean, SD)	81.2 (12.2)
Race/Ethnicity	

Characteristic	n (%)
White	78,669 (83.2%)
Black	3,996 (4.2%)
Hispanic	3,882 (4.1%)
Asian or Pacific Islander	926 (1.0%)
Multiracial or other	7,124 (7.5%)
Final Setting of Care	
Home	55,906 (57.9%)
Nursing Home	19,341 (20.0%)
Acute Care Hospital	4,881 (5.1%)
Hospice Inpatient Unit	15,538 (16.1%)
Other	847 (0.9%)
Length of final episode of hospice care	
Less than 1 week	28,864 (25.0%)
1 to less than 2 weeks	20,326 (17.6%)
2 to less than 4 weeks	18,472 (16.0%)
1 to less than 2 months	16,570 (14.4%)
2 to less than 4 months	13,415 (11.6%)
4 to less than 6 months	5,725 (5.0%)
6 or more months	11,916 (10.3%)
Primary Diagnosis	
Dementia	13,832 (13.1%)
Parkinson's and other degenerative diseases	2,469 (2.3%)
Solid tumor cancers	22,463 (21.3%)
Acute Lymphomas	230 (0.2%)
Chronic Lymphomas	810 (0.8%)
Acute Lymphoid or Myeloid or Monocytic Leukemia	408 (0.4%)
Chronic Lymphoid or Myeloid or Monocytic Leukemia	277 (0.3%)
All other blood cancers	751 (0.7%)
Other cancers	14,980 (14.2%)
Non-cancer end organ, except ESRD	25,926 (24.6%)
ESRD	1,667 (1.6%)

Characteristic	n (%)
Pneumonias and other infectious lung diseases	3,284 (3.1%)
Cerebrovascular accident/Stroke	7,821 (7.4%)
Other	10,483 (10.0%)
Highest grade or level of school	
8th grade or less	10,054 (9.1%)
Some high school but did not graduate	11,010 (10.0%)
High school graduate or GED	42,431 (38.4%)
Some college or 2-year degree	23,925 (21.7%)
4-year college graduate	10,963 (9.9%)
More than 4-year college degree	12,048 (10.9%)
Payer for healthcare services	
Medicare only	80,506 (72.9%)
Medicaid only or Medicaid/Private	2,002 (1.8%)
Medicare/Medicaid	4,680 (4.2%)
Private only	5,160 (4.7%)
Private/Medicare	3,905 (3.5%)
Uninsured/no payer	453 (0.4%)
Other	13,787 (12.5%)
Caregiver Characteristics	
Sex	
Male (%)	31,660 (27.9%)
Age	
18 to 24	132 (0.1%)
25 to 34	1039 (0.9%)
35 to 44	3225 (2.8%)
45 to 54	13,527 (11.9%)
55 to 64	34,426 (30.2%)
65 to 74	34,290 (30.1%)
75 to 84	19,699 (17.3%)
85 or older	7,763 (6.8%)
Highest grade or level of school	
8th grade or less	1,543 (1.4%)
Some high school but did not graduate	4,712 (4.2%)

Characteristic	n (%)
High school graduate or GED	30,602 (27.2%)
Some college or 2-year degree	36,137 (32.1%)
4-year college graduate	18,295 (16.3%)
More than 4-year college degree	21,197 (18.8%)
Home Language	
English	111,300 (98.2%)
Spanish	1,360 (1.2%)
Chinese	133 (0.1%)
Some other language	558 (0.5%)
Relation to decedent (decedent was caregiver's):	
Spouse or partner	46,375 (40.3%)
Parent	51,051 (44.4%)
Mother-in-law or father-in-law	2,539 (2.2%)
Grandparent	981 (0.9%)
Aunt or Uncle	2,692 (2.3%)
Sister or Brother	4,963 (4.3%)
Child	2,115 (1.8%)
Friend	1,864 (1.6%)
Other	2,421 (2.1%)

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.

The following analyses were conducted with data from Quarter 1 and Quarter 2, 2015: descriptives (Questions 1.5 and 1.6), risk adjustment (Question 2b4), and missing data analysis (Question 2b7).The following analyses were restricted to Quarter 2 only: reliability testing (Question 2a.2), validity (Question 2b2), and identification of statistically significant and meaningful differences in performance (Question 2b5). This difference in analytic datasets is due to the fact that Quarter 1 data were collected during a "dry run" period in which hospices and survey vendors were becoming familiar with new data collection requirements, and in which hospices were required to participate for only one month. There were potential concerns about data quality from this period. Analyses using both quarters were conducted after a statistical evaluation indicated that results were consistent across quarters. For example, we examined risk adjustment in each quarter and then compared results across quarters. We found no evidence of differences in risk adjustment quantities or recommendations between quarters and thus quarters were combined to improve precision.

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*)

To assess measure reliability, we calculate:

1) The internal consistency of the multi-item measures, measured by Cronbach's alpha. Cronbach's alpha, a 0 to 1 index of the magnitude of internal consistency for multi-item measures, increases with the number of items in a multi-item measure and their average correlation with each other. Higher values indicate more precise measurement of the underlying construct that the multi-item measure is intended to represent. Alpha coefficients of at least 0.70 are usually considered acceptable for making assessments at the person level (Nunnally and Bernstein, 1994); a somewhat lower threshold is acceptable for making assessments at the entity level, as is intended for the CAHPS Hospice Survey.
2) The person-level Pearson item-total correlation, the correlation between a given item and the total multi-item measure score with the given item removed. This metric reflects how related each item is to all other items in the measure.
3) Inter-unit (i.e., hospice-level) reliability, which refers to the degree to which measure scores are able to precisely distinguish between the performances of hospices. 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, multi-item measure reliability greater than 0.70 is commonly considered adequate (Hargraves, Hays & Cleary, 2003; Nunnally 1994).

Citations:

Allen M, Yen W. *Introduction to Measurement Theory*. Monterey, CA: Brooks/Cole;1979.
Nunnally JC, Bernstein IH. Psychometric Theory. New York: McGraw Hill; 1994.
Raudenbush SW, Bryk AS. Hierarchical Linear Models. 2nd ed. Thousand Oaks, CA: Sage; 2002.

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*)

Table 2a2.3a shows the Cronbach's alpha for each multi-item measure, as well as the item-total Pearson correlations (i.e., the correlation between each item and the multi-item measure with the given item removed). Table 2a2.3b shows the hospice-level reliability for each measure when n = 200 completed surveys.

Table 2a2.3a. Cronbach's Alpha Reliability Coefficients and Item-total Pearson Correlations for CAHPS Hospice Survey Measures, Quarter 2 2015

Multi-Item Measures and items	Item-total correlation

Hospice team communication (alpha = 0.84)	
How often did the hospice team listen carefully to you when you talked with them about problems with your family member's hospice care?	0.82
While your family member was in hospice care, how often did the hospice team listen carefully to you?	0.71
While your family member was in hospice care, how often did the hospice team explain things in a way that was easy to understand?	0.71
While your family member was in hospice care, how often did the hospice team keep you informed about your family member's condition?	0.68
While your family member was in hospice care, how often did the hospice team keep you informed about when they would arrive to care for your family member?	0.53
While your family member was in hospice care, how often did anyone from the hospice team give you confusing or contradictory information about your family member's condition or care?	0.32
Getting timely care (alpha = 0.60)	
While your family member was in hospice care, when you or your family member asked for help from the hospice team, how often did you get help as soon as you needed it?	0.44
How often did you get the help you needed from the hospice team during evenings, weekends, or holidays?	0.44
Treating your family member with respect (alpha = 0.69)	
While your family member was in hospice care, how often did the hospice team treat your family member with dignity and respect?	0.59
While your family member was in hospice care, how often did you feel that the hospice team really cared about your family member?	0.59
Getting emotional and religious support (alpha = 0.66)	
In the weeks after your family member died, how much emotional support did you get from the hospice team?	0.54
While your family member was in hospice care, how much emotional support did you get from the hospice team?	0.46
Support for religious or spiritual beliefs includes talking, praying, quiet time, or other ways of meeting your religious or spiritual needs. While your family member was in hospice care, how much support for your religious and spiritual beliefs did you get from the hospice team?	0.46
Getting help for symptoms (alpha = 0.74)	
How often did your family member receive the help he or she needed from the hospice team for feelings of anxiety or sadness?	0.62
How often did your family member get the help he or she needed for trouble with constipation?	0.57
How often did your family member get the help he or she needed for trouble breathing?	0.53
Did your family member get as much help with pain as he or she needed?	0.52
Getting hospice care training (alpha = 0.86)	
Did the hospice team give you enough training about how to help your family member if he or she had trouble breathing?	0.69
Did the hospice team give you enough training about what side effects to watch for from pain medicine?	0.76

Did the hospice team give you enough training about what to do if your family member became restless or agitated?	0.68
Did the hospice team give you enough training about if and when to give more pain medicine to your family member?	0.65
Side effects of pain medicine include things like sleepiness. Did any member of the hospice team discuss side effects of pain medicine with you or your family member?	0.66

Table 2a2.3b. Hospice-Level Reliability for CAHPS Hospice Survey Measures, Quarter 2 2015

Multi-item or single-item measure	Intraclass Correlation Coefficient (ICC)	Hospice reliability @ N=200 per hospice
Multi-Item Measures		
Getting Hospice Care Training (5-items)	.017	.78
Hospice Team Communication (6-items)	.013	.72
Getting Timely Care (2-items)	.012	.71
Getting Emotional and Religious Support (3-items)	.011	.70
Getting Help for Symptoms (4-items)	.008	.62
Treating Family Member with Respect (2-items)	.008	.61
Global Measures		
Rating of Hospice	.011	.68
Willingness to Recommend	.017	.78

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?*)

Multi-item measures with the highest Cronbach's alpha reliability (internal consistency) coefficients were *Getting Hospice Care Training* (alpha = 0.86)*, Hospice Team Communication* (alpha = 0.84)*,* and *Getting Help for Symptoms* (alpha = 0.74; Table 2a2.3a). Cronbach's alpha is closely related to the number of items in a multi-item measure; therefore, it is not suprising that internal consistency is somewhat lower for the shorter multi-item measures *Getting Timely Care* (two items)*, Getting Emotional and Religious Support* (three items) *and Treating Your Family Member with Respect* (two items). Although these shorter measures have internal consistencies less than 0.70, they are designed for use at the hospice-level (not the person-level, for which that 0.70 threshold is generally applied).

As indicated by the item-total correlations presented in Table 2a2.3a – all of which are smaller than the Cronbach's alpha for their multi-item measures – there are no multi-item measures for which deletion of an item or items would result in a higher Cronbach's alpha for the multi-item measure.

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 Timeline Care, Getting Emotional and Religious Support,* and *Willingness to Recommend the Hospice*; Table 2a2.3b*)*. At a reliability of 0.68, *Rating of Hospice* is close to achieving the adequate reliability threshold. The remaining two measures, *Getting Help for Symptoms* and *Treating Family Member with Respect*, exhibit hospice-level reliabilities that

suggest somewhat less ability to distinguish between hospices' performance (0.62 and 0.61, respectively).

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)*

At the hospice level, we examined the relationships between each multi-item measure's top box score and the top box score for the global measure of "Using any number from 0 to 10, where 0 is the worst hospice care possible and 10 is the best hospice care possible, what number would you use to rate your family member's hospice care?" We used linear regressions that predicted the global measure from the multi-item measures while adjusting for case mix and mode of survey administration. These analyses are designed to look for evidence of construct validity – that is, that caregivers' responses to the multi-item measures are related to their global ratings of hospice care.

We also examined Pearson correlations among the multi-item measures to assess the magnitude of association between the measures, as a proxy for understanding the extent to which they measure different constructs. Moderate intercorrelations are to be expected, given that all measures assess aspects of care experiences. Very high intercorrelations may indicate that measures should be combined, while low intercorrelations may be seen as evidence of divergent validity (i.e., that measure content that we would not expect to be related is in fact not related).

In addition, we conducted validity testing looking at Willingness to Recommend as an outcome. (Willingness to Recommend is also a single item measure.) The results are shown in Table 2b2.3a. The addition of Willingness to Recommend to the table is highlighted for ease of use.

The associations of Willingness to Recommend and the Multi-Item measures are similar to those for Rating of Hospice, with Hospice Team Communication exhibiting the strongest relationship to Willingness to Recommend and Getting Hospice Care Training the weakest. As is usually observed, the associations between composites and Willingness to Recommend are somewhat weaker than those between composites and Rating of Hospice, since Willingness to Recommend may be affected by factors such as geographic location of the respondent.

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

Table 2b2.3a. Results of Analysis Assessing Construct Validity Between Multi-Item Measures and Rating of Hospice, Q2 2015

	β (Standard Error)
Hospice Team Communication	0.853 (0.005)*

Treating Family Member with Respect	0.812 (0.006)*
Getting Emotional and Religious Support	0.746 (0.007)*
Getting Help for Symptoms	0.609 (0.006)*
Getting Timely Care	0.517 (0.004)*
Getting Hospice Care Training	0.472 (0.004)*

*p<.001

Table. 2b2.3a. Continued Results of Analysis Assessing Construct Validity Between Composites and Willingness to Recommend Hospice, Q2 2015

	β (Standard Error)
Hospice Team Communication	0.781 (0.005)*
Treating Family Member with Respect	0.776 (0.006)*
Getting Emotional and Religious Support	0.707 (0.008)*
Getting Help for Symptoms	0.539 (0.006)*
Getting Timely Care	0.464 (0.004)*
Getting Hospice Care Training	0.432 (0.004)*

*p<.001

Table 2b2.3b. Pearson Correlations between CAHPS Hospice Survey Multi-Item Measures, Q2 2015

	Hospice Team Communication	Treating Family Member with Respect	Getting Hospice Care Training	Getting Timely Care	Getting Emotional and Religious Support	Getting Help for Symptoms
Hospice Team Communication	1					
Treating Family Member with Respect	.68	1				
Getting Hospice Care Training	.55	.39	1			
Getting Timely Care	.61	.52	.36	1		
Getting Emotional and Religious Support	.47	.45	.31	.35	1	
Getting Help for Symptoms	.56	.48	.47	.46	.37	1

Note: All two-tailed p-values < .001.

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 in Table 2b2.3a demonstrate that three CAHPS Hospice Survey multi-item measures strongly related to respondents' rating of hospice (*Hospice Team Communication, Treating Family Member with Respect,* and *Getting Emotional and Religious Support* with β = 0.853, β = 0.812 and β = 0.746, respectively), while the remaining multi-item measures are moderately related (range: β = 0.472 to 0.609).These results demonstrate the construct validity of the proposed measures.

Table 2b2.3b indicates that CAHPS Hospice Survey measures are generally moderately intercorrelated. No single measure or cluster of measures stands out as being very highly or weakly intercorrelated, suggesting that they measure unique but related constructs. Intercorrelations are highest between *Hospice Team Communication* and *Treating Family Member with Respect* (r = 0.68) because these two measures assess forms of communication. The somewhat lower intercorrelations between some of the multi-item measures (e.g., *Getting Hospice Care Training* and *Getting Emotional and Religious Support*, r =0.31) provides some evidence of divergent validity (i.e., that measures that we would not expect to be related are in fact not related).

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*)

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*)

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*)

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

☐ **Other,**

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.

N/A

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient 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 and not related to disparities*)

Previous research, on both CAHPS surveys and other types of surveys, has identified respondent characteristics that are not under the control of the entities being assessed but tend to be related to survey responses. For example, individuals who are older, those with less education, and those in better overall and mental health generally tend to give more positive ratings and reports of care in Medicare CAHPS (Elliott, Swartz et al. 2001; Zaslavsky, Zaborski et al. 2001; Elliott, Zaslavsky et al. 2009). Hence, entities with disproportionate numbers of respondents with such characteristics (favorable case mix) are advantaged relative to those with a less favorable case mix. To ensure that comparisons between hospices reflect differences in performance rather than differences in patient and/or caregiver characteristics, hospice scores should be adjusted for variations of such characteristics across hospices. CMS has not yet publicly released case mix adjustment information for the CAHPS Hospice Survey. The analyses described here use Quarter 1 and Quarter 2 2015 data to identify an initial case-mix adjustment model; in keeping with longstanding practice in CAHPS initiatives, the the parameterization of the case-mix adjustment model will be re-evaluated periodically, and the coefficients of the model will be reestimated regularly, as new data become available.

We identified patient and caregiver characteristics as candidates for case-mix adjustment if they were available in hospice administrative or survey response data and were not within the hospice's control. For each potential case-mix adjustor, we examined (a) variation among hospices using intraclass correlation coefficients (ICCs), (b) bivariate and multivariate association with selected CAHPS Hospice Survey outcomes, (c) impact on adjustment, and (d) appropriate parameterization of adjustors. We selected outcomes for assessment that had substantial variation across hospices, different response scales, and addressed conceptually distinct aspects of hospice care experiences.

Citations:

Elliott, M. N., R. Swartz, J. Adams, K. L. Spritzer and R. Hays (2001). "Case-mix adjustment of the National CAHPS® Benchmarking Data 1.0: A violation of model assumptions?" Health Services Research 36(3): 555-574.

Elliott, M. N., A. M. Zaslavsky, E. Goldstein, W. Lehrman, K. Hambarsoomian, M. K. Beckett and L. Giordano (2009). "Effects of survey mode, patient mix, and nonresponse on CAHPS Hospital Survey scores." Hlth Serv Res 44(2): 501-508.

Zaslavsky AM, Zaborski LB, Ding L, Shaul JA, Cioffi MJ, Cleary PD. Adjusting performance measures to ensure equitable plan comparisons. Health Care Financing Review; 2001; 22(3): 109-126.

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

Intraclass correlation coefficients (ICCs) were used to assess variation in respondent and decedent characteristics among hospices (Table 2b4.4). Adjustors with ICCs greater than 0.012 were considered to have moderate or more variation across hospices, as an ICC of 0.012 approximately corresponds to hospice-level reliability of 0.70 or higher if n = 200 completes for each hospice. There was moderate variation in primary diagnosis, particularly for primary diagnoses of dementia or solid tumor cancers versus other diagnoses. There was substantial variation among hospices in response percentile, payer for hospice care (particularly for 'Medicare only' and 'Medicare and private' versus other payers), and survey language/respondent's home language (particularly for Spanish survey language versus other categories).

Table 2b4.4a. Hospice-Level Intraclass Correlation Coefficients of Potential Case-Mix Adjustors for CAHPS Hospice Survey Measures, Q1 and Q2 2015

Potential CMA	ICC estimate
Response percentile	0.2050
Decedent characteristic	
Age	
18-54	0.0033
55-64	0.0076
65-69	0.0036
70-74	0.0031
75-79	0.0025
80-84	0.0005
85-89	0.0019
90+	0.0252
Payer for hospice care	
Medicare only	0.4146
Medicaid only/Medicaid and private	0.0205
Medicare and Medicaid	0.1732
Private only	0.0458
Medicare and private	0.3611
Uninsured/no payer	0.1321
Other	0.6821
Primary diagnosis	
Dementia	0.0730
Parkinson's and other degenerative diseases	0.0098
Solid tumor cancers	0.0530
Acute Lymphomas	0.0060
Chronic Lymphomas	0.0040

Potential CMA	ICC estimate
Acute Lymphoid or Myeloid or Monocytic Leukemia	0.0024
Chronic Lymphoid or Myeloid or Monocytic Leukemia	0.0030
All other blood cancers	0.0037
Other cancers	0.0231
Non-cancer end organ	0.0326
ESRD	0.0119
Pneumonias and other infectious lung diseases	0.0266
Cerebrovascular accident/stroke	0.0357
Other	0.1139
Length of final episode of hospice care	
2-5 days	0.0229
6-12 days	0.0053
13-29 days	0.0032
30-80 days	0.0052
81 days or longer	0.0346
Respondent characteristic	
Age	
18-24	0.0006
25-34	0.0021
35-44	0.0033
45-54	0.0044
55-64	0.0037
65-74	0.0026
75-84	0.0048
85+	0.0041
Education	
Less than 8th grade	0.0200
Some HS	0.0134
HS	0.0347
Some College	0.0116
BA	0.0153
More than BA	0.0261
Relationship (decedent was respondent's:)	
Spouse or partner	0.0300
Parent	0.0192
Mother-in-law or father-in-law	0.0010
Grandparent	0.0009
Aunt or uncle	0.0032
Sister or brother	0.0021
Child	0.0024
Friend	0.0025
Other	0.0029

Potential CMA	ICC estimate
Survey language/respondent's home language	
Spanish survey language	0.5787
Home language is Spanish but survey language was not Spanish	0.0271
Chinese survey language	0.0020
Home language is Chinese but survey language was not Chinese	0.0120
All others	0.4304

We used bivariate and multivariate regression models to examine the predictive power of each candidate case-mix adjustor. Table 2b4.4.b presents results for the *Rating of Hospice* measure (one of the outcomes used for case-mix adjustment analyses). Several candidate case-mix adjustors were significantly predictive of *Rating of Hospice*. Higher response percentile and younger decedent age were associated with lower rating. For payer for hospice care, 'Medicaid only/Medicaid and private,' 'Medicare and Medicaid,' and 'private only' payers were associated with lower ratings compared with the reference group, 'Medicare only.' Compared with the reference group of dementia, having a primary diagnosis of 'other cancer' or ESRD was associated with a lower rating of care. Length of final episode of care was associated with worse ratings in the middle range (6 to 80 days) compared with the longest episodes of care (81+ days); ratings for the shortest final episodes of care (2 to 5 days) were not statistically significantly different than for the longest. Younger and more educated respondents rated care significantly worse. Ratings were significantly higher if the decedent was the respondent's spouse/partner or mother-in-law/father-in-law than if the decedent was the respondent's parent. Ratings were higher if the survey language was Spanish or the respondent's home language was in Spanish but the survey was not in Spanish, and lower if the home language was Chinese but the survey was not in Chinese, both compared to the reference group (primarily English speakers). Examining results for the multivariate model, response percentile, payer, primary diagnosis, length of final episode of hospice care, respondent's age and education, patient's relationship to caregiver, and language were still significantly associated with *Rating of Hospice*. The magnitude of some but not all associations was attenuated in multivariate models.

Table 2b4.4b. Summary of Relationship between Potential Case Mix Adjusters and Overall Rating of Hospice Applying Top Box Scoring, Q1 and Q2 2015

	Standardized Beta Coefficients (Standard Error)	
	Bivariate models	Multivariate model
Response percentile	-0.1850 (0.0246) ***	-0.1766 (0.0245) ***
Decedent characteristic		
Age		
18-54	-0.0689 (0.0175) ***	0.0143 (0.0237)
55-64	-0.0389 (0.0122) **	0.0199 (0.0176)
65-69	-0.0144 (0.0127)	-0.0014 (0.0155)
70-74	0.0194 (0.0114)	0.0198 (0.0137)
75-79	0.0192 (0.0106)	0.0201 (0.0122)
80-84	0.0013 (0.0095)	0.0093 (0.0106)
85-89	-0.0052 (0.0089)	0.0039 (0.0093)
90+ (ref)		
Payer for hospice care		
Medicare only (ref)		
Medicaid only/Medicaid and private	-0.0866 (0.0232) ***	-0.0660 (0.0254) **

Medicare and Medicaid	-0.1029 (0.0165) ***	-0.1110 (0.0166) ***
Private only	-0.0769 (0.0148) ***	-0.0450 (0.0177) *
Medicare and private	-0.0087 (0.0219)	-0.0071 (0.0218)
Uninsured/no payer	0.0156 (0.0483)	0.0305 (0.0488)
Other	-0.0387 (0.0149) **	-0.0266 (0.0153)
Primary diagnosis		
Dementia (ref)		
Parkinson's and other degenerative diseases	-0.0175 (0.0222)	-0.0170 (0.0223)
Solid tumor cancers	-0.0098 (0.0113)	-0.0082 (0.0119)
Acute Lymphomas	0.0019 (0.0676)	0.0092 (0.0674)
Chronic Lymphomas	0.0350 (0.0368)	0.0292 (0.0368)
Acute Lymphoid or Myeloid or Monocytic Leukemia	-0.0925 (0.0508)	-0.1009 (0.0508) *
Chronic Lymphoid or Myeloid or Monocytic Leukemia	-0.1255 (0.0614) *	-0.1362 (0.0612) *
All other blood cancers	-0.0148 (0.0379)	-0.0183 (0.0379)
Other cancers	-0.0175 (0.0122)	-0.0127 (0.0126)
Non-cancer end organ	-0.0066 (0.0108)	-0.0193 (0.0109)
ESRD	-0.0475 (0.0264)	-0.0693 (0.0264) **
Pneumonias and other infectious lung diseases	0.0047 (0.0199)	-0.0128 (0.0199)
Cerebrovascular accident/Stroke	0.0286 (0.0145) *	0.0192 (0.0145)
Other	0.0077 (0.0137)	0.0013 (0.0137)
Length of final episode of hospice care		
2-5 days	0.0286 (0.0094) **	0.0324 (0.0095) ***
6-12 days	-0.0029 (0.0095)	-0.0005 (0.0096)
13-29 days	-0.0363 (0.0095) ***	-0.0344 (0.0096) ***
30-80 days	-0.0438 (0.0095) ***	-0.0427 (0.0095) ***
81 days or longer (ref)		
Respondent characteristic		
Age		
18-24	-0.3401 (0.0872) ***	-0.3779 (0.0878) ***
25-34	-0.2387 (0.0316) ***	-0.2506 (0.0331) ***
35-44	-0.1192 (0.0185) ***	-0.1196 (0.0196) ***
45-54	-0.0317 (0.0102) **	-0.0383 (0.0107) ***
55-64 (ref)		
65-74	0.0595 (0.0077) ***	0.0549 (0.0084) ***
75-84	0.0427 (0.0090) ***	0.0221 (0.0114)
85+	-0.0169 (0.0127)	-0.0289 (0.0152)
Education		
Less than 8th grade	0.0106 (0.0266)	0.0164 (0.0268)
Some HS	0.0478 (0.0158) **	0.0541 (0.0158) ***

HS (ref)			
Some College	-0.0627 (0.0079) ***	-0.0606 (0.0079) ***	
BA	-0.1436 (0.0095) ***	-0.1393 (0.0097) ***	*p<0.05; **p<0.01, ***p<0.001
More than BA	-0.1729 (0.0092) ***	-0.1767 (0.0093) ***	
Relationship (decedent was respondent's:)			
Spouse or partner	0.0276 (0.0066) ***	-0.0250 (0.0106) *	
Parent (ref)			
Mother-in-law or father-in-law	0.0945 (0.0205) ***	0.0932 (0.0204) ***	
Grandparent	-0.0498 (0.0325)	0.0683 (0.0340) *	
Aunt or uncle	0.0098 (0.0200)	0.0052 (0.0199)	
Sister or brother	-0.0012 (0.0151)	-0.0388 (0.0173) *	
Child	0.0014 (0.0225)	-0.0337 (0.0263)	
Friend	-0.0143 (0.0240)	-0.0384 (0.0243)	
Other	-0.0508 (0.0211) *	-0.0634 (0.0214) **	
Survey language/respondent's home language			
Spanish survey language	0.1649 (0.0471) ***	0.1474 (0.0473) **	
Home language is Spanish but survey language was not Spanish	-0.0111 (0.0450)	-0.0140 (0.0450)	
Chinese survey language	-0.3920 (0.2772)	-0.3986 (0.2762)	
Home language is Chinese but survey language was not Chinese	-0.4383 (0.0915) ***	-0.3913 (0.0912) ***	
All others (ref)			

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*)

(Revised response) To investigate CMA's overall effect on each CAHPS Hospice Survey measure, we compared hospice-level estimates without adjustment versus hospice-level estimates after adjusting for case mix using the multivariate model. We calculated Kendall's tau, a measure of rank correlation, which expresses the proportion of hospice pairs whose relative rankings were reversed by adjustment, scaled from 1 for no changes to −1 for a complete reversal of rankings. A tau value near 0 would indicate very little correlation between the unadjusted and adjusted scores and a large tau value near 1 would indicate almost perfect correlation between the scores. A tau estimate equal to 1 would indicate that case-mix adjustment has no effect on the hospice-level scores, which would be concerning since case-mix adjustment is expected to have some effect. A tau estimate very close to -1 would indicate that case-mix adjustment almost completely re-ranked all hospices, which would also be concerning since case-mix adjustment would not be expected to have such a dramatic effect. Based on prior CAHPS work, tau for these types of measures is expected to be between 0.80 and 0.95. In addition, Kendall's tau is directly related to the proportion of pairs of hospices that would switch ordering as a consequence of case-mix adjustment. If Kendall's tau is denoted as K, a value of K would indicate that [(1-K)/2] % of hospice pairs switched rankings due to case–mix adjustment.

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

Table 2b4.5. Summary of Impact of Case-Mix Adjustment Variables on Rating of Hospice, Q2 2015

	Kendall's Tau Comparing Unadjusted and Case-Mix Adjusted Hospice-level Means
Composite Measures	
Hospice Team Communication	0.9014
Getting Timely Care	0.9164
Treating Family Member with Respect	0.9141
Support for Emotional and Religious Beliefs	0.9092
Getting Help for Symptoms	0.9045
Getting Hospice Care Training	0.8628
Global Measures	
Rating of Hospice	0.9248
Willingness to Recommend	0.9406

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

N/A

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

N/A

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

N/A

2b4.9. Results of Risk Stratification Analysis:

N/A

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*)

Results of the comparison between adjustments for the multivariate model versus the null model are shown in Table 2b4.5. Kendall's tau comparing scores between null and multivariate model adjustments for each measure are in the expected range, from 0.86 for *Getting Hospice Care Training* to 0.94 for *Willingness to Recommend.* This means that applying the full set of recommended adjustors, seven percent of hospice pairs would switch in terms of relative rankings for *Getting Hospice Care Training* and three percent of hospice pairs would switch in terms of relative rankings for *Willingness to Recommend*.

For the purposes of providing hospice-level scores for hospices in the primary phases of national implementation, the recommended case-mix adjustment model includes the following:

- response percentile
- decedent age
- payer for hospice care

- primary diagnosis
- length of final episode of hospice care
- respondent age
- respondent education
- relationship of decedent to caregiver
- language (five categories: survey language is Spanish, home language is Spanish but survey language is not Spanish, survey language is Chinese, home language is Chinese but survey language is not Chinese, everything else).

`

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)*

To examine the ability of CAHPS Hospice Survey measures to identify high or low performing hospices, we calculated the number and percentage of hospices that were significantly above or below the national hospice average for each multi-item and global measure. Since CMS does not intend to publicly report or score these measures until at least one year of data are available, and one year of data is not yet available, we used the existing data to approximate what these numbers and percentages would be expected to look like with a full year of data. Specifically, we used Quarter 2 national implementation data and calculated the expected mean scores and standard errors if each hospice had data over a full year. The expected number of respondents for a full year was calculated as four times the number of respondents in Quarter 2. All scores were adjusted for mode and case mix and were scored using top-box scoring. A two-sided alpha=0.05 level test was used to test for significance.

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*)

Table 2b5.2a Number of Hospices Significantly Above or Below the Expected 2015 National Hospice Average (Projection Based on Q2 2015)

Hospice Multi-Item Measure/Items	Count of Hospices Significantly Above	Count of Hospices Significantly Below	% Statistically Different from Hospice National Average
Hospice Team Communication	781	449	49%
Getting Timely Care	712	443	46%
Treating Your Family Member with Respect	864	290	46%
Getting Help with Symptoms	716	490	48%
Getting Emotional and Religious Support	850	330	47%
Getting Hospice Care Training	738	448	47%
Rating of Hospice	717	399	44%
Willingness to Recommend	754	413	46%

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?*)

Table 2b5.2a indicates that CAHPS Hospice Survey measures can discrimate between hospices' performance as compared to a national average. For each of multi-item and global measures, nearly half of all hospices score either significantly above or below the national average. Among hospice scores that were significantly above or below the average, the mean absolute difference between the hospice's score and the average was 10.3 points on a 0 to 100 scale. The range of the absolute difference was 1.2 to 66.6. For reference, analyses of data from the field test of the CAHPS Hospice Survey indicated that for each 3-point increase in a caregiver's *Rating of Hospice* on a 0 to100 scale, the odds that the caregiver definitely recommended that hospice increased by 70%.

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 criterion is directed 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).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of 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*)

Hospices participating in national implementation of the CAHPS Hospice Survey may choose from one of three modes of survey administration: Mail Only, Telephone Only, or Mixed Mode (mail with telephone follow-up). To assess the effect of mode on response rates, response patters, and determine whether survey mode adjustments were needed to fairly compare CAHPS Hospice Survey scores across hospices using different modes of administration, we conducted a randomized mode experiment, sampling 17,121 decedent/caregiver pairs from 59 large hospice programs in Quarter 1 and Quarter 2, 2015. Each hospice provided a sample of decedents/caregivers who met the eligibility criteria for national implementation of the CAHPS Hospice Survey. Within each hospice, we randomly assigned one-third of cases to each of the three modes of data collection. One vendor collected survey data for all hospices using a standardized data collection protocol. To evaluate possible survey mode effects, we conducted linear regression analysis predicting each of the CAHPS Hospice Survey outcomes (on a 0-100 scale) from survey mode (Mail Only as reference category), case-mix adjustors and hospice indicators, scoring the outcomes using the top box approach.

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*)

We found significant effects of survey mode on responses to several survey outcomes. Telephone Only respondents reported significantly worse experiences of care than Mail Only respondents for 10 of the 28 evaluative items on the survey ($p<0.05$); there were no outcomes for which the reverse was true. For example, Telephone Only respondents tended to give lower hospice ratings (regression coefficient for Telephone Only = -3.93 compared to Mail Only, $p<0.01$).

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of 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*)

To ensure fair comparisons across hospices, CAHPS Hospice Survey scores must be adjusted for mode of survey administration, which affects scores but is not related to quality of hospice care.

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*)

Survey Non-Response

The *CAHPS Hospice Survey Quality Assurance Guidelines* outlines the survey response rate calculation, as follows:

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

The Total Number of Completed Surveys is the total number of surveys for which the caregiver respondent answers at least 50 percent of the questions applicable to all decedents/caregivers. The Total Number of Surveys Fielded is the total sample, and Total Number of Ineligible Surveys is the total number of surveys for which it is determined that the decedent/caregiver did not meet the eligibility criteria outlined above in Section 1.6; the respondent has a language

barrier, mental/physical incapacity, or is institutionalized or deceased; or the respondent indicates that they were never involved in decedent care.

The following are **not** removed from the denominator of the response rate calculation: break-off surveys, refusals, non-response after maximum attempts at data collection, bad or no address or telephone number, incomplete caregiver or decedent name, and respondents' disavowal that the decedent received care from any hospice or the named hospice.

The *CAHPS Hospice Survey Quality Assurance Guidelines* provide advice for maximizing survey response rates, including:

- For Mail Only and Mixed Mode survey administration, vendors must perform address updates for missing or incorrect information, including working with client hospices to obtain the most current caregiver contact information, using the National Change of Address and the United States Postal Service CASS Certified Zip+4 software, and using other means such as commercial software and internet search options to locate current addresses.
- For Telephone Only and Mixed Mode survey administration, vendors must use commercial software or other means to update telephone numbers provided by the hospice for sampled caregivers. This includes running update program software against the sample file just before or after uploading data to survey management systems, utilizing commercial software, Internet directories and/or directory assistance, and contacting the hospice to request updated telephone numbers.
- For Mail Only and Mixed Mode survey administration, send all mailings with first class postage or indicia to ensure delivery in a timely manner and to maximize response rates, as first class mail is more likely to be opened.
- For Telephone Only and Mixed Mode survey administration, make telephone attempts at various times of the day, on different days of the week, and in different weeks to maximize the probability that the survey vendor will contact the caregiver.
- Survey vendors should make every reasonable effort to achieve optimal telephone response rates, such as thoroughly familiarizing interviewers with the study purpose, carefully supervising interviewers, retraining those interviewers having difficulty enlisting cooperation, and re- contacting reluctant respondents at different times until the data collection protocol is completed.

We assessed the association between survey nonresponse and several caregiver and decedent characteristics, including relationship of caregiver to the decedent, and decedent age at death, sex, race/ethnicity, payer for hospice care, final setting of care, length of final episode of hospice care, and primary diagnosis.

Item Non-Response

In 2b7.2, we present nonresponse to evaluative items among unit respondents. Specifically, we report the proportion of respondents that skipped each item *appropriately* (i.e., dictated by the survey's skip logic instructions directed them to do so), *inappropriately* (i.e., *not* dictated by the survey's skip logic instructions), as well as the total proportion of missing data for each evaluative item on the survey.

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*)

The overall CAHPS Hospice Survey response rate in Quarter 1 and Quarter 2 2015 was 32.22 percent. The vast majority of surveys (92.68 percent) were administered via Mail Only administration, with a response rate of 31.69 percent. Response rates were 33.32 percent for Telephone Only Mode and 43.69 percent for Mixed Mode. In addition to survey mode, the factors most strongly associated with the odds of response were the caregiver's relationship to the decedent, with spouses almost twice as likely as other relations to respond to the survey; decedent age, with caregivers of older decedents more likely to respond than younger decedents; and race/ethnicity, with caregivers of white decedents more than twice as likely to respond as caregivers of black and Hispanic decedents. In addition, caregivers of decedents with Medicaid were substantially less likely to respond than those with Medicare, and caregivers of decedents whose final setting of care was a nursing home, acute care hospital or other setting were less likely to respond than those whose final setting was at home. In addition, caregivers of decedents with shorter final episodes were less likely to respond than those with longer episodes.

Table 2b7.2a. CAHPS Hospice Survey Missing Data, Q1 and Q2 2015

CAHPS Hospice Survey Evaluative Item	% Missing due to Appropriate Skip	% Missing due to Inappropriate Skip	% Missing (Total)
Hospice Team Communication			
While your family member was in hospice care, how often did the hospice team keep you informed about when they would arrive to care for your family member? (Q6)	0	2.2	2.2
While your family member was in hospice care, how often did the hospice team explain things in a way that was easy to understand? (Q8)	0	1.1	1.1
How often did the hospice team listen carefully to you when you talked with them about problems with your family member's hospice care? (Q14)	57.2	3.2	60.3
While your family member was in hospice care, how often did the hospice team keep you informed about your family member's condition? (Q9)	0	1.3	1.3
While your family member was in hospice care, how often did the hospice team	0	2.9	2.9

CAHPS Hospice Survey Evaluative Item	% Missing due to Appropriate Skip	% Missing due to Inappropriate Skip	% Missing (Total)
listen carefully to you? (Q35)			
While your family member was in hospice care, how often did anyone from the hospice team give you confusing or contradictory information about your family member's condition or care? (Q10)	0	1.6	1.6
Getting Timely Care			
While your family member was in hospice care, when you or your family member asked for help from the hospice team, how often did you get help as soon as you needed it? (Q7)	0	2.2	2.2
How often did you get the help you needed from the hospice team during evenings, weekends, or holidays? (Q5)	33.1	2.1	35.2
Treating Family Member with Respect			
While your family member was in hospice care, how often did the hospice team treat your family member with dignity and respect? (Q11)	0	1	1
While your family member was in hospice care, how often did you feel that the hospice team really cared about your family member? (Q12)	0	1.1	1.1
Support for Emotional and Religious Beliefs			
While your family member was in hospice care, how much emotional support did you get from the hospice team? (Q37)	0	3.5	3.5
In the weeks after your family member died, how much emotional support did you get from the hospice team? (Q38)	0	5.7	5.7
Support for religious or spiritual beliefs includes talking, praying, quiet time, or other ways of meeting your religious or spiritual needs. While your family member was in hospice care, how much support for your religious and spiritual beliefs did you get from the hospice team? (Q36)	0	5.9	5.9
Getting Help for Symptoms			
Did your family member get as much help with pain as he or she needed? (Q16)	27.4	2.8	30.2
How often did your family member get the help he or she needed for trouble	45.2	3.4	48.7

CAHPS Hospice Survey Evaluative Item	% Missing due to Appropriate Skip	% Missing due to Inappropriate Skip	% Missing (Total)
breathing? (Q22)			
How often did your family member get the help he or she needed for trouble with constipation? (Q25)	54.8	4.5	59.2
How often did your family member receive the help he or she needed from the hospice team for feelings of anxiety or sadness? (Q27)	41.7	4.5	46.2
Getting Hospice Care Training			
Did the hospice team give you enough training about what side effects to watch for from pain medicine? (Q19)	9	4.3	13.3
Did the hospice team give you the training you needed about if and when to give more pain medicine to your family member? (Q20)	33.1	4.4	37.5
Did the hospice team give you the training you needed about how to help your family member if he or she had trouble breathing? (Q23)	58.5	3.4	62
Did the hospice team give you the training you needed about what to do if your family member became restless or agitated? (Q29)	38.5	4.2	42.7
Side effects of pain medicine include things like sleepiness. Did any member of the hospice team discuss side effects of pain medicine with you or your family member? (Q18)	9	3.4	12.4
Rating of Hospice			
Using any number from 0 to 10, where 0 is the worst hospice care possible and 10 is the best hospice care possible, what number would you use to rate your family member's hospice care? (Q39)	0	2	2
Willingness to Recommend			
Would you recommend this hospice to your friends and family? (Q40)	0	1.8	1.8

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)

Unit response rates for the CAHPS Hospice Survey are comparable to those observed in other facility-level experience of care surveys; for reference, national mean hospital-level response rate for HCAHPS was 30 percent during the same period (HCAHPS Online). Although our analyses indicate that response propensity varies by certain caregiver and decedent characteristics, previous work in other CAHPS settings has demonstrated that nonresponse weighting to account for potential bias is not needed after case-mix adjustment (see, for example, Elliott, Edwards et al. 2005 and Elliott, Zaslavsky et al. 2009). When patient-mix adjustment suffices to address nonresponse bias, it generally does so with greater statistical efficiency than nonresponse weighting, resulting in estimates of equal reliability and precision with smaller sample sizes than would be required with nonresponse weighting.

Across evaluative items, less than 6 percent of respondents inappropriately skipped items. Item missingness tended to be higher for respondents for whom the survey was longer (i.e., respondents who were eligible for more survey items due to skip logic) and for those items that appeared later in the survey instrument (data not shown). These findings suggest that it is unlikely that CAHPS Hospice Survey item results are biased due to systematic skipping of items by respondents. The CAHPS Hospice Survey employs skip logic to promote appropriate skipping among respondents who are not qualified to answer an item Rodriguez et al., 2009.

Citation:

Centers for Medicare & Medicaid Services. Summary of HCAHPS Survey Results, April 2014 to March 2015 Discharges. Online at: http://www.hcahpsonline.com/Files/December_2015_Summary_Analyses_Survey_Results.pdf.

Elliott, M. N., C. Edwards, J. Angeles and R. D. Hays (2005). "Patterns of unit and item non-response in the CAHPS® Hospital Survey." Hlth Serv Res **40**(6): 2096-2119.

Elliott, M. N., A. M. Zaslavsky, E. Goldstein, W. Lehrman, K. Hambarsoomian, M. K. Beckett and L. Giordano (2009). "Effects of survey mode, patient mix, and nonresponse on CAHPS Hospital Survey scores." Hlth Serv Res **44**(2): 501-508.

Rodriguez HP, Glahn Tv, Li A, Rogers WH, Safran DG. The effect of item screeners on the quality of patient survey data: a randomized experiment of ambulatory care experience measures. *Patient*. 2009 Jun 1;2(2):135-41.

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.**

Other

If other: Survey of caregivers of hospice patients.

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.**

This is an experience of care survey. The best measure of individual´s experience is to ask them. Thus a survey is the best data collection option.

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.

At the time of the field test we had 3 closely related questionnaires differentiated by setting. As a result of the field test we consolidated our questionnaire into a single instrument in order to increase the ease of administering the survey for vendors and to increase usability for respondents.

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)*.

No fees or licensing.

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)
Public Reporting Quality Improvement with Benchmarking (external benchmarking to multiple organizations)	

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

All uses are planned not current.

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?*)

This survey started national implementation in January 2015 with a dry run covering the first quarter (January, February and March). Monthly data collection started April 2015. CMS plans to publicly report data after at least four quarters of data have been submitted to the CAHPS Hospice Survey Data Warehouse (excluding the dry run). Data will be collected for quarters 2, 3 and 4 in 2015 and quarter 1 of 2016 before public reporting can begin. CMS anticipates that the data will be publicly reported on Medicare.gov in a format similar to the other "Compare" web sites such as Hospital Compare.

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.*)

National implementation started January, 2015. CMS has not finalized plans for public reporting. However, CMS anticipates collecting at least four quarters of data before public reporting can begin. CMS expects that public reporting is likely to begin in 2017.

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**

Initial endorsement request.

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.

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.**
The field test of this questionnaire did not reveal any unintended negative consequences. National implementation has, thus far, revealed no unintended negative consequences.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0208 : Family Evaluation of Hospice Care
1623 : Bereaved Family Survey

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?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
1623 Bereaved Family Survey´s target population is families of veterans. The CAHPS Hospice Survey targets primary caregivers of patients who died under hospice care without regard to veteran status.

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.)
0208 Family Evaluation of Hospice Care.
The Family Evaluation of Hospice Care Survey (FEHC) is maintained by the NHPCO. NHPCO operated a voluntary repository that provided hospice programs with national benchmarks for FEHC measures. With the national implementation of the CAHPS Hospice Survey, NHPCO has shut down the voluntary repository, with the exception of those hospice programs that do not meet CMS's minimum threshold for participation in the CAHPS Hospice Survey. Once CMS publishes national benchmarks for the CAHPS Hospice Survey, NHPCO is no longer planning to support the FEHC or the voluntary repository.

The FEHC was created nearly 20 years ago. The CAHPS Hospice Survey covers similar domains, but represents important methodological improvement in the response task, and is adjusted for case mix and mode. Additionally, more stringent survey administration guidelines are in place to permit public reporting of the survey results and valid comparison across hospice programs.

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 **Attachment:** Appendix_A-635920971414908558.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare and Medicaid Services
Co.2 **Point of Contact:** Debra, Dean-Whittaker, debra.dean-whittaker@cms.hhs.gov, 410-786-0848-
Co.3 **Measure Developer if different from Measure Steward:** Centers for Medicare and Medicaid Services
Co.4 **Point of Contact:** Debra, Dean-Whittaker, debra.dean-whittaker@cms.hhs.gov, 410---

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.
Technical Expert Panel for Development of Hospice CAHPS Survey.
David Casarett, MD, MA TEP Co-Chair Director of Hospice and Palliative Care and Associate Professor of Medicine, University of Pennsylvania Health System

Paul Cleary, PhD TEP Co-Chair Dean, Yale School of Public Health

Bradley Beukema, MS Consultant and Chaplain, Montgomery Hospice

Karen Mikula, RN, BSN, CPHQ Senior Director of Quality Initiatives, VITAS Innovative Hospice Care

Naomi Naierman, MPA President and Chief Executive Officer, American Hospice Foundation

Scott Shreve, DO National Director of Hospice and Palliative Care, Department of Veterans Affairs

Eugenia Smither, RN, CHC, CHP, CHE Corporate Compliance Officer/Vice President of Compliance and Quality Improvement, Hospice of the Bluegrass

Shoshanna Sofaer, DrPH Robert P. Luciano Professor of Health Care Policy, School of Public Affairs, Baruch College

Carol Spence, PhD Vice President, Research and Quality, National Hospice and Palliative Care Organization

John Thoma CEO, Hospice of Wake County, Hospice of Harnett County, Horizons Palliative Care, and Horizons Home Care

In addition, there was a TEP for National Implementation of the survey. A list of attendees follows:

Julie Brown Senior Study Director and Director of the Survey Research Group RAND Corporation

Diane Burkom Program Manager Battelle Memorial Institute Health and Analytics Unit

Teresa Craig Chief Executive Officer Suncoast Solutions

Brad Edwards Vice President and Deputy Area Director Westat

Donna Farley, PhD Adjunct Senior Health Policy Analyst RAND Corporation

Ellen Martin, RN Texas Association for Home Health and Hospice

Scott Shreve, DO National Director of Hospice and Palliative Care Department of Veterans Affairs

Eugenia Smither, RN, CHC, CHP, CHE Corporate Compliance Officer/Vice President of Compliance and Quality Improvement Hospice of the Bluegrass

Carol Spence, PhD Vice President, Research and Quality National Hospice and Palliative Care Organization

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2014

Ad.3 **Month and Year of most recent revision:** 10, 2014

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

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

Ad.6 **Copyright statement:** This questionnaire is not copyrighted. However, CAHPS is a copyright of the Agency for Healthcare Research and Quality (AHRQ)

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:** We have no currently scheduled update for this measure.

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 #: 0210
Measure Title: Proportion of patients who died from cancer receiving chemotherapy in the last 14 days of life
Measure Steward: American Society of Clinical Oncology
Brief Description of Measure: Proportion of patients who died from cancer receiving chemotherapy in the last 14 days of life
Developer Rationale: There is evidence that demonstrates that patients receive unnecessary treatments at the end of life, which can negatively impact the patient and caregiver experience. Patients continue to receive chemotherapy treatments at the end of life even when it is recognized that it is unnecessary. For example, more than 15% of patients with metastatic lung and colorectal cancer received chemotherapy in the last month of life (Mack, 2015). In addition, receipt of chemotherapy at the end of life can increase the potential for hospitalizations and intensive care admissions (El-Jawahri, 2015), which can negatively impact the patient's and caregiver's experience.

El-Jawahri, A. R., G. A. Abel, et al. (2015). "Health care utilization and end-of-life care for older patients with acute myeloid leukemia." Cancer 121(16): 2840-2848.

Mack, J. W., A. Walling, et al. (2015). "Patient beliefs that chemotherapy may be curative and care received at the end of life among patients with metastatic lung and colorectal cancer." Cancer 121(11): 1891-1897.

Numerator Statement: Patients who died from cancer and received chemotherapy in the last 14 days of life
Denominator Statement: Patients who died from cancer.
Denominator Exclusions: None

Measure Type: Process
Data Source: Administrative claims, Electronic Clinical Data : Registry
Level of Analysis: Clinician : Group/Practice

IF Endorsement Maintenance – Original Endorsement Date: Aug 10, 2009 **Most Recent Endorsement Date:** Aug 09, 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**

Summary of prior review in 2012

- The developer cited individual studies indicating that continuing chemotherapy near death does not prolong survival but instead often results in toxicity, inconvenience, and increased costs and that ratings of quality of end-of-life care are higher when there is more than a 14-day interval between chemotherapy and death. They also referenced an underlying consensus statement from 2003 that identified potential indicators of quality of end-of-life cancer care using administrative data.
- In the previous evaluation of the measure, the Committee noted that sometimes administration of chemotherapy in the last 14 days of life is appropriate, but agreed that this measure is useful for detecting patterns in practice and variation in performance, identifying outliers when comparing similar practices with similar patient populations, addressing patient preference and overtreatment at the end of life, and reflecting disparities in access to care and the capacity of the local healthcare system to treat patients appropriately at the end of life.

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 provides a diagram of the relationship of this process of care (discontinuing chemotherapy in last 2 weeks of life) to better patient outcomes (fewer ED and ICU visits, in-hospital deaths, reduced costs, improved death experience and improved satisfaction with care).
- The developers reference four sources for evidence. However, none of these supports the relationship of discontinuing chemotherapy in the last 2 weeks of life to desired patient outcomes.
    - El-Jawahri, 2015, found that receipt of intense chemotherapy at the end of life (compared to less intense chemotherapy) increases the likelihood of hospitalization and ICU admission, which can negatively impatient patient and caregiver experience.
    - Mack, 2015, found that more than 15% of patients with metastatic lung and colorectal cancer received chemotherapy in the last month of life. Developers note that this article indicates that "*Patients continue to receive chemotherapy treatments at the end of life even when it is recognized that it is unnecessary*."
    - A 2013 Cochrane Collaborative systematic review evaluated the impact of home-based palliative care services on several patient and caregiver outcomes. This review found that for patients with cancer, home-based palliative care services increases the chance of dying at home and reduces symptom burden, without impacting on caregiver grief. However, findings regarding home-based palliative care are not synonymous with discontinuation of chemotherapy and so this systematic review is not directly applicable to this measure.
    - A 2012 provisional clinical opinion from the American Society of Clinical Oncology addressed integration of palliative care services into standard oncology care. This panel concluded that while there is need for additional research, several of the available studies show evidence of benefit, and recommended that palliative care should be considered early in the course of illness for any patient with metastatic cancer and/or with high symptom burden.

Exception to evidence – N/A

Guidance from the Evidence Algorithm

Process measure, but no systematic review of evidence relevant to the process being measured (Box 3) → Empirical evidence without systematic review (Greer, Landrum) (Box 7) → Unknown whether articles cited includes all studies in

the body of evidence (Box 8) → Moderate or Low, depending on whether the body of evidence is represented

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?*
- *Does the Committee know of additional empirical evidence supporting this process to desired patient outcomes?*

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

Rationale: If the articles cited include all studies in the body of evidence, the measure is eligible for a MODERATE rating if the Committee agrees there is a high certainty that benefits clearly outweigh undesirable effects.

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.

- Data from the ASCO Quality Oncology Practice Initiative registry (QOPI) for 2013-2015 were provided, as follows, although it is unclear whether the above data represent all the patients in the registry or a sample.

	2013	2014	2015
Number of practices	180	172	222
Number of charts	4,951	5,021	7,239
Total patient population	10.16%	11.43%	11.80%
Mean	11.47	12.92	13.16
SD	11.87	12.58	11.5
25th percentile	3.35	9.88	15.81
Median	9.88	11.45	11.95
75th percentile	15.81	17.07	16.66
90th percentile	24.26	21.88	23.08

Disparities: Although patient-level disparities data from the QOPI registry were provided, practice-level data are needed.

Questions for the Committee:

- *Are there enough practices and patients represented to make a determination about opportunity for improvement?*
- *Does it appear that performance has gotten worse over time?*
- *Is there a gap in care that warrants a national performance measure?*
- *Does the Committee have evidence that there are practice-level racial or other disparities in chemotherapy administration in the last 2 weeks of life for cancer patients?*

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.

* The evidence that has been presented is mostly recent literature. The literature is tangential in that none of the articles shows the relationship between discontinuation of chemotherapy and desired patient outcomes, that of negatively impacting the patients' and caregivers' experience. Although the recommendation of palliative care early in

the process of treating a metastatic patient makes sense they would still be able to receive all treatments, including chemotherapy, radiation and/or biologics. The transition to hospice before the last 2 weeks of life would be critical and this is not addressed. I am not aware of any other evidence.
* The "chemo within 14 days of death" metric is a proximal process measure which seeks to touch upon the ultimate outcomes of "positive death experience/improved patient satisfaction/reduced utilization." The relationship between the process measure and final outcome has face validity, but the evidence cited is of low-modest quality. Curiously, two important studies were not included in the submission:
1. Prigerson HG, Bao Y, Shah MA, et al. Chemotherapy Use, Performance Status, and Quality of Life at the End of Life. JAMA Oncol. 2015;1(6):778-784. doi:10.1001/jamaoncol.2015.2378.
and 2. Patients' Expectations about Effects of Chemotherapy for Advanced Cancer. Jane C. Weeks, M.D., Paul J. Catalano, Sc.D., Angel Cronin, M.S., Matthew D. Finkelman, Ph.D., Jennifer W. Mack, M.D., M.P.H., Nancy L. Keating, M.D., M.P.H., and Deborah Schrag, M.D., M.P.H. N Engl J Med 2012; 367:1616-1625October 25, 2012DOI: 10.1056/NEJMoa1204410.

The evidence of chemotherapy harm at the end of life is most robust for advanced solid tumor colon and lung...how relevant is this quality measure for less toxic "targeted" agents and biologics?

Fundamentally, this metric is a surrogate for informed decision making and goals of care clarity in advanced cancer care...balancing "aggressive" therapy near the end of life with the costs involved (potential for harm, decreased quality of life, financial toxicity, etc.)

1b.
* Performance data was provided but the disparities were not clear. Although the number of patients was presented the racial breakdown was not clear. It is not clear if performance has gotten better because the numbers indicate that more patients are being reported on but not what the results are.
* Performance data was provided and there is a "gap" in performance. Nevertheless, it is unclear what the ideal state should be.

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 sources specified in the submission include administrative claims and registry.

Specifications:

- This measure is specified for the clinician group/practice level of analysis.
- A lower score indicates better quality.
- Care settings for this measure include clinician offices or clinics, hospices, or hospitals/acute care facilities.
- Registry data are collected through the ASCO Quality Oncology Practice Initiative (QOPI®) registry. No data collection instrument has been provided.
- For the measure numerator, RxNORM, CPT, and SNOMEDCT codes were provided to identify chemotherapy administration in administrative claims data. It is not clear how these codes are being used to compute the numerator, but presumably the date of last chemotherapy claim is compared to the date of death.
- For registry data, the numerator is calculated by comparing the date of death to the date of last chemotherapy administration.
- The denominator (patients who died from cancer) is easily identified in the registry data, but it is unclear how it would be identified through claims data. The developers briefly mention use of the “death registry” but do not explain how this is used in conjunction with claims data.
- There are no exclusions to the measure.
- A calculation algorithm is provided.
- This measure is not risk-adjusted.

Questions for the Committee :

- *Is use of death registry data required if the measure is calculated from claims data? If so, what is the timeliness of this dataset? If the death registry is not used, can the measure denominator be identified reliably through administrative data?*
- *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:

- For registry data, data element reliability testing was conducted by comparing QOPI data to data re-abstracted from medical records.
- For claims data, data element validity testing was conducted by comparing claims data to the full medical record. NQF guidance indicates that if data element validity is demonstrated, additional data element reliability testing is not required.

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

- For registry data:
  - In 2008, data from the QOPI registry was compared to data that were re-abstracted from medical records by QOPI nurse abstractors. The testing sample included 264 records from 44 sites.
  - **If** the QOPI nurse abstractors who did the re-abstraction are considered expert abstractors—and their re-abstracted data to be the gold standard—then this testing would also meet the requirements for data element validity testing.
- For claims data, see method(s) of validity testing, below.

Results of reliability testing

- For registry data:
  - Developers report a single kappa value of 0.818.
  - 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 0.818 means that the raters agreed 81.8% 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.
  - NQF guidance indicates that data element testing should be conducted for all critical data elements, although at minimum, results about the numerator, denominator, and exclusions should be provided. For this measure, only a single kappa value is reported, and thus it is unclear whether applies to the numerator or denominator data element. It likely applies to the numerator, as the developer presumably sampled only cancer decedents from the registry.
- For claims data, see results of validity testing, below.

Guidance from the Reliability Algorithm:
For registry data: Precise specifications (Box 1) → empirical reliability testing conducted for registry data (Box 2) → testing at score level not conducted (Box 4) → data element testing conducted with appropriate method, but need to verify testing was for the numerator (Box 9) → Moderate

For claims data: Fairly precise specifications, except for denominator (Box 1) → no empirical reliability testing conducted for claims data (Box 2) →empirical validity testing conducted (Box 3) → FROM VALDITY ALGORITHM: data element testing for numerator (Box 10) → Accuracy of the denominator data element not demonstrated for claims data, but assume okay if data derived from national death registry (Box 11) → Moderate

Questions for the Committee:

- *Is death from cancer be accurately recorded in the QOPI registry?*
- *Can claims adequately identify cancer deaths? If not, is use of death registry data required?*
- *No updated testing information is presented. The prior testing demonstrated moderate reliability. Does the Committee think there is a need to 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**

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:

- For claims, data element validity testing was done by comparing claims data to the full medical record. This is an appropriate method of data element validation.
- For registry data, data element reliability testing was conducted by comparing QOPI data to data re-abstracted from medical records. This would be considered an appropriate method of data element validation testing **only if** the QOPI nurse abstractors who did the re-abstraction are considered expert abstractors and thus their re-abstracted data considered to be the gold standard.

Describe any updates to validity testing:

- No updated testing was provided.

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 methods (from previous evaluation):

- For administrative claims data:
  - Claims data for 150 consecutive patients treated for advanced cancer at Boston's Dana-Farber Cancer Institute and Brigham and Women´s Hospital were compared to data from the full medical record. Numerator data elements were compared ***but not the denominator data element***. Dates for the data examined in testing were not provided.
- Face validity assessment:
  - The developer states that face validity was assessed by conducting focus group and structured interviews with end-of-life cancer patients and bereaved caregivers, and then vetted by an expert panel of cancer providers. According to NQF guidance, the face validity of the measure score as a quality indicator may be adequate if it explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. There is not enough information provided to know if the face validity assessment conforms to NQF's requirements.

Validity testing results (from previous evaluation):

- For administrative claims data:
  - Sensitivity=0.92; Specificity=0.94
  - *Sensitivity* measures the proportion of actual positives that are correctly identified as such. A sensitivity value of 0.92 reflects the accuracy of identifying a chemotherapy administration within 14 days of death in the claims data when it present in the medical record data (the authoritative source).
  - *Specificity* measures the proportion of actual negatives that are correctly identified as such. A specificity value of 0.94 reflects the accuracy of the absence chemotherapy administration within 14 days of death in the claims data when this is not recorded in the medical record data (the authoritative source).
- Face validity assessment:
  - The developer does not provide results from the face validity assessment.

Question for the Committee:

- *The developer has not provided any new validity testing, even though there is no testing for the denominator element (cancer deaths) for claims data. Does the developer expect the denominator to be identified via a death registry? If so, should these data be assumed valid? If not, is there evidence that claims adequately identify cancer deaths?*
- *Is the test sample adequate to generalize for widespread implementation?*
- *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:

- There are no exclusions for this measure.

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

- Typically, process measures are not risk-adjusted, although theoretically they can be.
- In the ***previous evaluation*** of the measure, the Committee raised concerns about how case mix would be accounted for in the measure and questioned whether facilities with a high number of patients enrolled in clinical trials would skew the measure results so that those facilities would appear not to do as well on the measure.
- The developer states that "*No risk adjustment or risk stratification is necessary because a) the measure is intended to be used for comparison among similar providers; unless there is a reason to believe that one providers' patients have significantly different risks than others, it will not affect relative comparisons, and b) comorbidity risks will if anything decrease the likelihood of experiencing this process of care*."
- No empirical analysis is provided to support lack of adjustment and nothing in the specifications requires

comparison of results across similar providers.

Question for the Committee:

- *Do you agree with the developer's rationale that risk-adjustment or stratification is not needed for this measure?*

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

- When initially developed (using 1991-1996 SEER-Medicare linked data), a benchmark of <10% was achieved by the highest performing SEER regions.
- Variation across providers is shown via descriptive statistics about performance results (see section 1b, above), although the high standard errors indicate that differences between providers may be due to chance.

Question for the Committee:

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

2b6. Comparability of data sources/methods:

- The developer has not provided analysis to assure comparability of results between claims versus registry (the analysis of claims data across time presented in the earlier evaluation does not speak to this topic).

2b7. Missing Data

- The developer describes how missing data are handled in section S.22. No information regarding extent of missing data in claims or registry data is provided.

Guidance from the Validity Algorithm:
For claims data: Specifications somewhat consistent with evidence (Box 1) → potential threats to not completely assessed (Box2) →Insufficient

Preliminary rating for validity: ☐ **High** ☐ **Moderate** ☐ **Low** ☒ **Insufficient**
Rationale: Additional analysis to assess threats to validity needed (risk-adjustment/comparability). For claims data, testing for the denominator data element for claims was not conducted, but may not be needed if data required to be obtained from a death registry. For registry data, results cannot be considered demonstration of validity unless the QOPI nurse re-abstraction is considered the gold standard and results are presented for numerator and denominator separately.

Committee pre-evaluation comments

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

2a, 2b.
Reliability – Specifications
* The main concern about the reliability is how the cause of death is reported, is it a result of cancer or treatment or other factors. Also should the parameters include all treatment including biologics and radiation along with chemotherapy.

Validity – Specifications
* The specifications are somewhat consistent with the evidence but the evidence is not as clear as it should be.

Reliability – Testing
* The expertise of the data abstractors is not clear and it is not clear if each one uses the same parameters.
* Fairly reliable, however, the issue around the definition of "chemotherapy" in the evolving era of targeted therapy and biologics needs clarification.

Validity Testing

* It is not clear if the data from this measure is an indicator of quality of care. It is not clear that the patients who died died from the cancer or other complications. A conclusion of better end of life care cannot be made from the data presented. PRO in this case family members may be more accurate.

Threats to Validity
* Not sure if the problem is missing data or insufficient data.

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.

- Most data elements are in defined fields in electronic clinical data, either in the ASCO Quality Oncology Practice Initiative registry, or in administrative claims. Most data elements are routinely generated during care delivery. The Chemotherapy Administration Value Set is recommended by the developer to identify chemotherapy treatment.
- Cancer decedents are easily identifiable in the registry, but may not be easily or accurately identifiable through claims.
- No feasibility concerns were raised by the Cancer Steering Committee during the NQF Maintenance of Endorsement review in 2012.

Questions for the Committee:
- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*
- *Is use of a death registry required? If not, how will the measure denominator be identified using claims data?*

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

Committee pre-evaluation comments
Criteria 3: Feasibility

* Use of a death registry would be helpful if the specific cause of death is indicated in the registry. The use of the ASCO QOPI represent consistency that may not be present in the administrative claims.

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:
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 measure is used the Quality Oncology Practice Initiative (QOPI), a practice-based quality improvement and benchmarking program. Data for this measure can be submitted to CMS to meet PQRS requirements. The developer notes that "*In Fall 2015, 36 practices and 3,124 patient charts were submitted to PQRS through QOPI*"

although it is unclear how many of these practices submitted data for this measure.

- While individual-level data will be publicly reported on Physician Compare in 2016, practice-level data (for which this measure is specified) will not be available until 2017.
- The measure is part of America's Health Insurance Plans (AHIP) Medical Oncology Core Measure Set. The AHIP effort is a collaboration of both public and private stakeholders to identify measures that are meaningful to patients, consumers, and physicians and to reduce variability in measure selection, collection burden, and cost. Payers involved in the collaboration have committed to using for reporting as soon as feasible. By virtue of being included in the AHIP measure set, CMS will consider this measure for inclusion in Medicare quality programs.

Improvement results:

- In 2013-2015, mean practice performance slightly worsened from 11.47% of patients receiving chemotherapy in last 14 days of life to 13.16%. These results are based on data from the QOPI registry and reflect slightly greater use of the registry during the timeframe, from 180 practices in 2013 to 222 in 2015.

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

Potential harms: No potential harms have been reported by the developer.

Feedback: Public comments received during the 2012 maintenance review of the measure expressed concern about the difficulties of prognosticating death and the appropriateness of chemotherapy in the last two weeks of life for some patients.

***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**

Committee pre-evaluation comments

Criteria 4: Usability and Use

* It is not apparent what if any unintended consequences would occur with the measure. The issue of quality of care at the end of life is not a measurable metric.

Criterion 5: Related and Competing Measures

Related measures:

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

0213: Proportion of patients who died from cancer admitted to the ICU in the last 30 days of life

0215: Proportion of patients who died from cancer not admitted to hospice

0216: Proportion of patients who died from cancer admitted to hospice for less than 3 days

Harmonization: N/A

Pre-meeting public and member comments

Name: Michele Galioto, RN, MSN

Organization: Oncology Nursing Society

Comment: Should the measure include immunotherapy? Does “chemotherapy” include hormonal and biotherapy? It may be more inclusive to refer to all as "antineoplastic therapy."

Should the measure include radiation therapy as well?

Does “death from cancer” include all death within 14 days? Death may be the result of infection, accident (e.g., fall), bleeding, etc. which could be tied to cancer or cancer treatment. Death attributed to side effects of therapy may be indistinguishable from cancer deaths. Is the intent that death occurs within a timeframe of receiving chemotherapy?

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*)**:** 0210

Measure Title: Proportion of patients who died from cancer receiving chemotherapy in last 14 days of life

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): Click here to name the PRO

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: Patients receiving chemotherapy in the last 14 days of life

☐ 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.

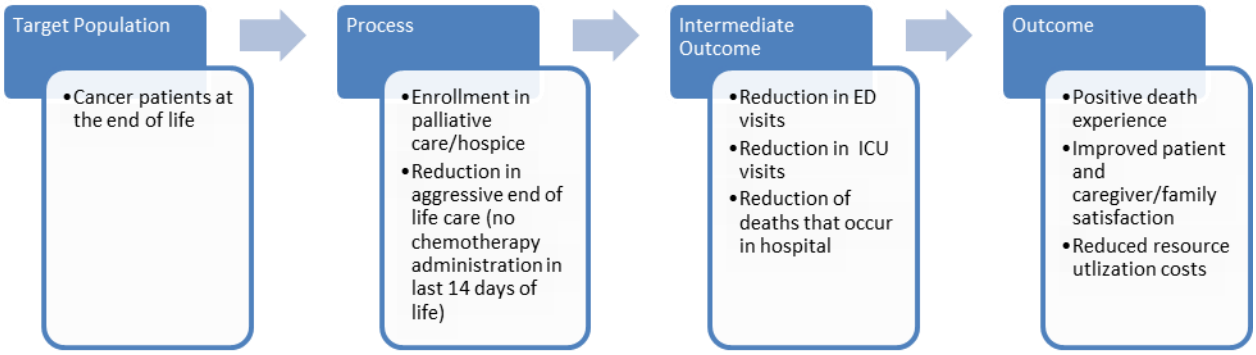
1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

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.

Use from



There is evidence that demonstrates patients receive unnecessary treatments at the end of life, which can negatively impact the patient and caregiver experience. Patients continue to receive chemotherapy treatments at the end of life even when it is recognized that it is unnecessary. For example, more than 15% of patients with metastatic lung and colorectal cancer received chemotherapy in the last month of life (Mack, 2015). In addition, receipt of chemotherapy at the end of life can increase the potential for hospitalizations and intensive care admissions (El-Jawahri, 2015), which can negatively impact the patient's and caregiver's experience.

El-Jawahri, A. R., G. A. Abel, et al. (2015). "Health care utilization and end-of-life care for older patients with acute myeloid leukemia." Cancer 121(16): 2840-2848.

Mack, J. W., A. Walling, et al. (2015). "Patient beliefs that chemotherapy may be curative and care received at the end of life among patients with metastatic lung and colorectal cancer." Cancer 121(11): 1891-1897.

A structural feature: regional availability of hospice, has been shown to correlate with a composite measure of the aggressiveness of cancer care near the end of life that contains this measure. Mostly it is a process measure indicating a possible inadequate focus on palliation and supportive care, that can affect quality of life.

The Process-Outcome link is that continuing chemotherapy near death does not prolong survival (Saito AM, Landrum MB, Neville BA, et al. The effect on survival of continuing chemotherapy to near death. BMC Palliat.Care. 10:14.:14, 2011) or quality of life (Temel JS, Greer JA, Muzikansky A, et al. Early palliative care for patients with metastatic non-small-cell lung cancer. N Engl J Med. 363:733-42, 2010; Greer et al in press J Clin Oncol 2011) and consequently provides only toxicity, inconvenience, and cost in most cases.

In the NIH-funded Cancer Care Outcomes Research and Surveillance Consortium, bereaved family members of 706 lung or colorectal cancer patients rated the quality of end-of-life care their loved one had received. Adjusted for age, sex, marital status, income, education, stage, comorbidity, health system type, census region, and the respondent´s relationship to the patient, respondents were significantly more likely to rate the end-of-life care to have been very good or excellent (86.0 vs 75.7%), with no unmet need for help with anxiety or depression (86.6 vs 78.0%) or breathing (86.3 vs 80.3%), and that they

died in their preferred location (77.6 vs 56.3%) if the patient had a composite measure of having spent at least 3 days in hospice, had 1 or fewer hospital admissions in the last month of life, or had an interval of more that 14 days between the last dose of chemotherapy and death(Landrum MB et al, under review).

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***

In 2012: Type of Evidence was selected individual studies (rather than entire body of evidence)

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*):

Smith TJ, Temin S, Alesi ER, et al. American Society of Clinical Oncology Provisional Clinical Opinion: The Integration of Palliative Care into Standard Oncology Care. J Clin Oncol 2012;30:880-887. Available at: http://www.instituteforquality.org/asco-provisional-clinical-opinion-integration-palliative-care-standard-oncology-care.

Gomes, B., N. Calanzani, et al. (2013). "Effectiveness and cost-effectiveness of home palliative care services for adults with advanced illness and their caregivers." Cochrane Database Syst Rev 6: CD007760. Available at: http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD007760.pub2/pdf.

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?

A 2012 American Society of Clinical Oncology (ASCO) Provisional Clinical Opinion (PCO) addresses the integration of palliative care (PC) services into standard oncology care at the time a person is diagnosed with metastatic cancer and/or high symptom burden.

A 2013 Cochrane Review, 'Effectiveness and cost-effectiveness of home palliative care services for adults with advanced illness and their caregivers', evaluated the impact of home palliative care services on outcomes for adults with advanced illness or their family caregivers, or both. The aim of the review was to quantify the effect of home palliative care services on a patients' odds of dying at home, examine the clinical effectiveness of home palliative care services on other outcomes such as symptom control, quality of life, caregiver distress and satisfaction with care, and comparing resource use and costs associated with these services.

2012 Submission: The argument is made that because providers cannot predict the future, measures based on decedent cohorts are unfair. However, as described above in 1a.a, the idea is for the measure to be seen as an overall indication of practice style and/or available palliative resources. An individual patient experiencing this process of care has not necessarily received poor quality care. If explanations other than practice style and resource availability, such as unusually poor prognostic ability on the part of the provider or unexpected toxic deaths (whether unavoidable, from overly aggressive treatment, or poor patient selection) are enough to influence the overall aggregate rates, it is still justifiable to consider it a 'red flag' that should prompt examination of the care provided.

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

2012 ASCO PCO (p. 881):

The American Society of Clinical Oncology (ASCO) has established a rigorous, evidence-based approach—the provisional clinical opinion (PCO)—to offer a rapid response to emerging data in clinical oncology. The PCO is intended to offer timely clinical direction to ASCO's oncologists after publication or presentation of potentially practice- changing data from major studies.

The PCO may serve in some cases as interim direction to the membership pending the development or updating of an ASCO clinical practice guideline. As such, the evidence is not graded in a PCO and is a result of expert consensus. A clinical guideline on palliative care integration with recommendations and the associated grading is under development.

2013 Cochrane Review (p. 12):

Two independent reviewers assessed all included studies for methodological quality using the standard criteria developed by the Cochrane EPOC Review Group for RCTs/CCTs, CBAs and ITSs. The checklist for RCTs/CCTs contains seven items qualified as done, unclear and not done for concealment of allocation, follow-up of professionals, follow up of patients or episodes of care, blinded assessment of primary outcome(s), baseline assessment, reliable primary outcome measure(s) and protection against contamination. Blinding and reliability of all outcomes were also assessed.

Each criterion was scored zero (not done), 0.5 (not clear or when scores varied across outcomes) and one (done). Total scores for RCTs/ CCTs ranged from zero to six; studies with a score of 3.5 or above were considered of high quality. Integration of the results of the quality assessment in data analysis was done in addition to meta-analyses with sensitivity analyses including only high quality studies.

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

See 1a.7.2 for this information

2012 Submission: Other, does not apply

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

2012 ASCO PCO: 2004-2012

2013 Cochrane Review: 1950 – November 2012

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*)

2012 ASCO PCO: 7 randomized controlled trials

2013 Cochrane Review: 5 randomized controlled trials and 2 controlled clinical trials

2012 Submisson: 4 studies were used in the body of evidence

1a.7.6. What is the overall quality of evidence <u>across studies</u> 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*)

2012 ASCO PCO:

This PCO did not provide an assessment of the overall quality of evidence across the studies. This analysis will be completed during the development of the upcoming clinical guideline.

2013 Cochrane Review :

p. 3: The direction of the effect was consistent across all studies but did not reach statistical significance in 3; ORs ranged from 1. 36 (95% CI 0.80 to 2.31) to 2. 86 (95% CI 0.78 to 10.53) Sensitivity analyses showed that exclusion of the 2 CCTs (both of Swedish hospital-based services with a pooled OR 3.44, 95% CI 0.60 to 19.57) and inclusion of only high quality RCTs resulted in a reduction of the OR to 1.28 (95% CI 1.28 to 2.33) and 1.75 (95% CI 1.24 to 2. 47) respectively, with more precision and less heterogeneity.

p. 22:

Pooled data from seven studies (five RCTs, three of high quality, and two CCTs with 1222 participants) showed that those receiving home palliative care had statistically significantly higher odds of dying at home than those receiving usual care (OR 2.21, 95% CI 1.31 to 3.71; Z = 2.98, P value = 0.003; Chi^2 = 20.57, degrees of freedom (df) = 6, P value = 0.002; I^2 = 71%). The study population control risk was of 307 home deaths per 1000 deaths; based on this ACR of 0.307, the NNTB was 5 (95% CI 3 to 14), meaning that for one additional patient to die at home five more would need to receive home palliative care as opposed to usual care. Assuming a medium cancer home death rate population ACR of 0.278 (i.e. 278 home deaths per 1000 cancer deaths), the NNTB was 6 (95% CI 3 to 15). This means that for one additional cancer patient to die at home in a population where there are 278 home deaths per 1000 cancer deaths, six more would need to receive home palliative care. NNTB estimates ranged from 9 patients (95% CI 5 to 16) when applied to a low home death rate population such as the one observed in Norway (128 home deaths per 1000 cancer deaths) to 5 patients (95% CI 3 to 13) when applied to a high home death rate population such as the one observed in the Netherlands (454 home deaths per 1000 cancer deaths).

2012 submission: The studies are qualitative and observational using administrative data, consequently there are limitations to the quality of the data.

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) <u>across studies</u> 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*)

2012 ASCO PCO (p. 884):

Seven published randomized trials demonstrate the feasibility of providing various components of PC alongside usual oncology care. There is, however, a dearth of data evaluating the integration of modern PC practices into standard oncology care, especially in concert with ongoing antitumor therapy. Overall, the addition of PC interventions to standard oncology care delivered via different models to patients with cancer provided evidence

of benefit.

2013 Cochran Review (p. 22):

The study population control risk was of 307 home deaths per 1000 deaths; based on this ACR of 0.307, the NNTB was 5 (95% CI 3 to 14), meaning that for one additional patient to die at home five more would need to receive home palliative care as opposed to usual care. Assuming a medium cancer home death rate population ACR of 0.278 (i.e. 278 home deaths per 1000 cancer deaths), the NNTB was 6 (95% CI 3 to 15). This means that for one additional cancer patient to die at home in a population where there are 278 home deaths per 1000 cancer deaths, six more would need to receive home palliative care. NNTB estimates ranged from 9 patients (95% CI 5 to 16) when applied to a low home death rate population such as the one observed in Norway (128 home deaths per 1000 cancer deaths) to 5 patients (95% CI 3 to 13) when applied to a high home death rate population such as the one observed in the Netherlands (454 home deaths per 1000 cancer deaths).

2012 Submission: All studies have shown similar results.

As per Ho TH, Barbera L, Saskin R, Lu H, Neville BA, Earle CC.Trends in the Aggressiveness of End-of-Life Cancer Care in the Universal Health Care System of Ontario, Canada. J Clin Oncol April 20, 2011 vol. 29 no. 12 1587-1591, although rates in Canada were lower, trends were similar over time in a comparison with U.S. Medicare patients.

A study examined the frequency and duration of chemotherapy use among Medicare cancer decedents in Massachusetts and California. Researchers found that 9% of Medicare cancer decedents in Massachusetts and California received chemotherapy in the last month of life. No difference was found by geographic region (Emanuel E, Young-Xu Y, Levinsky, N, et al. Chemotherapy use among medicare beneficiaries at the end-of-life. Annals of Internal Medicine. 2003; 138(8): 639-643.).

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

2012 ASCO PCO (p. 884-885):

No harm to any patient was observed in any trial, even with discussions of EOL planning, such as hospice and ADs. Two of five trials measuring change in symptoms, two of five studies measuring QOL, two of three studies measuring patient/caregiver satisfaction, and one of three studies measuring survival found statistically significant improvements with PC. Three of six studies measuring mood, two of five studies measuring resource use, and one of four studies measuring outcomes of advance care planning found statistically significant differences, and one outcome of borderline significance was also found in each of these three areas, Therefore, most trials showed benefits ranging from equal to improved overall survival, reduced depression, improved caregiver and/or patient QOL, and overall lower resource use and cost because EOL hospitalizations were avoided.

2013 Cochrane Review: Discussion of harms was not addressed.

2012 Submission: A shift from aggressive anti-cancer treatment to supportive care can improve the quality of death. There are no known harms to not giving chemotherapy near death.

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.

No relevant studies have been conducted and published since the systematic reviews.

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.

2012 Submission: The underlying evidence was obtained by expert consensus, as described in Earle CC, Park ER, Lai B, Weeks JC, Ayanian JZ, Block S. Identifying potential indicators of the quality of end of life cancer care from administrative data. J Clin Oncol. 2003;21(6):1133-8. The panel consisted of oncologists, nurses, palliative care specialists, etc, and used a modified Delphi process to evaluate measures.

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 #: 0210
De.2. **Measure Title:** Proportion of patients who died from cancer receiving chemotherapy in the last 14 days of life
Co.1.1. **Measure Steward:** American Society of Clinical Oncology
De.3. **Brief Description of Measure:** Proportion of patients who died from cancer receiving chemotherapy in the last 14 days of life
1b.1. **Developer Rationale:** There is evidence that demonstrates that patients receive unnecessary treatments at the end of life, which can negatively impact the patient and caregiver experience. Patients continue to receive chemotherapy treatments at the end of life even when it is recognized that it is unnecessary. For example, more than 15% of patients with metastatic lung and colorectal cancer received chemotherapy in the last month of life (Mack, 2015). In addition, receipt of chemotherapy at the end of life can increase the potential for hospitalizations and intensive care admissions (El-Jawahri, 2015), which can negatively impact the patient's and caregiver's experience.

El-Jawahri, A. R., G. A. Abel, et al. (2015). "Health care utilization and end-of-life care for older patients with acute myeloid leukemia." Cancer 121(16): 2840-2848.

Mack, J. W., A. Walling, et al. (2015). "Patient beliefs that chemotherapy may be curative and care received at the end of life among patients with metastatic lung and colorectal cancer." Cancer 121(11): 1891-1897.

S.4. **Numerator Statement:** Patients who died from cancer and received chemotherapy in the last 14 days of life
S.7. **Denominator Statement:** Patients who died from cancer.
S.10. **Denominator Exclusions:** None

De.1. **Measure Type:** Process
S.23. **Data Source:** Administrative claims, Electronic Clinical Data : Registry
S.26. **Level of Analysis:** Clinician : Group/Practice

IF Endorsement Maintenance – Original Endorsement Date: Aug 10, 2009 **Most Recent Endorsement Date:** Aug 09, 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
0210_Evidence_Form_3.15.16.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)*
There is evidence that demonstrates that patients receive unnecessary treatments at the end of life, which can negatively impact the patient and caregiver experience. Patients continue to receive chemotherapy treatments at the end of life even when it is recognized that it is unnecessary. For example, more than 15% of patients with metastatic lung and colorectal cancer received chemotherapy in the last month of life (Mack, 2015). In addition, receipt of chemotherapy at the end of life can increase the potential for hospitalizations and intensive care admissions (El-Jawahri, 2015), which can negatively impact the patient's and caregiver's experience.

El-Jawahri, A. R., G. A. Abel, et al. (2015). "Health care utilization and end-of-life care for older patients with acute myeloid leukemia." Cancer 121(16): 2840-2848.

Mack, J. W., A. Walling, et al. (2015). "Patient beliefs that chemotherapy may be curative and care received at the end of life among patients with metastatic lung and colorectal cancer." Cancer 121(11): 1891-1897.

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.*
This data was produced from the QOPI® registry and data was abstracted for a sample of patients seen with the data collection period. Data is reported at the practice level.

In 2013, 180 practices reported on 4951 charts.

In 2014, 172 practices reported on 5021 charts.

In 2015, 222 practices reported on 7239 charts.

	2013	2014	2015
Total Patient Population (%)	10.16	11.43	11.80
Mean	11.47	12.92	13.16
Minimum	0	0	0
Maximum	100	100	100
Standard Deviation	11.87	12.58	11.5
Percentiles			
10	0	0	0
25	3.35	9.88	15.81
50	9.88	11.45	11.95
75	15.81	17.07	16.66
90	24.26	21.88	23.08
95	28.57	28.13	32.14

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.*
This data was produced from the QOPI® registry and data was abstracted for a sample of patients seen with the data collection

period. Data is reported at the chart level as practice level disparity data is currently not calculated.

In 2013, 180 practices reported on 4951 charts.

In 2014, 172 practices reported on 5021 charts.

In 2015, 222 practices reported on 7239 charts.

	2013 (n=4951)	2014 (n=5021)	2015 (n=7239)
Total Patient Population	10.16	11.43	11.80
Female	10.31	11.38	10.84
Male	10.03	11.48	12.64
Hispanic	8.67	12.50	11.36
White	10.78	11.40	12.14
Black	9.57	11.81	10.76
Other	6.19	13.49	10.04

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**

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.**

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

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

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.)*
No webpage available

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)*
Attachment **Attachment:** Chemotherapy.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
No changes have been made since last endorsement

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 who died from cancer and received chemotherapy in the last 14 days of life

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.)*
14 days prior to death

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.
Claims: see attached chemotherapy code set.
Registry: Date of death – date of last chemotherapy administration </= 14 days

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Patients who died from cancer.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
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)*
Claims: Patients in the death registry with cancer as their cause of death. In the cited analyses by the measure submitter, this is a field in the cancer registry or denominator file not requiring specific codes. This may be different in other administrative data sets.

Registry: Deceased = Yes, patient is deceased as a consequence of his/her cancer or cancer treatment.

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)*
None

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)*
Not applicable

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)*
Not applicable

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 = Lower 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.)*
Performance is calculated as:
1. Identify those patients that meet the denominator criteria defined in the measure.
2. Subtract those patients with a denominator exclusion from the denominator. Note: this measure does not have exclusions.
3. From the patients who qualify for the denominator (after any exclusions are removed), identify those who meet the numerator criteria.
4. Calculation: Numerator/Denominator-Denominator Exclusions

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.

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.
Not applicable

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
This measure is specified with defined criteria and data elements. If a patient record does not include one or more of these components for the initial patient population or denominator, then patients are not considered eligible for the measure and not included.
If data to determine whether a patient should be considered for the numerator or exclusions is missing, then the numerator or exclusions not considered to be met and the practice will not get credit for meeting performance for that patient.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Electronic Clinical Data : Registry

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.
ASCO Quality Oncology Practice Initiative (QOPI®)

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)*
Clinician : Group/Practice

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic, Hospice, 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.)*
Not applicable

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
0210_MeasureTesting_MSF5.0_Data_Update.doc

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 0210	**NQF Project:** Cancer Project

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)***:**

The measure was developed using the Medicare claims of all continuously-enrolled patients who died of cancer after having been diagnosed in a SEER region between 1991 and 1996.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)***:**

Evaluation was carried out on 150 consecutive patients treated for advanced cancer at Dana-Farber Cancer Institute and Brigham and Women´s Hospital in Boston. The percent accuracy of death ascertainment for includion into this cohort is unknown but is likely high as the cancer registry regularly uses the death index for ascertainment. Ascertainment would be expected to be highly specific. Hospital billing cClaims were obtained and analyzed and the accuracy was compared to detailed medical record review.

2a2.3 Testing Results *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**

Sensitivity 0.92, Specificity 0.94 where sensitivity = # true positives (both claims and charts)/(# true positives + # false negatives, i.e., not in claims but present in charts) and specificity = # true negatives/(# true negatives + false positives, i.e., present in claims but not in charts).

2b. VALIDITY. Validity, Testing, including all Threats to Validity: 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:**

They are identical

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)***:**

1) Evaluation was carried out on 150 consecutive patients treated for advanced cancer at Dana-Farber Cancer Institute and Brigham and Women´s Hospital in Boston. Claims were obtained and analyzed and the accuracy was compared to detailed medical record review.

2) In QOPI nurse abstractors did a re-abstraction of 264 medical records at 44 sites in 2008.

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

1) Face validity was determined by focus groups and structured interviews with end-of-life cancer patients and bereaved caregivers, and then vetted by an expert panel of cancer providers. The percent agreement between claims and medical record review was calculated.

2) Inter-rater reliability was calculated using Kappa statistics

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)***:**

1) The measure was 92% accurate (percent true positives + true negatives).

2) The Kappa in the QOPI validation study was 0.818

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)***:**

N/A

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

N/A

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)***:**

N/A

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

N/A

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)***:**

N/A

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment: No risk adjustment or risk stratification is necessary because a) the measure is intended to be used for comparison among similar providers; unless there is a reason to believe that one providers' patients have significantly different risks than others, it will not affect relative comparisons, and b) comorbidity risks will if anything decrease the likelihood of experiencing this process of care.

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)***:**

We used the Medicare claims of all 28,777 continuously-enrolled patients who died of cancer after having been diagnosed in a SEER region between 1991 and 1996. This was an analysis of SEER-Medicare linked data obtained from NCI (http://healthservices.cancer.gov/seermedicare/).

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

Benchmarks were established to identify the outlying 10th decile of practice: The proportion of patients experiencing each process of care in each Health Care Service Area (HCSA) was computed and ranked from best (least aggressive) to worst. A new cohort was created by sequentially adding HCSAs in order starting with the least aggressive until they contained at least 10% of the original cohort and the proportion experiencing each process of care was then recalculated to arrive at the 'Achievable Benchmark of Care). More detail on this, as well as a reference for the Achievable Benchmark of Care method can be found in our publication: Earle CC, Neville BA, Landrum ME, Souza JE, Weeks JC, Block SD, Grunfeld E, Ayanian JZ. Evaluating claims-based indicators of the intensity of end-of-life cancer care. Int J Qual Health Care. 2005;17(6):505-9.

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)*:

A benchmark target of < 10% of patients receiving chemotherapy in the last 14 days of life corresponds to that achieved by the highest performing regions in the country.

Year	Mean	St. Dev.	Min	10th	25th	50th	75th	90th	Max
2013	11.47	11.87	0	0	3.35	9.88	15.81	24.27	100
2014	12.92	12.58	0	0	5.80	11.45	17.07	21.88	100
2015	13.16	11.5	0	6.45	11.95	11.95	16.67	23.08	100

The range of performance across years and across practices suggests there's clinically meaningful variation across physicians' performance.

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

Comparability of measure results using administrative claims versus registry data had not been analyzed as of March 2016.

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)*:

Administrative claims and chart review, as described above: 77 entities (HCSAs), 215,484 patients, between 1991 and 2000.

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

We have also assessed the stability of these measures over time by examining the stability of relative aggressive care over time. If the relative aggressiveness of a provider or organization´s practice appeared to change from year to year, then these measures might not be assessing a stable property of practice. To investigate this, we used hierarchical regression models to estimate regional variation in both levels and trends of each measure. We used as our geographic unit of analysis the Health Care Service Area (HCSA). HCSAs are groupings of Metropolitan Statistical Areas defined by the Centers for Medicare & Medicaid Services (CMS) based on observed patient flow patterns in Medicare for tertiary care. As such, each HCSA can be considered to be a self-contained regional health system with a related group of providers. We ranked each region according to the model-estimated rate of each indicator and computed the correlation among relative ranks of each region during the 10-year study period.

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)*:

We observed significant variation both in levels of aggressive care and in trends in aggressiveness over time, but generally stability of regional practice patterns: Year to year correlation on this measure was 0.94, and over a 5 year span was 0.66. This provides supportive evidence of the reliability of these measures.

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:

N/A

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.**

Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), 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)*

ALL data elements are in defined fields in electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)

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.**

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.**

No feasibility assessment **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.

The measure and its specifications have been in place for several years and ASCO continues to monitor and ensure that the measure and its specifications are up-to-date for widespread use.

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)*.

Not applicable

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)
Public Reporting	Payment Program http://www.instituteforquality.org/qopi/pqrs-measures-0 CMS Physician Quality Reporting Program Qualified Clinical Data Registry Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Quality Oncology Practice Initiative (QOPI®) http://www.instituteforquality.org/qopi/manual-qopi-measures

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

Quality Oncology Practice Initiative:
In 2002, the American Society of Clinical Oncology established the Quality Oncology Practice Initiative (QOPI®). QOPI® is a practice-based quality assessment and improvement program designed to foster a culture of self-examination and improvement in oncology. Collection rounds are offered twice per year, in spring and fall, for an eight week period. QOPI® continues to be a successful program in the United States and 12 other countries, with 441, 313, 361 and 256 unique practices participating in Fall 2013, Spring 2014, Spring 2015 and Fall 2015 respectively.

PQRS Qualified Clinical Data Registry:
In addition to the current use for quality improvement with benchmarking in the QOPI® registry, this measure has been reported to CMS by the registry as a Qualified Clinical Data Registry. QOPI® is approved to be a Qualified Clinical Data Registry and this measure is included in the list of measures that can be reported by participating practices. QOPI® was deemed as a QCDR to report to PQRS in 2015 and 2016. Eligible professionals will be considered to have satisfactorily participated in PQRS if they submit quality measures data or results to CMS via a qualified clinical data registry. In Fall 2015, 36 practices and 3,124 patient charts were submitted to PQRS through QOPI.

In addition to the current use for quality improvement with benchmarking in the QOPI® registry, this measure has been reported to CMS by the registry as a Qualified Clinical Data Registry. QOPI® is approved to be a Qualified Clinical Data Registry and this measure is included in the list of measures that can be reported by participating practices. QOPI® was deemed as a QCDR to report to PQRS in 2015 and 2016. Eligible professionals will be considered to have satisfactorily participated in PQRS if they submit quality measures data or results to CMS via a qualified clinical data registry. In Fall 2015, 36 practices and 3,124 patient charts were submitted to PQRS through QOPI®. In late 2016, all 2015 individual-level QCDR data will be available for public reporting on Physician Compare. In 2017, both individual and group-level QCDR measures will be available for public reporting.

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?*)
As described above, this measure is included in the CMS PQRS program. Additionally, although the measure is currently in use, we will continue to seek opportunities to advocate for expanded use of this measure in government or other programs.

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 has also been included in America´s Health Insurance Plans Medical Oncology Core Measure Set. The purpose of this program is to reduce variability in measure selection, specifications and implementation. The measures will be implemented nationally by private health plans using a phased-in approach. Contracts between physicians and private payers are individually negotiated and therefore come up for renewal at different points in time depending on the duration of the contract. It is anticipated that private payers will implement these core sets of measures as and when contracts come up for renewal or if existing contracts allow modification of the performance measure set. CMS is also working to align measures across public programs. They intend to include, for broad input, the agreed upon draft measure sets in the Physician Fee Schedule and other proposed rules. For measures that are not currently in CMS programs, CMS would go through the annual pre-rulemaking and rulemaking processes to solicit stakeholder and public input. Depending on public response, these measures will be included in a timeframe determined by the Agency.

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**

The performance rates show variation with no trend of improvement. These rates indicate the opportunity for continued performance improvement.

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.**

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.**
There have been no reports of unintended consequences with this measure.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

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

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

5a. Harmonization

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.)

Appendix

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

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** American Society of Clinical Oncology
Co.2 **Point of Contact:** Tayyaba, Shehzadi, Tayyaba.Shehzadi@asco.org, 571-483-1673-
Co.3 **Measure Developer if different from Measure Steward:** American Society of Clinical Oncology
Co.4 **Point of Contact:** Tayyaba, Shehzadi, Tayyaba.Shehzadi@asco.org, 571-483-1673-

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.
ASCO Measure Development Panel

The panel is responsible for reviewing evidence and maintaining measures

Tracey Evans, MD (Chair)
University of Pennsylvania

Craig Earle, MD, FASCO (Co-Chair)
Institute for Clinical Evaluative Science

Katherine Ast, MSW, LCSW
American Academy of Hospice and Palliative Medicine

Amy Berman
The John A. Hartford Foundation

Kathleen Bickel, MD, MPhil
White River Junction VA Medical Center

Eduardo Bruera, MD
The University of Texas MD Anderson Cancer Center

Sydney Dy, MD
Johns Hopkins

Esme Finlay, MD
University of New Mexico Cancer Research and Treatment Center

Arif Kamal, MD, MHS, FAAHPM
Duke University

Kristen McNiff, MPH
Dana-Farber Cancer Institute

Michael Neuss, MD, FASCO
Vanderbilt Ingram Cancer Center

John Sprandio, MD
Consultant in Med Onc and Hem Inc

Holley Stallings, RN
Norton Cancer Institute

Jamie Von Roenn, MD, FASCO
American Society of Clinical Oncology

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2005
Ad.3 **Month and Year of most recent revision:** 11, 2015
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?** 12, 2017

Ad.7 **Disclaimers:** These clinical indicators and quality measures are not intended to and should never supplant independent physician judgment with respect to particular patients or clinical situations. Patient care is always subject to the independent physician judgment with respect to particular patients or clinical situations. Patient care is always subject to the independent professional judgment of the treating physician.

Accordingly, QOPI participants' adherence to quality measures contained in this research report is strictly voluntary and discretionary, with the ultimate determination regarding their application to be made by the treating physician in his or her professional judgment and in light of each patient's individual circumstances. ASCO does not endorse the QOPI® measures as guidelines for standards of practice or 'best practices.'

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

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 #: 0211
Measure Title: Proportion of patients who died from cancer with more than one emergency department visit in the last 30 days of life
Measure Steward: American Society of Clinical Oncology
Brief Description of Measure: Proportion of patients who died from cancer with more than one emergency department visit in the last 30 days of life
Developer Rationale: Studies suggest that cancer treatments and care continue to be more aggressive than desired for patients at the end of life. Emergency department (ED) visits in the last 30 days of life are one indicator that supportive care may not be provided effectively to these patients (Guadagnolo, 2015). In general, unnecessary ED visits should be avoided for those concerns that can be addressed at the practice or clinic. For example, a study at Memorial Cancer Institute found that 48% of ED visits occurred during office hours in patients with cancer and many were for concerns that did not require the use of ED services (Hunis, 2016). For patients with cancer at the end of life, the use of unnecessary services such as the ED can negatively impact a patient and family's quality of life and satisfaction with end of life care (Barbera, 2010).

Barbera, L., C. Taylor, et al. (2010). "Why do patients with cancer visit the emergency department near the end of life?" CMAJ 182(6): 563-568.

Guadagnolo, B. A., K. P. Liao, et al. (2015). "Variation in Intensity and Costs of Care by Payer and Race for Patients Dying of Cancer in Texas: An Analysis of Registry-linked Medicaid, Medicare, and Dually Eligible Claims Data." Med Care 53(7): 591-598.

Hunis, B., A. J. Alencar, et al. (2016). "Making steps to decrease emergency room visits in patients with cancer: Our experience after participating in the ASCO Quality Training Program." J Clin Oncol 34, 2016 (suppl 7S; abstr 51) Presented at the ASCO Quality Care Symposium, February 26th, 2016, Phoenix, AZ.

Numerator Statement: Patients who died from cancer and had at least one emergency department visit in the last 30 days of life
Denominator Statement: Patients who died from cancer
Denominator Exclusions: None

Measure Type: Intermediate Clinical Outcome
Data Source: Administrative claims, Electronic Clinical Data : Registry
Level of Analysis: Clinician : Group/Practice

IF Endorsement Maintenance – Original Endorsement Date: Aug 10, 2009 **Most Recent Endorsement Date:** Aug 09, 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**

Summary of prior review in 2012 :

- The developer cited one retrospective cohort study that examined trends in the aggressiveness of end-of-life (EOL) cancer care in Ontario, Canada, between 1993 and 2004. Aggressiveness of EOL care was defined as the occurrence of: chemotherapy received within 14 days of death, more than one emergency department (ED) visit within 30 days of death, more than one hospitalization within 30 days of death, or at least one intensive care unit (ICU) admission within 30 days of death.
- The developer also referenced an underlying expert consensus statement from 2003 that identified potential indicators of quality of end-of-life cancer care using administrative data.
- In the previous evaluation of the measure, the Committee noted that in some cases more than one visit to the ED during the last days of life is appropriate, but agreed that in most cases, overutilization of ED services for the actively dying is inappropriate and distressing for patients. Members agreed that this measure is useful for detecting patterns in practice and variation in performance, identifying outliers when comparing similar practices with similar patient populations, addressing patient preference and overtreatment at the end of life, and reflecting disparities in access to care and the capacity of the local healthcare system to treat patients appropriately at the end of life.

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 provides a diagram of the relationship of this intermediate outcome (more than one ED visit in the last 30 days of life) to better patient outcomes (positive death experience, improved patient and caregiver/family satisfaction, and reduced resource utilization and costs).
- The developers reference five sources as evidence. However, none of these supports the relationship of reduced ED visits to desired patient outcomes.
    - Guadagnolo, 2015, examined EOL care in last 30 days of life for Medicaid, Medicare, and dually eligible beneficiaries dying of cancer in Texas from 2000 to 2008. Results indicate that >1 ED visit was more likely for Medicaid patients (compared to Medicare patients), dually eligible patients (compared to Medicare-only beneficiaries), and Black and Hispanic patients (compared to white patients).
    - Hunis, 2016, after finding that 48% of their oncology patients' ED visits happened during office hours, implemented a triage system at their patient access center, which resulted in a 60% decrease in the number of patients utilizing the ED.
    - Barbera, 2010, examined the diagnosis codes for Ontario cancer patients who visited the ED during the last 6 months of life. They concluded that many such visits may be avoidable.
    - A 2013 Cochrane Collaborative systematic review evaluated the impact of home-based palliative care services on several patient and caregiver outcomes. This review found that for patients with cancer, home-based palliative care services increases the chance of dying at home and reduces symptom burden, but for the most part did not significantly reduce the number of ED visits.
    - A 2012 provisional clinical opinion from the American Society of Clinical Oncology addressed integration of palliative care services into standard oncology care. This panel concluded that while there is need for

additional research, several of the available studies show evidence of benefit, and recommended that palliative care should be considered early in the course of illness for any patient with metastatic cancer and/or with high symptom burden.

Exception to evidence: None of the empirical evidence submitted links reduced ED visits to desired patient outcomes. However, a similar measure was recommended by an expert panel charged with exploring potential indicators of the quality of end-of-life services for cancer patients that could be monitored using existing administrative data (Earle, et al).

Guidance from the Evidence Algorithm
Intermediate clinical outcome measure, but no systematic review of evidence relevant to the process being measured (Box 3) → Empirical evidence without systematic review, but does not link ED visits to desired outcomes (Box 7) → Systematic assessment of expert opinion regarding measure (Earle, et al.) (Box 11) → vote INSUFFICIENT WITH EXCEPTION

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?*
- *Does the Committee know of any empirical evidence supporting this process to desired patient outcomes?*

Preliminary rating for evidence: ☐ **High** ☐ **Moderate** ☐ **Low** ☒ **Insufficient**
Rationale: Use of the evidence exception may be warranted unless empirical evidence linking this measure to desired health outcomes is identified.

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.

- Data from 2 integrated health systems (IHS) was provided. These appear to be system-level results rather than clinician group/practice performance results, although the measure is specified for clinician groups/practices. ***The performance results from the two sources are substantially different***.
  - The first IHS used the sampling methodology from ASCO's Quality Oncology Practice Initiative (QOPI®) Registry (a minimum of 40 cases twice each year). No data on the number of patients represented in these results were provided.

	Fall 2011	Spring 2012	Fall 2012	Spring 2013
Overall Performance %	35.00	47.50	55.00	43.90

  - The second IHS used data derived from a combination of electronic clinical data and tumor registry data for patients who died of cancer between June 2013 and May 2015.

Numerator	801
Denominator	15,098
% of Total	5.31%
Mean (2 year)	5.38%
Standard Deviation (2 year)	0.91%
Minimum (2 year)	4.05%
Maximum (2 year)	7.68%

Disparities

- The developer provides data from 2 integrated health systems.
- Results from Guadagnolo, et al. (2015) indicate patient-level disparities by racial/ethnic status and payer type

Questions for the Committee:

- *Do these results demonstrate a gap in care for clinician group/practices that warrants a national performance*

measure?

- *The results from the two integrated systems are very different. Does the variation—and particularly the very high rates for system 1—give any cause for concern?*
- *Does the Committee have additional evidence that there are practice-level disparities in ED visits in the last 30 days of life for cancer patients?*

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.

* Updated evidence was provided in the form of 5 sources. The developer also provided a diagram of the relationship.None of the evidence links reduced ED visits to desired patient outcomes. It is not clear how a "positive death" is measured. The evidence is not directly applicable to the desired patient outcomes.
* Intermediate clinical outcome. Evidence supporting a link between this outcome and better patient care is weak and poorly supported by available evidence.
* The desired outcome is appropriate care at the end of life. Authors posit that unnecessary or avoidable ED visits negatively impact pt/family QoL. None of the studies directly linked ED visits to QoL except for consensus opinion by Earle et al. Measure does not look just at avoidable ED visits but all ED visits.
* This is a measure of an intermediate outcoe. the evidence previously focused on one retrospective cohort study and expert consensus.

The additional citations include a provisional cliical opinion from ASCO re integration of palliative care services into oncology care and a cochrane collaborative review that conculdes that patients with cancer benefit from home-based pallaitve care but for the most part did not reduce ED visits. The strenth of evidence for reduction of ED visits was raised. Qualitative subanalysis of Edes et al's JAGS publication about the outcomes of home-based primary care for older adults with multimorbidity and serious illness (not namely cancer) did describe reduced anxiety by having alternatives to ED visit, and most of the cost-savings was related to a 25% reduction in hospitalization (in part presumably albeit not directly stated, due to a reduction in ED visits)

1b.

*The data presented from 2 separate IHS are substantially different. No data on the number of patients represented was provided. Disparities data was provided. The discrepancies in the data raises questions of validity of the measure.
* Variation in ED visits in the last month of life have been demonstrated by the Dartmouth Atlas and in data provided by the developer. Reasons for this variation have not been well studied and may result from poor provider/system performance or may be due to community factors beyond the provider's control. Data on average baseline rates would be helpful.
* Moderate evidence for a performance gap--rates ranging from 5% to 55%.
* data from 2 integrated health systems is consistent with systems level gaps. There is also data on racial and ethnic status.

The committee was asked for additioanl evidence for practice-level disparities:
While not ED nor cancer specific Teno's 2013 analysis of Medicare's rise in the number of care transitions, most certainly includes a rise i ED use.

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 sources for this measure include administrative claims and registry. Unlike several other ASCO measures (e.g., #0210, #0215, #0216), this measure is not currently in use in ASCO's Quality Oncology Practice Initiative (QOPI®) Registry.

Specifications:

- The measure is specified at clinician group/practice level of analysis.
- A lower score on the measure is indicative of better quality.
- Care settings for this measure include clinician offices or clinics, hospices, and hospitals/acute care facilities.
- The numerator includes those with at least one emergency department visit in the last 30 days of life.
- The denominator includes patients who died from cancer. Presumably these patients can be identified in a registry. It is unclear if/how the denominator would be identified through claims data. The developers briefly mention use of the "death registry" but do not explain how this is used in conjunction with claims data.
- There are no exclusions for this measure.
- A calculation algorithm is provided.
- This measure is not risk-adjusted.

Questions for the Committee :

- *Is use of death registry data required if the measure is calculated from claims data? If so, what is the timeliness of this dataset? If the death registry is not used, can the measure denominator be identified reliably through administrative data?*
- *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:

- For claims data, data element validity testing was conducted by comparing claims data to the full medical record. NQF guidance indicates that if data element validity is demonstrated, additional data element reliability testing is not required.

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:

- For claims data, see method(s) of validity testing, below.
- For registry data, no testing results were provided.

Results of reliability testing:

- For claims data, see results of validity testing, below.

Guidance from the Reliability Algorithm

For registry data: Precise specifications (Box 1) → no empirical reliability testing conducted (Box 2) → no empirical validity testing conducted (Box 3) → Insufficient

For claims data: Fairly precise specifications, except for denominator (Box 1) → no empirical reliability testing conducted for claims data (Box 2) →empirical validity testing conducted (Box 3) → FROM VALDITY ALGORITHM: data

element testing for numerator (Box 10) → Accuracy of the denominator data element not demonstrated for claims data, but assume okay if data derived from national death registry (Box 11) → Moderate

Questions for the Committee:

- *Can claims adequately identify cancer deaths? If not, is use of death registry data required?*
- *Is the test sample adequate to generalize for widespread implementation?*

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

Rationale: For the claims data, the measure is eligible for a MODERATE rating. However, testing with registry data is required if the developer wants the measure endorsed for the registry data source. One option would be for the developer to remove the registry option from the measure specifications.

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**

Specification not completely consistent with evidence

- None of the empirical evidence submitted links reduced ED visits to desired patient outcomes, but supported by expert consensus.

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:

- For claims, data element validity testing was done by comparing claims data to the full medical record. This is an appropriate method of data element validation.

Describe any updates to validity testing:

- No updated testing was provided.

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:

- For administrative claims data:
  - Claims data for 150 consecutive patients treated for advanced cancer at Boston's Dana-Farber Cancer Institute and Brigham and Women´s Hospital were compared to data from the full medical record. Numerator data elements were compared ***but not the denominator data element***. Dates for the data examined in testing were not provided.
- Face validity assessment:

- The developer states that face validity was assessed by conducting focus group and structured interviews with end-of-life cancer patients and bereaved caregivers, and then vetted by an expert panel of cancer providers. According to NQF guidance, the face validity of the measure score as a quality indicator may be adequate if it explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. There is not enough information provided to know if the face validity assessment conforms to NQF's requirements.

Validity testing results:

- For administrative claims data:
  - The developer states "*The measure was 89% accurate (percent true positives + true negatives).*"
  - Sensitivity and Specificity values were not provided.
- Face validity assessment:
  - The developer does not provide results from the face validity assessment.

Questions for the Committee:

- *The developer has not provided any new validity testing, even though there is no testing for the denominator element (cancer deaths) for claims data. Does the developer expect the denominator to be identified via a death registry? If so, should these data be assumed valid? If not, is there evidence that claims adequately identify cancer deaths?*
- *For claims data, is the test sample adequate to generalize for widespread implementation?*
- *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:

- There are no exclusions for this measure.

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

Conceptual rationale for SDS factors included? ☐ **Yes** ☒ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary

- In the ***previous evaluation*** of the measure,
  - The developer stated that "*No risk adjustment or risk stratification is necessary because the measure is intended to be used for comparison among similar providers; unless there is a reason to believe that one providers' patients have significantly different risks than others, it will not affect relative comparisons. Since, however, comorbidity risks could increase the likelihood of experiencing this process of care, stratification or adjustment as described above can be considered.*"
  - The Committee agreed that this measure is useful for detecting variations in performance and identifying outliers when comparing similar practices with similar patient populations.
- The developer is now considering whether risk-adjustment for this measure is appropriate.
- No empirical analysis is provided to support lack of adjustment and nothing in the specifications requires comparison of results across similar providers.

Questions for the Committee:

- *Do you believe risk-adjustment is needed for this measure?*
- *Do you have thoughts about which (if any) SDS factors should be considered if the measure is risk-adjusted?*

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

- When initially developed (using 1991-1996 SEER-Medicare linked data), a benchmark of <4% was achieved by the highest performing SEER regions.
- Variation within 2 integrated health systems is shown via descriptive statistics about performance results (see section 1b, above). It is not clear, however, whether either set of results reflect clinician group/practice performance.

Question for the Committee:

○ *Does this measure identify meaningful differences about quality for clinician groups/practices?*

2b6. Comparability of data sources/methods:

- The developer has not provided analysis to assure comparability of results between claims versus registry (the analysis of claims data across time presented in the earlier evaluation does not speak to this topic).

2b7. Missing Data

- The developer describes how missing data are handled in section S.22.

Guidance from the Validity Algorithm:
For claims data: Specifications somewhat consistent with evidence (Box 1) → potential threats to not completely assessed (Box2) →Insufficient OR, if no concerns about risk-adjustment/comparability, → empirical testing conducted (Box 3) → no score-level testing (Box 6) → data element testing conducted (Box 10) → Accuracy of the denominator data element not demonstrated for claims data, but assume okay if data derived from national death registry (Box 11) → Moderate

For registry data: Specifications somewhat consistent with evidence (Box 1) → potential threats to not completely assessed (Box2) →Insufficient OR, if no concerns about risk-adjustment/comparability, → no empirical testing conducted (Box 3) → face validity conducted, but not enough information to know if it conforms to NQF requirements (Box 4) → Insufficient

Preliminary rating for validity: ☐ **High** ☐ **Moderate** ☐ **Low** ☒ **Insufficient**
Rationale: Additional analysis to assess threats to validity needed (risk-adjustment/comparability). For claims data, testing for the denominator data element for claims was not conducted, but may not be needed if data required to be obtained from a death registry (and potentially eligible for a MODERATE rating). For registry data, no testing data presented; unclear if face validity testing meets NQF requirements. Concerns with testing and comparability become moot if the developer removes the registry option from the measure specifications.

Committee pre-evaluation comments

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

2a, 2b.
Reliability – Specifications
* This measure unlike the others is not currently using the ASCO QOPI Registry. The definition of the data element "died from cancer" is not clear. The algorithm is clear lack of supportive care leads to increased ED visits.
* Identifying the denominator may be problematic if not available in claims. Absence of risk adjustment is concerning.
* Numerator specifications are clearly defined. Denominator specifications could be ascertained through a registry but not claims data.
* Performance is calculated as:
1. Identify those patients that meet the denominator criteria defined in the measure.
2. Subtract those patients with a denominator exclusion from the denominator. Note: this measure does not have exclusions.
3. From the patients who qualify for the denominator (after any exclusions are removed), identify those who meet the numerator criteria.
4. Calculation: Numerator/Denominator-Denominator Exclusions

I can get that, especially if there are no exclusion.

Validity – Specifications
* The evidence provided is opinion and does not link ED visits to desired patient outcomes.
* Acceptable. Expert consensus (including patient groups and CMS) consider unwanted ED visits an adverse patient outcome
* The specifications are for ALL ED visits not those that are unwanted, avoidable or unnecessary.
* It is potentially concerning that the claims data woudn't identify cancer deaths
89% accuracy figure noted. No new testing.

Reliability – Testing
* It is not clear if the test sample is adequate to generalize for widespread implementation due to the lack of number of patients.
* Death registry data will likely be required for denominator determination.
* For registry data, no testing results provided. For claims data, no empirical reliability testing for claims data.
* The concerns raised merit discussion. A little unclear to me.

Validity Testing
* It is not clear if the report from the focus group can be used to distinguish good from poor quality. Sensitivity and specificity were not provided but accuracy was reported to be 89%.
* Absence of denominator data are concerning.
* Numerator elements were compared to the EMR data; similar comparison not conducted for denominator data element. Face validity tested through focus groups but enough data to fully evaluate.
* I appreciate that Box 2 claims appears Insufficient, Box 3 has no core level testing and for claims Box 2 appears insufficient and box 4 appears insufficient.

Threats to Validity
* The developers stated that risk adjustment and risk stratification were not necessary because the measure is used to compare similar providers. They are not clear on what the similarities are. They are now considering if risk adjustment is appropriate.
* Missing death data from claims represents a threat.
* Missing data on denominator for claims data.
* Risk Adjustment is being considered but not done. The differences between 2 diferrent IHS areas is concerning but the issue could be as simple as difference in the sampling methodology not true care process/ quality issues such as earlier delivery of hospice care in a system leading to better home-based care with associated ED use reduction.
claims data test sample updated information of 1 institution 150 patients

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.

- All data elements are in defined fields in electronic clinical data. Most data elements are routinely generated during care delivery.
- Cancer decedents should be easily identifiable in the registry, but may not be easily or accurately identifiable through claims. Use of a death registry may be needed if measure calculated via claims.
- No feasibility concerns were raised by the Cancer Steering Committee during the NQF Maintenance of Endorsement review in 2012.

Questions for the Committee:
- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*
- *Is use of a death registry required? If not, how will the measure denominator be identified using claims data?*

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

Committee pre-evaluation comments

Criteria 3: Feasibility

* The required data elements are routinely generated during care delivery. Cancer decedents can be easily identified in the registry but not by claims. The elements are available in the EHR.
* Yes with the exception of mortality data
* No feasibility concerns.
* Data elements seem reasonably available in electronic data.

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

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 measure is part of America's Health Insurance Plans (AHIP) Medical Oncology Core Measure Set. The AHIP effort is a collaboration of both public and private stakeholders to identify measures that are meaningful to patients, consumers, and physicians and to reduce variability in measure selection, collection burden, and cost. Payers involved in the collaboration have committed to using for reporting as soon as feasible. By virtue of being included in the AHIP measure set, CMS will consider this measure for inclusion in Medicare quality programs.

Improvement results: No performance trend can be inferred from the data presented: four semesters from Spring 2011-Fall 2013 in one integrated health system.

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: From previous evaluation:

- Steering Committee members raised concerns about use of the measure given the [then] current systemic issues with access to quality hospice facilities. The Committee believed patients may utilize emergency department services when good hospice care is not available. In areas where performance of the measure is poor, it will call attention to a lack of resources available for patients at the end of life.

***Questions for the Committee**:*

- *NQF guidelines strongly recommended measures be publicly reported following six years of endorsement. Does the measure have a viable path to public reporting and eventual accountability applications?*
- *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**

Committee pre-evaluation comments

Criteria 4: Usability and Use

* The measure is not being publicly reported but is is part of AHIP and will be included in the Medicare quality programs. There is a viable path to public reporting. No potential harms were reported.
* Lack of risk adjustment may unfairly penalize providers in poorly serviced areas with inadequate hospice services or poor access to care. Unclear what the baseline performance level should be.
* In AHIP. No usability concerns.
* The measure is part of America's Health Insurance Plans medical oncology core measure set.

the measure is not publicly reported

I think one of the attractive things with this measure is the focus on >1 ED visit. There is to me a meaningful difference between the need for 1 visit to ED, such as an acute symptom vs multiple ED visits which implies complex process failures.

I am thinking of the unexpected aspect of ED visits and a recent patient event comes to mind: a non-emergent procedure needed to be performed and because the provider couldn't manage to arrange for it to occur, the family was able to meet the provider in the ED to expedite the event. Is that good care? the Patient got what was needed and went home. The provider was able to get around clinic shortage.
I think the more important thing in this measure is the ability to demonstrate variation in social support and when the ED is used for lower quality EOL care, but in the example above that might have been good delivery with limited resources.

Criterion 5: Related and Competing Measures

Related or competing measures

0210: Proportion of patients who died from cancer receiving chemotherapy in the last 14 days of life
0213: Proportion of patients who died from cancer admitted to the ICU in the last 30 days of life
0215: Proportion of patients who died from cancer not admitted to hospice
0216: Proportion of patients who died from cancer admitted to hospice for less than 3 days

Harmonization : N/A

Pre-meeting public and member comments

Name: Michele Galioto, RN, MSN

Organization: Oncology Nursing Society

Comment: Should the measure include immunotherapy? Does “chemotherapy” include hormonal and biotherapy? It may be more inclusive to refer to all as "antineoplastic therapy."

Should the measure include radiation therapy as well?

Does “death from cancer” include all death within 14 days? Death may be the result of infection, accident (e.g., fall), bleeding, etc. which could be tied to cancer or cancer treatment. Death attributed to side effects of therapy may be indistinguishable from cancer deaths. Is the intent that death occurs within a timeframe of receiving chemotherapy?

Does “emergency room” apply to other urgent care facilities?

How is this data captured if they are seen out of network?

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.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0211

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

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): Click here to name the PRO

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

☒ Intermediate clinical outcome (*e.g., lab value*): Patients who died from cancer with more than one emergency department visit in the 30 last days of life

☐ 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*

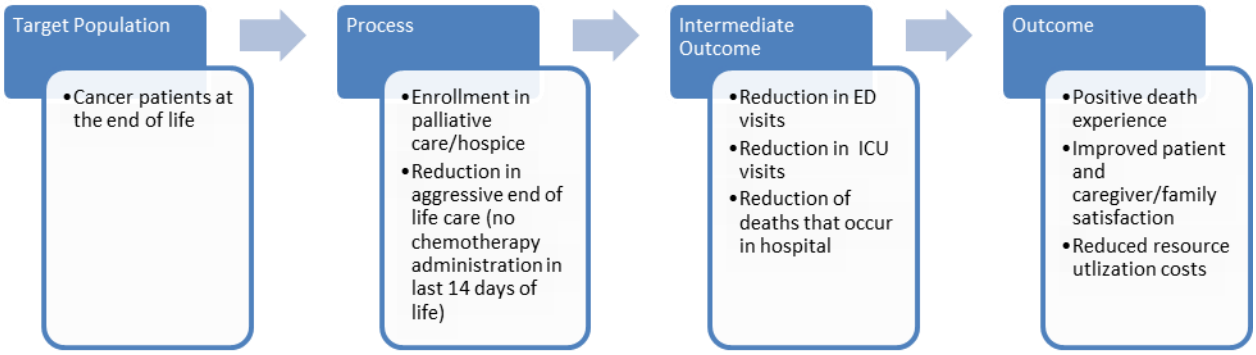
1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

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*).

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.



Studies suggest that cancer treatments and care continue to be more aggressive than desired for patients at the end of life. Emergency department (ED) visits in the last 30 days of life are one indicator that supportive care may not be provided effectively to these patients (Guadagnolo, 2015). In general, unnecessary ED visits should be avoided for those concerns that can be addressed at the practice or clinic. For example, a study at Memorial Cancer Institute found that 48% of ED visits occurred during office hours in patients with cancer and many were for concerns that did not require the use of ED services (Hunis, 2016). For patients with cancer at the end of life, the use of unnecessary services such as the ED can negatively impact a patient and family's quality of life and satisfaction with end of life care (Barbera, 2010).

Barbera, L., C. Taylor, et al. (2010). "Why do patients with cancer visit the emergency department near the end of life?" CMAJ 182(6): 563-568.

Guadagnolo, B. A., K. P. Liao, et al. (2015). "Variation in Intensity and Costs of Care by Payer and Race for Patients Dying of Cancer in Texas: An Analysis of Registry-linked Medicaid, Medicare, and Dually Eligible Claims Data." Med Care 53(7): 591-598.

Hunis, B., A. J. Alencar, et al. (2016). "Making steps to decrease emergency room visits in patients with cancer: Our experience after participating in the ASCO Quality Training Program." J Clin Oncol 34, 2016 (suppl 7S; abstr 51) Presented at the ASCO Quality Care Symposium, February 26th, 2016, Phoenix, AZ.

2012 Submission:

A structural feature: regional availability of hospice, has been shown to correlate with a composite measure of the aggressiveness of cancer care near the end of life that contains this measure. Mostly it is a process measure indicating a possible inadequate focus on palliation and supportive care, that can affect quality of life.

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***

2012 Submission: Selected individual studies (rather than entire body of evidence)

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*):

Smith TJ, Temin S, Alesi ER, et al. American Society of Clinical Oncology Provisional Clinical Opinion: The Integration of Palliative Care into Standard Oncology Care. J Clin Oncol 2012;30:880-887. Available at: http://www.instituteforquality.org/asco-provisional-clinical-opinion-integration-palliative-care-standard-oncology-care.

Gomes, B., N. Calanzani, et al. (2013). "Effectiveness and cost-effectiveness of home palliative care services for adults with advanced illness and their caregivers." Cochrane Database Syst Rev 6: CD007760 Available at: http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD007760.pub2/pdf.

2012 Submission: The underlying evidence was obtained by expert consensus, as described in Earle CC, Park ER, Lai B, Weeks JC, Ayanian JZ, Block S. Identifying potential indicators of the quality of end of life cancer care from administrative data. J Clin Oncol. 2003;21(6):1133-8. The panel consisted of oncologists, nurses, palliative care specialists, etc, and used a modified Delphi process to evaluate measures.

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?

A 2012 American Society of Clinical Oncology (ASCO) Provisional Clinical Opinion (PCO) addresses the integration of palliative care (PC) services into standard oncology care at the time a person is diagnosed with metastatic cancer and/or high symptom burden.

A 2013 Cochrane Review, 'Effectiveness and cost-effectiveness of home palliative care services for adults with advanced illness and their caregivers', evaluated the impact of home palliative care services on outcomes for adults with advanced illness or their family caregivers, or both. The aim of the review was to quantify the effect of home palliative care services on a patients' odds of dying at home, examine the clinical effectiveness of home palliative care services on other outcomes such as symptom control, quality of life, caregiver distress and satisfaction with care, and comparing resource use and costs associated with these services.

2012 Submission: The argument is made that because providers cannot predict the future, measures based on decedent cohorts are unfair. However, as described above in 1a.a, the idea is for the measure to be seen as an overall indication of practice style and/or available palliative resources. An individual patient experiencing this process of care has not necessarily received poor quality care. If explanations other than practice style and resource availability, such as unusually poor prognostic ability on the part of the provider or unexpected toxic deaths (whether unavoidable, from overly aggressive treatment, or poor patient selection) are enough to influence the overall aggregate rates, it is still justifiable to consider it a 'red flag' that should prompt examination of the care provided.

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

2012 ASCO PCO (p. 881):

The American Society of Clinical Oncology (ASCO) has established a rigorous, evidence-based approach—the provisional clinical opinion (PCO)—to offer a rapid response to emerging data in clinical oncology. The PCO is intended to offer timely clinical direction to ASCO's oncologists after publication or presentation of potentially practice- changing data from major studies.

The PCO may serve in some cases as interim direction to the membership pending the development or updating of an ASCO clinical practice guideline. As such, the evidence is not graded in a PCO and is a result of expert consensus. A clinical guideline on palliative care integration with recommendations and the associated grading is under development,

2013 Cochrane Review (p. 12):

Two independent reviewers assessed all included studies for methodological quality using the standard criteria developed by the Cochrane EPOC Review Group for RCTs/CCTs, CBAs and ITSs. The checklist for RCTs/CCTs contains seven items qualified as done, unclear and not done for concealment of allocation, follow-up of professionals, follow up of patients or episodes of care, blinded assessment of primary outcome(s), baseline assessment, reliable primary outcome measure(s) and protection against contamination. Blinding and reliability of all outcomes were also assessed.

Each criterion was scored zero (not done), 0.5 (not clear or when scores varied across outcomes) and one (done). Total scores for RCTs/ CCTs ranged from zero to six; studies with a score of 3.5 or above were considered of high quality. Integration of the results of the quality assessment in data analysis was done in addition to meta-analyses with sensitivity analyses including only high quality studies.

2012 Submission: The studies are qualitative and observational using administrative data, consequently there are limitations to the quality of the data.

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

See 1a.7.2 for this information.

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

2012 ASCO PCO: 2004-2012

2013 Cochrane Review: 1950 – November 2012

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*)

2012 ASCO PCO: 7 randomized controlled trials

2013 Cochrane Review: 5 randomized controlled trials and 2 controlled clinical trials

2012 Submission: 4

1a.7.6. What is the overall quality of evidence <u>across studies</u> 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*)

2012 ASCO PCO:

This PCO did not provide an assessment of the overall quality of evidence across the studies. This analysis will be completed during the development of the upcoming clinical guideline.

2013 Cochrane Review :

p. 3: The direction of the effect was consistent across all studies but did not reach statistical significance in 3; ORs ranged from 1. 36 (95% CI 0.80 to 2.31) to 2. 86 (95% CI 0.78 to 10.53) Sensitivity analyses showed that exclusion of the 2 CCTs (both of Swedish hospital-based services with a pooled OR 3.44, 95% CI 0.60 to 19.57) and inclusion of only high quality RCTs resulted in a reduction of the OR to 1.28 (95% CI 1.28 to 2.33) and 1.75 (95% CI 1.24 to 2. 47) respectively, with more precision and less heterogeneity.

p. 22:

Pooled data from seven studies (five RCTs, three of high quality, and two CCTs with 1222 participants) showed that those receiving home palliative care had statistically significantly higher odds of dying at home than those receiving usual care (OR 2.21, 95% CI 1.31 to 3.71; Z = 2.98, P value = 0.003; Chi^2 = 20.57, degrees of freedom (df) = 6, P value = 0.002; I^2 = 71%). The study population control risk was of 307 home deaths per 1000 deaths; based on this ACR of 0.307, the NNTB was 5 (95% CI 3 to 14), meaning that for one additional patient to die at home five more would need to receive home palliative care as opposed to usual care. Assuming a medium cancer home death rate population ACR of 0.278 (i.e. 278 home deaths per 1000 cancer deaths), the NNTB was 6 (95% CI 3 to 15). This means that for one additional cancer patient to die at home in a population where there are 278 home deaths per 1000 cancer deaths, six more would need to receive home palliative care. NNTB estimates ranged from 9 patients (95% CI 5 to 16) when applied to a low home death rate population such as the one observed in Norway (128 home deaths per 1000 cancer deaths) to 5 patients (95% CI 3 to 13) when applied to a high home death rate population such as the one observed in the Netherlands (454 home deaths per 1000 cancer deaths).

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) <u>across studies</u> 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*)

2012 ASCO PCO (p. 884):

Seven published randomized trials demonstrate the feasibility of providing various components of PC alongside usual oncology care. There is, however, a dearth of data evaluating the integration of modern PC practices into standard oncology care, especially in concert with ongoing antitumor therapy. Overall, the addition of PC interventions to standard oncology care delivered via different models to patients with cancer provided evidence of benefit.

2013 Cochrane Review (p. 22):

The study population control risk was of 307 home deaths per 1000 deaths; based on this ACR of 0.307, the

NNTB was 5 (95% CI 3 to 14), meaning that for one additional patient to die at home five more would need to receive home palliative care as opposed to usual care. Assuming a medium cancer home death rate population ACR of 0.278 (i.e. 278 home deaths per 1000 cancer deaths), the NNTB was 6 (95% CI 3 to 15). This means that for one additional cancer patient to die at home in a population where there are 278 home deaths per 1000 cancer deaths, six more would need to receive home palliative care. NNTB estimates ranged from 9 patients (95% CI 5 to 16) when applied to a low home death rate population such as the one observed in Norway (128 home deaths per 1000 cancer deaths) to 5 patients (95% CI 3 to 13) when applied to a high home death rate population such as the one observed in the Netherlands (454 home deaths per 1000 cancer deaths).

2012 Submission: All studies have shown similar results. As per Ho TH, Barbera L, Saskin R, Lu H, Neville BA, Earle CC.Trends in the Aggressiveness of End-of-Life Cancer Care in the Universal Health Care System of Ontario, Canada. J Clin Oncol April 20, 2011 vol. 29 no. 12 1587-1591, rates were similar in Canada and among U.S. Medicare patients.

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

2012 ASCO PCO (p. 884-885):

No harm to any patient was observed in any trial, even with discussions of EOL planning, such as hospice and ADs. Two of five trials measuring change in symptoms, two of five studies measuring QOL, two of three studies measuring patient/caregiver satisfaction, and one of three studies measuring survival found statistically significant improvements with PC. Three of six studies measuring mood, two of five studies measuring resource use, and one of four studies measuring outcomes of advance care planning found statistically significant differences, and one outcome of borderline significance was also found in each of these three areas, Therefore, most trials showed benefits ranging from equal to improved overall survival, reduced depression, improved caregiver and/or patient QOL, and overall lower resource use and cost because EOL hospitalizations were avoided.

2013 Cochrane Review: Discussion of harms was not addressed.

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.

No relevant studies have been conducted and published since the systematic reviews.

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.

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
0211_Evidence_Form_3.16.16.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)*
Studies suggest that cancer treatments and care continue to be more aggressive than desired for patients at the end of life. Emergency department (ED) visits in the last 30 days of life are one indicator that supportive care may not be provided effectively to these patients (Guadagnolo, 2015). In general, unnecessary ED visits should be avoided for those concerns that can be addressed at the practice or clinic. For example, a study at Memorial Cancer Institute found that 48% of ED visits occurred during office hours in patients with cancer and many were for concerns that did not require the use of ED services (Hunis, 2016). For patients with cancer at the end of life, the use of unnecessary services such as the ED can negatively impact a patient and family's quality of life and satisfaction with end of life care (Barbera, 2010).

Barbera, L., C. Taylor, et al. (2010). "Why do patients with cancer visit the emergency department near the end of life?" CMAJ 182(6): 563-568.

Guadagnolo, B. A., K. P. Liao, et al. (2015). "Variation in Intensity and Costs of Care by Payer and Race for Patients Dying of Cancer in Texas: An Analysis of Registry-linked Medicaid, Medicare, and Dually Eligible Claims Data." Med Care 53(7): 591-598.

Hunis, B., A. J. Alencar, et al. (2016). "Making steps to decrease emergency room visits in patients with cancer: Our experience after participating in the ASCO Quality Training Program." J Clin Oncol 34, 2016 (suppl 7S; abstr 51) Presented at the ASCO Quality Care Symposium, February 26th, 2016, Phoenix, AZ.

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.*
Performance data on this measure was obtained from two independent integrated healthcare delivery systems in the United States. Both systems are located in the South.

The first integrated healthcare delivery system manually abstracted data from their EMR using the sampling methodology from ASCO's Quality Oncology Practice Initiative (QOPI®) Registry (a minimum of 40 cases twice each year).

Integrated delivery system #1
ED visits in the last 30 days of life

	Fall 2011	Spring 2012	Fall 2012	Spring 2013
Overall Performance %	35.00	47.50	55.00	43.90

Integrated delivery system #2

The second integrated healthcare delivery system´s scores were derived from death notification data in a tumor registry combined with electronic clinical data for patients who died from cancer between June 2013 and May 2015:

Based on reported deaths 6/1/2013-5/31/2015 (2 years rolling)

ED visits in the last 30 days of life

Numerator	801
Denominator	15098
% of Total	5.31%
Mean (2 year)	5.38%
Standard Deviation (2 year)	0.91%
Minimum (2 year)	4.05%
Maximum (2 year)	7.68%

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.*
Performance data on this measure was obtained from two independent integrated healthcare delivery systems in the United States. Both systems are located in the South.

The first integrated healthcare delivery system manually abstracted data from their EMR using the sampling methodology from ASCO's Quality Oncology Practice Initiative (QOPI®) Registry (a minimum of 40 cases twice each year). Data is reported at the practice level.

ED visits in the last 30 days of life

	Fall 2011	Spring 2012	Fall 2012	Spring 2013
Female	23.53	55.56	37.50	62.50
Male	43.48	40.91	66.67	17.65
Hispanic	53.85	65.00	70.59	45.45
White	18.75	31.25	45.45	35.71
Black	20.00	25.00	44.44	60.00
Other	33.33	0.00	50.00	0.00

The second integrated healthcare delivery system´s scores were derived from death notification data in a tumor registry combined with electronic clinical data for patients who died from cancer between June 2013 and May 2015:

Based on reported deaths 6/1/2013-5/31/2015 (2 years rolling)

ED visits in the last 30 days of life

N Female 6903
N Male 8195
% of Total Female 5.49%
% of Total Male 5.15%
Medicare numerator 260
Medicare denominator 6249
% Medicare 4.16%
Non-Medicare numerator 541
Non-Medicare denominator 8849
% Non-Medicare 6.11%

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**

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.**

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

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

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.)
No webpage available

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 have been made to the measure since the last endorsement

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 who died from cancer and had at least one emergency department visit in the last 30 days of life

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.)*
Last 30 days of life

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.
ED visits documented in MEDPAR claims

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Patients who died from cancer

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
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)*
Claims:Patients in the death registry with cancer as their cause of death. In the cited analyses by the measure submitter, this is a field in the cancer registry or denominator file not requiring specific codes. This may be different in other administrative data sets.

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)*
None

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)*

Not applicable

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)*
Not applicable

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 = Lower 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.)*
Performance is calculated as:
1. Identify those patients that meet the denominator criteria defined in the measure.
2. Subtract those patients with a denominator exclusion from the denominator. Note: this measure does not have exclusions.
3. From the patients who qualify for the denominator (after any exclusions are removed), identify those who meet the numerator criteria.
4. Calculation: Numerator/Denominator-Denominator Exclusions

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.
Not applicable

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.
Not applicable

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
This measure is specified with defined criteria and data elements. If a patient record does not include one or more of these components for the denominator, then patients are not considered eligible for the measure and not included.

If data to determine whether a patient should be considered for the numerator or exclusions is missing, then the numerator or

exclusions not considered to be met and the practice will not get credit for meeting performance for that patient.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Electronic Clinical Data : Registry

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.

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)*
Clinician : Group/Practice

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic, Hospice, 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.)*
Not applicable

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
0211_MeasureTesting_MSF5.0_Data_Update.doc

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.**
Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), 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)*
ALL data elements are in defined fields in electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)

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.**

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.
No feasibility assessment **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.
The measure and its specifications have been in place for several years and ASCO continues to monitor and ensure that the measure and its specifications are up-to-date for widespread use.

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)*.
Not applicable

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 0211	**NQF Project:** Cancer Project

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)***:**

The measure was developed using the Medicare claims of all continuously-enrolled patients who died of cancer after having been diagnosed in a SEER region between 1991 and 1996.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)***:**

Evaluation was carried out on 150 consecutive patients treated for advanced cancer at Dana-Farber Cancer Institute and Brigham and Women´s Hospital in Boston. The percent accuracy of death ascertainment for includion into this cohort is unknown but is likely high as the cancer registry regularly uses the death index for ascertainment. Ascertainment would be expected to be highly specific. Hospital billing claims were obtained and analyzed and the accuracy was compared to detailed medical record review.

2a2.3 Testing Results *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**

Sensitivity 0.82, Specificity 0.96, where sensitivity = # true positives (both claims and charts)/(# true positives + # false negatives, i.e., not in claims but present in charts) and specificity = # true negatives/(# true negatives + false positives, i.e., present in claims but not in charts).

2b. VALIDITY. Validity, Testing, including all Threats to Validity: 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:**

They are identical

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)***:**

Evaluation was carried out on 150 consecutive patients treated for advanced cancer at Dana-Farber Cancer Institute and Brigham and Women´s Hospital in Boston. Claims were obtained and analyzed and the accuracy was compared to detailed medical record review.

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

Face validity was determined by focus groups and structured interviews with end-of-life cancer patients and bereaved caregivers, and then vetted by an expert panel of cancer providers. The percent agreement between claims and medical record review was calculated.

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)***:**

The measure was 89% accurate(percent true positives + true negatives).

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)***:**

N/A

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

N/A

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)***:**

N/A

2b4.2 Analytic Method (*Describe methods and rationale for development and testing of risk model or risk stratification

including selection of factors/variables):

Traditionally, ASCO believed that risk adjustment or risk stratification was not critical for these palliative care measures. The measures are used for comparison among similar providers and there is no expectation that performance will be 0%.

ASCO is exploring whether risk adjustment is appropriate and if so how it could be accomplished for our outcome measures. ASCO's palliative care measures would be included in that assessment. While several studies indicate potential variables for consideration, additional time is needed to properly define a statistical risk model method and variables, including sociodemographic (SDS) factors now that they may be considered.

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)*:

N/A

2b4.4 **If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment:** No risk adjustment or risk stratification is necessary because the measure is intended to be used for comparison among similar providers; unless there is a reason to believe that one providers' patients have significantly different risks than others, it will not affect relative comparisons. Since, however, comorbidity risks coule increase the likelihood of experiencing this process of care, stratification or adjustment as described above can be considered.

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)*:

We used the Medicare claims of all 28,777 continuously-enrolled patients who died of cancer after having been diagnosed in a SEER region between 1991 and 1996. This was an analysis of SEER-Medicare linked data obtained from NCI (http://healthservices.cancer.gov/seermedicare/).

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

Benchmarks were established to identify the outlying 10th decile of practice: The proportion of patients experiencing each process of care in each Health Care Service Area (HCSA) was computed and ranked from best (least aggressive) to worst. A new cohort was created by sequentially adding HCSAs in order starting with the least aggressive until they contained at least 10% of the original cohort and the proportion experiencing each process of care was then recalculated to arrive at the 'Achievable Benchmark of Care). More detail on this, as well as a reference for the Achievable Benchmark of Care method can be found in our publication: Earle CC, Neville BA, Landrum ME, Souza JE, Weeks JC, Block SD, Grunfeld E, Ayanian JZ. Evaluating claims-based indicators of the intensity of end-of-life cancer care. Int J Qual Health Care. 2005;17(6):505-9.

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):

A benchmark target of < 4% of patients having multiple ER visits in the last 30 days of life corresponds to that achieved by the highest performing regions in the country.

Integrated Health System #2 (2 year range)	Mean	St. Dev.	Min	Max
	5.38	0.91	4.05	7.68

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)***:**

Administrative claims and chart review, as described above: 77 entities (HCSAs), 215,484 patients, between 1991 and 2000.

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

We have also assessed the stability of these measures over time by examining the stability of relative aggressive care over time. If the relative aggressiveness of a provider or organization´s practice appeared to change from year to year, then these measures might not be assessing a stable property of practice. To investigate this, we used hierarchical regression models to estimate regional variation in both levels and trends of each measure. We used as our geographic unit of analysis the Health Care Service Area (HCSA). HCSAs are groupings of Metropolitan Statistical Areas defined by the Centers for Medicare & Medicaid Services (CMS) based on observed patient flow patterns in Medicare for tertiary care. As such, each HCSA can be considered to be a self-contained regional health system with a related group of providers. We ranked each region according to the model-estimated rate of each indicator and computed the correlation among relative ranks of each region during the 10-year study period.

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)***:**

We observed significant variation both in levels of aggressive care and in trends in aggressiveness over time, but generally stability of regional practice patterns: Year to year correlation on this measure was 0.91, and over a 5 year span was 0.71. This provides supportive evidence of the reliability of these measures.

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:**

N/A
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

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)
Payment Program	Quality Improvement (Internal to the specific organization) Multiple Integrated Delivery Systems Not applicable

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

Multiple integrated delivery systems:

In use in multiple integrated delivery systems across the United States for quality improvement purposes. Because it is internal to the specific organization, we are unable to provide any additional information.

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?*)

We are continuously seeking opportunities to advocate for expanded use of this measure in government or other programs, including those intended for accountability or public reporting. For example, this measure was recently selected for inclusion in a Medical Oncology Core Measure Set supported by America's Health Insurance Plans and CMS. See section 4a.3. below for additional details.

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 has also been included in America´s Health Insurance Plans Medical Oncology Core Measure Set. The purpose of this program is to reduce variability in measure selection, specifications and implementation. The measures will be implemented nationally by private health plans using a phased-in approach. Contracts between physicians and private payers are individually negotiated and therefore come up for renewal at different points in time depending on the duration of the contract. It is anticipated that private payers will implement these core sets of measures as and when contracts come up for renewal or if existing contracts allow modification of the performance measure set. CMS is also working to align measures across public programs. They intend to include, for broad input, the agreed upon draft measure sets in the Physician Fee Schedule and other proposed rules. For measures that are not currently in CMS programs, CMS would go through the annual pre-rulemaking and rulemaking processes to solicit stakeholder and public input. Depending on public response, these measures will be included in a timeframe determined by the Agency.

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**

The performance rates show variation with no trend of improvement. There are differences across each measurement period, but given the limited data available conclusions about the significance of these variations cannot be determined.

These rates indicate the opportunity for continued performance improvement.

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.**

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.**
There have been no reports of unintended consequences with this measure.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

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

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

5a. Harmonization

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.)

Appendix

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

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** American Society of Clinical Oncology
Co.2 **Point of Contact:** Tayyaba, Shehzadi, Tayyaba.Shehzadi@asco.org, 571-483-1673-
Co.3 **Measure Developer if different from Measure Steward:** American Society of Clinical Oncology
Co.4 **Point of Contact:** Tayyaba, Shehzadi, Tayyaba.Shehzadi@asco.org, 571-483-1673-

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.
Palliative Measure Development Panel
The panel is responsible for reviewing evidence and maintaining measures

Tracey Evans, MD (Chair)
University of Pennsylvania

Craig Earle, MD, FASCO (Co-Chair)
Institute for Clinical Evaluative Science

Katherine Ast, MSW, LCSW
American Academy of Hospice and Palliative Medicine

Amy Berman
The John A. Hartford Foundation

Kathleen Bickel, MD, MPhil
White River Junction VA Medical Center

Eduardo Bruera, MD
The University of Texas MD Anderson Cancer Center

Sydney Dy, MD
Johns Hopkins

Esme Finlay, MD
University of New Mexico Cancer Research and Treatment Center

Arif Kamal, MD, MHS, FAAHPM
Duke University

Kristen McNiff, MPH
Dana-Farber Cancer Institute

Michael Neuss, MD, FASCO
Vanderbilt Ingram Cancer Center

John Sprandio, MD
Consultant in Med Onc and Hem Inc

Holley Stallings, RN
Norton Cancer Institute

Jamie Von Roenn, MD, FASCO
American Society of Clinical Oncology

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2005

Ad.3 **Month and Year of most recent revision:** 11, 2015

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?** 12, 2017

Ad.7 **Disclaimers:** These clinical indicators and quality measures are not intended to and should never supplant independent physician judgment with respect to particular patients or clinical situations. Patient care is always subject to the independent physician judgment with respect to particular patients or clinical situations. Patient care is always subject to the independent professional judgment of the treating physician.

Accordingly, QOPI participants' adherence to quality measures contained in this research report is strictly voluntary and discretionary, with the ultimate determination regarding their application to be made by the treating physician in his or her professional judgment and in light of each patient's individual circumstances. ASCO does not endorse the QOPI® measures as guidelines for standards of practice or 'best practices.'

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

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 #: 0213
Measure Title: Proportion of patients who died from cancer admitted to the ICU in the last 30 days of life
Measure Steward: American Society of Clinical Oncology
Brief Description of Measure: Proportion of patients who died from cancer admitted to the ICU in the last 30 days of life
Developer Rationale: Studies suggest that over time, cancer care is becoming more aggressive near the end of life. Intensive care unit (ICU) admissions in the last 30 days of life are deemed as "aggressive care" and often used as an indicator of lower quality of care (Barbera, 2015). A higher quality of life has been predicted in patients who avoid aggressive measures such as ICU stays in the last week of life (Zhang, 2012). Furthermore, a longitudinal population-based study found patients who enrolled in hospice (long-or short-term) vs. those who did not receive hospice services had a reduced likelihood of being admitted to an ICU in the last 30 days of life by approximately 75% (Kao, 2015). ICU admissions, particularly those that result in a patient dying in the ICU, are more likely to result in physical and emotional distress as well as a less positive death experience (Wright, 2010).

Despite limited evidence of improved patient outcomes, nearly 25% of Medicare expenditures are spent of intensive care in the final month of life (Wright, 2010). A reduction in health care expenditures can be achieved by reduced utilization of hospital services including ICU stays and a greater focus on palliative care and hospice services (Langton, 2014).

Zhang B, Nilsson ME, Prigerson HG. Factors important to patients´quality of life at the end of life. Arch Intern Med 2012;172:1133-1142.Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3806298/

Wright AA, Keating NL, Balboni TA, et al. Place of death: correlations with quality of life of patients with cancer and predictors of bereaved caregivers' mental health. J Clin Oncol 2010; 28:4457–4464. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2988637/

Langton JM, Blanch B, Drew AK, et al. Retrospective studies of end of-life resource utilization and costs in cancer care using health administrative data: a systematic review. Palliat Med 2014;28:1167-1196. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24866758.

Kao YH, Chiang JK. Effect of hospice care on quality indicators of end-of-life care among patients with liver cancer: a national longitudinal population based study in Taiwan 2000-2011. BMC Palliat Care 2015: 14:39. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4545784/#CR5

Barbera L, Seow H, et al. Quality of end-of-life cancer care in Canada: a retrospective four-province study using administrative health care data. Curr Oncol 2015 Oct: 22(5): 341-355. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4608400/

Numerator Statement: Patients who died from cancer and were admitted to the ICU in the last 30 days of life
Denominator Statement: Patients who died from cancer
Denominator Exclusions: None

Measure Type: Intermediate Clinical Outcome
Data Source: Administrative claims, Electronic Clinical Data : Registry
Level of Analysis: Clinician : Group/Practice

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

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**

Summary of prior review in 2012:

- The developer cited one retrospective cohort study that examined trends in the aggressiveness of end-of-life (EOL) cancer care in Ontario, Canada, between 1993 and 2004. Aggressiveness of EOL care was defined as the occurrence of: chemotherapy received within 14 days of death, more than one emergency department (ED) visit within 30 days of death, more than one hospitalization within 30 days of death, or at least one intensive care unit (ICU) admission within 30 days of death. The developer also referenced an underlying expert consensus statement from 2003 that identified potential indicators of quality of end-of-life cancer care using administrative data.
- In the previous evaluation of the measure, the Committee noted that in some cases use of the ICU near death is appropriate, but agreed that patients overwhelmingly would prefer to not die in the ICU and noted that an ICU stay is distressing for the patient and the patient's family. Members agreed that this measure is useful for detecting patterns in practice and variation in performance, identifying outliers when comparing similar practices with similar patient populations, addressing patient preference and overtreatment at the end of life, and reflecting disparities in access to care and the capacity of the local healthcare system to treat patients appropriately at the end of life.

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 provides a diagram of the relationship of this intermediate outcome (admission to ICU in the last 30 days of life) to better patient outcomes (positive death experience, improved patient and caregiver/family satisfaction, and reduced resource utilization and costs).
- The developers reference the following as evidence for these relationships.
  - A 2013 Cochrane Collaborative systematic review evaluated the impact of home-based palliative care services on several patient and caregiver outcomes. This review found that for patients with cancer, home-based palliative care services increases the chance of dying at home and reduces symptom burden.
  - A 2012 provisional clinical opinion from the American Society of Clinical Oncology addressed integration of palliative care services into standard oncology care. This panel concluded that while there is need for additional research, several of the available studies show evidence of benefit, and recommended that

palliative care should be considered early in the course of illness for any patient with metastatic cancer and/or with high symptom burden.

- Individual articles by Barbera (2015), Zhang (2012), Kao (2015), Wright (2010), and Langton (2014). Of these, the articles by Zhang and Wright support the relationship of reduced ICU visits to desired patient outcomes.

Exception to evidence: N/A

Guidance from the Evidence Algorithm
Intermediate clinical outcome measure, but no systematic review of evidence relevant to the process being measured (Box 3) →Empirical evidence without systematic review (Zhang, Wright) (Box 7) →Unknown whether articles cited includes all studies in the body of evidence (Box 8) →Moderate or Low, depending on whether the body of evidence is represented

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?*
- *Does the Committee know of additional empirical evidence supporting this process to desired patient outcomes?*

Preliminary rating for evidence: ☐ **High** ☐ **Moderate** ☒ **Low** ☐ **Insufficient**
Rationale: If the articles cited include all studies in the body of evidence, the measure is eligible for a MODERATE rating if the Committee agrees there is a high certainty that benefits clearly outweigh undesirable effects.

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.

- Data from 2 integrated health systems (IHS) was provided. These appear to be system-level results rather than clinician group/practice performance results, although the measure is specified for clinician groups/practices. ***The performance results from the two sources are substantially different***.
  - The first IHS used the sampling methodology from ASCO's Quality Oncology Practice Initiative (QOPI®) Registry (a minimum of 40 cases twice each year). No data on the number of patients represented in these results were provided.

	Fall 2011	Spring 2012	Fall 2012	Spring 2013
Overall Performance %	20.00	25.00	40.00	37.00

  - The second IHS used data derived from a combination of electronic clinical data and tumor registry data for patients who died of cancer between June 2013 and May 2015.

Numerator	1,342
Denominator	14,988
% of Total	8.95%
Mean (2 year)	9.02%
Standard Deviation (2 year)	1.22%
Minimum (2 year)	6.86%
Maximum (2 year)	11.27

Disparities

- The developer provides data from 2 integrated health systems.

Questions for the Committee:

- *Do these results demonstrate a gap in care for clinician group/practices that warrants a national performance*

measure?

- *The results from the two integrated systems are very different. Does the variation—and particularly the very high rates for system 1—give any cause for concern?*
- *Does the Committee have additional evidence that there are practice-level disparities in ICU use in the last 30 days of life for cancer patients?*

Preliminary rating for opportunity for improvement: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**
Rationale: Clinician group/practice-level results are required; it is unclear whether the data provided meet this requirement.

Committee pre-evaluation comments

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

1a.
* Updated evidence was provided. A diagram of the relationship to better patient outcomes was included. The developers referenced 5 sources for evidence but no systematic review of the evidence relevant to the process being measured.
* There is no systematic review of evidence specifically related to the process being measured and patient outcomes. Evidence does suggest higher levels of PTSD in caregivers of patients who died in the ICU. Tangentially, patient/caregiver outcomes more positive if patient dies at home.
* Intermediate outcome. Supported by expert consensus and recent studies.
* Intermediate clinical outcome:

The updated information is helpful.
A positive death experience, improved satisfaction, and reduced resource utilization are reasonably well linked in individual studies to the reduced ICU visit outcome. It is remarkable of the 5 individual articles cited the relationship of reduced ICU visits to desired patient outcomes is linked to only Zhang and Wright. Hwang in Oct 2015 in Palliative and Supportive Care notes terminally ill cancer patients preferences re: ICU care are not stable over time.

The cochrane 2013 review, while not directly measuring the outcome of ICU admission, is such that it is fair to state that remaining in the home does avoid an ICU admission if hospital stay is avoided.

1b.
* The performance results from the 2 sources are substantially different. It appears that the data are system level not practice level and no data on the number of patients is provided.
* Variability in performance gap between 8.95% and 40%. Population subgroup data not provided. It does not appear that practice/provider data was described either. Other study showing variability in ICU care that includes cancer patients: Hart JL JAMA Intern Med. 2015 Jun 1; 175(6): 1019–1026.
* There is a large performance gap with extensive variation in rates of ICU deaths across the country. Additional data are available in the Dartmouth atlas that show this performance gap.
* I reviewed the 2 integrated health systems programs data. The 2 systems have different results. As for the variation, it gives one pause. I question if regional variation plays a role and wonder if the numbers were adjusted for the number of ICU beds in the region if that would make a difference.

a key question is the impact of practice-level disparity-
I am unaware of evidence directly relating to practice-level disparity in ICU use, but the Dartmouth atlas shows substantial state to state variability. http://www.dartmouthatlas.org/data/table.aspx?ind=14 This data from 2012, but has been recently updated I believe.

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 sources for this measure include administrative claims and registry. Unlike several other ASCO measures (e.g., #0210, #0215, #0216), this measure is not currently in use in ASCO's Quality Oncology Practice Initiative (QOPI®) Registry.

Specifications:

- The measure is specified at clinician group/practice level of analysis.
- A lower score on the measure is indicative of better quality.
- Care settings for this measure include clinician offices or clinics, hospices, and hospitals/acute care facilities.
- The numerator includes those admitted to the ICU in the last 30 days of life.
- The denominator includes patients who died from cancer. Presumably these patients can be identified in a registry. It is unclear if/how the denominator would be identified through claims data. The developers briefly mention use of the "death registry" but do not explain how this is used in conjunction with claims data.
- There are no exclusions for this measure.
- A calculation algorithm is provided.
- This measure is not risk-adjusted.

Questions for the Committee :

- *Is use of death registry data required if the measure is calculated from claims data? If so, what is the timeliness of this dataset? If the death registry is not used, can the measure denominator be identified reliably through administrative data?*
- *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:

- For claims data, data element validity testing was conducted by comparing claims data to the full medical record. NQF guidance indicates that if data element validity is demonstrated, additional data element reliability testing is not required.

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

- For claims data, see method(s) of validity testing, below.
- For registry data, no testing results were provided.

Results of reliability testing

- For claims data, see results of validity testing, below.

Guidance from the Reliability Algorithm

For registry data: Precise specifications (Box 1) → no empirical reliability testing conducted (Box 2) → no empirical validity testing conducted (Box 3) → Insufficient

For claims data: Fairly precise specifications, except for denominator (Box 1) → no empirical reliability testing conducted for claims data (Box 2) → empirical validity testing conducted (Box 3) → FROM VALDITY ALGORITHM: data element testing for numerator (Box 10) →Accuracy of the denominator data element not demonstrated for claims data, but assume okay if data derived from national death registry (Box 11) → Moderate

Questions for the Committee:

- *Can claims adequately identify cancer deaths? If not, is use of death registry data required?*
- *Is the test sample adequate to generalize for widespread implementation?*

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

Rationale: For the claims data, the measure is eligible for a MODERATE rating. However, testing with registry data is required if the developer wants the measure endorsed for the registry data source. One option would be for the developer to remove the registry option from the measure specifications.

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?*

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:

- For claims, data element validity testing was done by comparing claims data to the full medical record. This is an appropriate method of data element validation.

Describe any updates to validity testing:

- No updated testing was provided.

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:

- For administrative claims data:
  - Claims data for 150 consecutive patients treated for advanced cancer at Boston's Dana-Farber Cancer Institute and Brigham and Women´s Hospital were compared to data from the full medical record. Numerator data elements were compared ***but not the denominator data element***. Dates for the data examined in testing were not provided.
- Face validity assessment:
  - The developer states that face validity was assessed by conducting focus group and structured interviews with end-of-life cancer patients and bereaved caregivers, and then vetted by an expert panel

of cancer providers. According to NQF guidance, the face validity of the measure score as a quality indicator may be adequate if it explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. There is not enough information provided to know if the face validity assessment conforms to NQF's requirements.

Validity testing results:

- For administrative claims data:
  - Sensitivity=0.87; Specificity=0.97
  - Sensitivity measures the proportion of actual positives that are correctly identified as such. A sensitivity value of 0.87 reflects the accuracy of identifying an ICU stay within 30 days of death in the claims data when it present in the medical record data (the authoritative source).
  - Specificity measures the proportion of actual negatives that are correctly identified as such. A specificity value of 0.97 reflects the accuracy of the absence of an ICU stay within 30 days of death in the claims data when this is not recorded in the medical record data (the authoritative source).
- Face validity assessment:
  - The developer does not provide results from the face validity assessment.

Questions for the Committee:

- *The developer has not provided any new validity testing, even though there is no testing for the denominator element (cancer deaths) for claims data and no testing using registry data. Does the developer expect the denominator to be identified via a death registry? If so, should these data be assumed valid? If not, is there evidence that claims adequately identify cancer deaths?*
- *For claims data, is the test sample adequate to generalize for widespread implementation?*
- *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:

- There are no exclusions for this measure.

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

Conceptual rationale for SDS factors included ? ☐ **Yes** ☒ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary

- In the ***previous evaluation*** of the measure,
  - The developer stated that "*No risk adjustment or risk stratification is necessary because the measure is intended to be used for comparison among similar providers; unless there is a reason to believe that one providers' patients have significantly different risks than others, it will not affect relative comparisons. Since, however, comorbidity risks could increase the likelihood of experiencing this process of care, stratification or adjustment as described above can be considered.."*
  - The Committee agreed that this measure is useful for detecting variations in performance and identifying outliers when comparing similar practices with similar patient populations.
- The developer is now considering whether risk-adjustment for this measure is appropriate.
- No empirical analysis is provided to support lack of adjustment and nothing in the specifications requires comparison of results across similar providers.

Questions for the Committee:

- *Do you believe risk-adjustment is needed for this measure?*

○ *Do you have thoughts about which (if any) SDS factors should be considered if the measure is risk-adjusted?*

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

- When initially developed (using 1991-1996 SEER-Medicare linked data), a benchmark of <4% was achieved by the highest performing SEER regions.
- Variation within 2 integrated health systems is shown via descriptive statistics about performance results (see section 1b, above). It is not clear, however, whether either set of results reflect clinician group/practice performance.

Question for the Committee:

○ *Does this measure identify meaningful differences about quality for clinician groups/practices?*

2b6. Comparability of data sources/methods:

- The developer has not provided analysis to assure comparability of results between claims versus registry (the analysis of claims data across time presented in the earlier evaluation does not speak to this topic).

2b7. Missing Data

- The developer describes how missing data are handled in section S.22.

Guidance from the Validity Algorithm:
For claims data: Specifications consistent with evidence (Box 1) → potential threats to not completely assessed (Box2) →Insufficient OR, if no concerns about risk-adjustment/comparability, → empirical testing conducted (Box 3) → no score-level testing (Box 6) → data element testing conducted (Box 10) → Accuracy of the denominator data element not demonstrated for claims data, but assume okay if data derived from national death registry (Box 11) → Moderate

For registry data: Specifications consistent with evidence (Box 1) → potential threats to not completely assessed (Box2) →Insufficient OR, if no concerns about risk-adjustment/comparability, → no empirical testing conducted (Box 3) → face validity conducted, but not enough information to know if it conforms to NQF requirements (Box 4) → Insufficient

Preliminary rating for validity: ☐ **High** ☐ **Moderate** ☐ **Low** ☒ **Insufficient**
Rationale: Additional analysis to assess threats to validity needed (risk-adjustment/comparability). For claims data, testing for the denominator data element for claims was not conducted, but may not be needed if data required to be obtained from a death registry. For registry data, no testing data presented; unclear if face validity testing meets NQF requirements.

Committee pre-evaluation comments

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

2a, 2b.
Reliability – Specifications
* The calculation algorithm is clear. It is unclear if the denominator can be identified clearly through claims data.
* For claims data, specifications are reasonable. Claims data testing with 150 consecutive patients treated for advanced cancer.
* Hospital deaths are coded in Medicare claims and as such, unlike measure 0211, the denominator does not need an additional death registry. Specification of the numerator is adequate.
* ICU admission is fairly clear. Key issue of concern with the inclusion: how will the cancer patients be determined? Referred to in the NQF reviewer comments as the challenge of the denominator If admitted for a CHF exacerbation while undergoing cancer treatment, will the ICU admission be counted as a cancer patient? Will there be extensive variation in how this is done from place to place?

There is no risk adjustment.

Validity – Specifications
* Specifications are consistent with the evidence.
* Claims based sensitivity-specificity:0.87 and 0.97. Face validity via focus group.
* Specs are consistent with the measure.
* What do we make of the patient who is admitted to the ICU with a large pleural effusion and no prior cancer diagnosis who dies in the next 20 days?
Presumably the ICU admission pre-diagnosis of cancer was quality care as opposed to a system break-down for end of life care.

While it is reasonable to assume that there would be some similarities from hospital to hospital with a similar baseline rate overall, would tertiary hospitals appear to have much worse outcomes when undiagnosed cancer patients with presumed cancer are referred for a diagnostic workup at the end of life?

Reliability – Testing
* No updated testing was provided. The first method does not indicate the number of patients. The registry data is a problem.
* No reliability testing at registry level.
* Initial reliability of the measure was demonstrated.
* On whole, the argument for claims data Moderate rating vs registry data insufficiency sees to sum it up well.

Validity Testing
* No new validity data was provided.There is not enough information provided on the face validity data.
* No validity testing for denominator element. In general, claims//emr review should adequately capture in-hospital deaths, but not necessarily all deaths.
Also no results provided for face validity assessment though qualitative work was conducted.
* Acceptable
* While I can in general accept the face validity argument fairly well and accept, the quasi-experimental assessment method, the comments of Katherine Ast MSW, LCSW raise a key question: if we are interested in measuring quality of life of all deaths, why limit the ICU in last 30 days admission to cancer? On the other hand, if one were not anticipated to die, is ICU death not a marker of reasonable health care delivery. The cancer limitation does imply a disease trajectory for which the measure as a marker of quality has greater validity that across all disease types with more widely varying prognostic trajectories.

Threats to Validity
* The developer has not provided analysis to assure that the claims data and registry data are comparable. Dates were not provided so the age of the data is unknown.
* Unclear based on data provided.
* no
* see above discussions re: cancer diagnosis inclusion

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.

- All data elements are in defined fields in electronic clinical data. Most data elements are routinely generated during care delivery.
- Cancer decedents should be easily identifiable in the registry, but may not be easily or accurately identifiable through claims. Use of a death registry may be needed if measure calculated via claims.
- No feasibility concerns were raised by the Cancer Steering Committee during the NQF Maintenance of Endorsement review in 2012.

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

○ *Is use of a death registry required? If not, how will the measure denominator be identified using claims data?*

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

Committee pre-evaluation comments

Criteria 3: Feasibility

* The data elements are routinely generated. Cancer decedents can be easily identified in the registry data but not in the claims data. The use of a death registry may be necessary.
* Only problem is reporting of death if using claims data.
* Data are routinely collected in claims and available
* see above discussions re: cancer diagnosis inclusion

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:

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 measure is part of America's Health Insurance Plans (AHIP) Medical Oncology Core Measure Set. The AHIP effort is a collaboration of both public and private stakeholders to identify measures that are meaningful to patients, consumers, and physicians and to reduce variability in measure selection, collection burden, and cost. Payers involved in the collaboration have committed to using for reporting as soon as feasible. By virtue of being included in the AHIP measure set, CMS will consider this measure for inclusion in Medicare quality programs.

Improvement results: No performance trend can be inferred from the data presented: four semesters from Spring 2011-Fall 2013 in one integrated health system.

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

Potential harms: No potential harms have been signaled by the developer.

Feedback: From previous evaluation:

- Steering Committee members raised concerns about use of the measure given the [then] current issues with access to quality hospice facilities. The Committee believed patients may utilize ICU services at the end of life when good hospice care is not available. Members agreed that in areas where performance of the measure is

poor, it will call attention to a lack of resources available for patients at the end of life.

***Questions for the Committee**:*

- *NQF guidelines strongly recommended measures be publicly reported following six years of endorsement. Does the measure have a viable path to public reporting and eventual accountability applications?*
- *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**

Committee pre-evaluation comments

Criteria 4: Usability and Use

* The measure is not being publicly reported but it is part of AHIP and will be included in the Medicare quality programs. No unexpected findings or potential harms were reported.
* Included in the AHIP measure set
* Not currently being publicly reported. No unintended consequences of note.
* This measure is used as part of America's Health Insurance Plans (AHIP) Medical Oncology Core Measure Set.No performance result trend was noted from 2011-2013 data.

The key benefit of this measure would be establishing a baseline rate of ICU use across institutions for cancer and to spot areas of potential vast over-use of ICUs. WHile past NQF comittee members believed patients might utilize ICU services at end of life when good hospice care is not available, it is fair now to thing of both outpatient and home palliative care delivery as well as hospice as a factor impacting this metric.

Criterion 5: Related and Competing Measures

Related measures:

0210: Proportion of patients who died from cancer receiving chemotherapy in the last 14 days of life
0211: Proportion of patients who died from cancer with more than one emergency department visit in the last 30 days of life
0215: Proportion of patients who died from cancer not admitted to hospice
0216: Proportion of patients who died from cancer admitted to hospice for less than 3 days

Harmonization: N/A

Pre-meeting public and member comments

Name: Michele Galioto, RN, MSN

Organization: Oncology Nursing Society

Comment: Should the measure include immunotherapy? Does “chemotherapy” include hormonal and biotherapy? It may be more inclusive to refer to all as "antineoplastic therapy."

Should the measure include radiation therapy as well?

Does “death from cancer” include all death within 14 days? Death may be the result of infection, accident (e.g., fall), bleeding, etc. which could be tied to cancer or cancer treatment. Death attributed to side effects of therapy may be indistinguishable from cancer deaths. Is the intent that death occurs within a timeframe of receiving chemotherapy?

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 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*)**:** 0213

Measure Title: Proportion of patients who died from cancer admitted to the ICU in the last 30 days of life

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): Click here to name the PRO

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

☒ Intermediate clinical outcome (*e.g., lab value*): Rate of ICU admissions in the last 30 days of life

☐ 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*

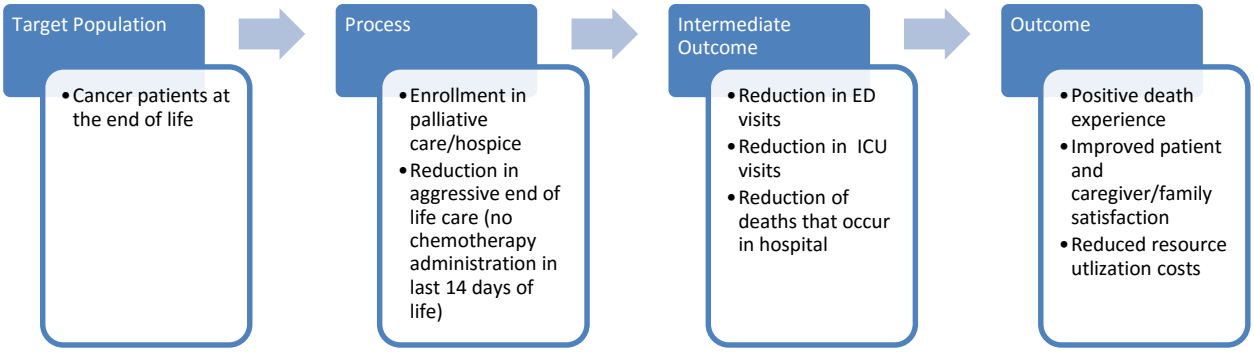
1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

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*).

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.



Studies suggest that over time, cancer care is becoming more aggressive near the end of life. Intensive care unit (ICU) admissions in the last 30 days of life are deemed as "aggressive care" and often used as an indicator of lower quality of care (Barbera, 2015). A higher quality of life has been predicted in patients who avoid aggressive measures such as ICU stays in the last week of life (Zhang, 2012). Furthermore, a longitudinal population-based study found patients who enrolled in hospice (long-or short-term) vs. those who did not receive hospice services had a reduced likelihood of being admitted to an ICU in the last 30 days of life by approximately 75% (Kao, 2015). ICU admissions, particularly those that result in a patient dying in the ICU, are more likely to result in physical and emotional distress as well as a less positive death experience (Wright, 2010).

Despite limited evidence of improved patient outcomes, nearly 25% of Medicare expenditures are spent on intensive care in the final month of life (Wright, 2010). A reduction in health care expenditures can be achieved by reduced utilization of hospital services including ICU stays and a greater focus on palliative care and hospice services (Langton, 2014).

Zhang B, Nilsson ME, Prigerson HG. Factors important to patients'quality of life at the end of life. ArchIntern Med 2012;172:1133-1142.Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3806298/

Wright AA, Keating NL, Balboni TA, et al. Place of death: correlations with quality of life of patients with cancer and predictors of bereaved caregivers' mental health. *J Clin Oncol* 2010; 28:4457–4464. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2988637/

Langton JM, Blanch B, Drew AK, et al. Retrospective studies of end of-life resource utilization and costs in cancer care using health administrative data: a systematic review. Palliat Med 2014;28:1167-1196. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24866758.

Kao YH, Chiang JK. Effect of hospice care on quality indicators of end-of-life care among patients with liver cancer: a national longitudinal population based study in Taiwan 2000-2011. BMC Palliat Care 2015: 14:39. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4545784/#CR5

Barbera L, Seow H, et al. Quality of end-of-life cancer care in Canada: a retrospective four-province study using administrative health care data. Curr Oncol 2015 Oct: 22(5): 341-355. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4608400/

2012 Submission: A structural feature: regional availability of hospice, has been shown to correlate with a composite measure of the aggressiveness of cancer care near the end of life that contains this measure. Mostly it is a process measure indicating a possible inadequate focus on palliation and supportive care, that can affect quality of life.

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***

2012 Submission: Selected individual studies (rather than entire body of evidence)

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*):

Smith TJ, Temin S, Alesi ER, et al. American Society of Clinical Oncology Provisional Clinical Opinion: The Integration of Palliative Care into Standard Oncology Care. J Clin Oncol 2012;30:880-887. Available at: http://www.instituteforquality.org/asco-provisional-clinical-opinion-integration-palliative-care-standard-oncology-care.

Gomes, B., N. Calanzani, et al. (2013). "Effectiveness and cost-effectiveness of home palliative care services for adults with advanced illness and their caregivers." Cochrane Database Syst Rev **6**: CD007760 Available at: http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD007760.pub2/pdf.

2012 Submission: The underlying evidence was obtained by expert consensus, as described in Earle CC, Park ER, Lai B, Weeks JC, Ayanian JZ, Block S. Identifying potential indicators of the quality of end of life cancer care from administrative data. J Clin Oncol. 2003;21(6):1133-8. The panel consisted of oncologists, nurses, palliative care specialists, etc, and used a modified Delphi process to evaluate measures.

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?

A 2012 American Society of Clinical Oncology (ASCO) Provisional Clinical Opinion (PCO) addresses the integration of palliative care (PC) services into standard oncology care at the time a person is diagnosed with metastatic cancer and/or high symptom burden.

A 2013 Cochrane Review, 'Effectiveness and cost-effectiveness of home palliative care services for adults with advanced illness and their caregivers', evaluated the impact of home palliative care services on outcomes for adults with advanced illness or their family caregivers, or both. The aim of the review was to quantify the effect of home palliative care services on a patients' odds of dying at home, examine the clinical effectiveness of home palliative care services on other outcomes such as symptom control, quality of life, caregiver distress and satisfaction with care, and comparing resource use and costs associated with these services.

2012 Submission: The argument is made that because providers cannot predict the future, measures based on decedent cohorts are unfair. However, as described above in 1a.a, the idea is for the measure to be seen as an overall indication of practice style and/or available palliative resources. An individual patient experiencing this process of care has not necessarily received poor quality care. If explanations other than practice style and resource availability, such as unusually poor prognostic ability on the part of the provider or unexpected toxic deaths (whether unavoidable, from overly aggressive treatment, or poor patient selection) are enough to influence the overall aggregate rates, it is still justifiable to consider it a 'red flag' that should prompt examination of the care provided.

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

2012 ASCO PCO (p. 881):

The American Society of Clinical Oncology (ASCO) has established a rigorous, evidence-based approach—the provisional clinical opinion (PCO)—to offer a rapid response to emerging data in clinical oncology. The PCO is intended to offer timely clinical direction to ASCO's oncologists after publication or presentation of potentially practice- changing data from major studies.

The PCO may serve in some cases as interim direction to the membership pending the development or updating of an ASCO clinical practice guideline. As such, the evidence is not graded in a PCO and is a result of expert consensus. A clinical guideline on palliative care integration with recommendations and the associated grading is under development,

2013 Cochrane Review (p. 12):

Two independent reviewers assessed all included studies for methodological quality using the standard criteria developed by the Cochrane EPOC Review Group for RCTs/CCTs, CBAs and ITSs. The checklist for RCTs/CCTs contains seven items qualified as done, unclear and not done for concealment of allocation, follow-up of professionals, follow up of patients or episodes of care, blinded assessment of primary outcome(s), baseline assessment, reliable primary outcome measure(s) and protection against contamination. Blinding and reliability of all outcomes were also assessed.

Each criterion was scored zero (not done), 0.5 (not clear or when scores varied across outcomes) and one (done). Total scores for RCTs/ CCTs ranged from zero to six; studies with a score of 3.5 or above were considered of high quality. Integration of the results of the quality assessment in data analysis was done in addition to meta-analyses with sensitivity analyses including only high quality studies.

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

See 1a.7.2 for this information.

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

2012 ASCO PCO: 2004-2012

2013 Cochrane Review: 1950 – November 2012

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*)

2012 ASCO PCO: 7 randomized controlled trials

2013 Cochrane Review: 5 randomized controlled trials and 2 controlled clinical trials

2012 Submission: 4

1a.7.6. What is the overall quality of evidence <u>across studies</u> 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*)

2012 ASCO PCO:

This PCO did not provide an assessment of the overall quality of evidence across the studies. This analysis will be completed during the development of the upcoming clinical guideline.

2013 Cochrane Review:

p. 3: The direction of the effect was consistent across all studies but did not reach statistical significance in 3; ORs ranged from 1. 36 (95% CI 0.80 to 2.31) to 2. 86 (95% CI 0.78 to 10.53) Sensitivity analyses showed that exclusion of the 2 CCTs (both of Swedish hospital-based services with a pooled OR 3.44, 95% CI 0.60 to 19.57) and inclusion of only high quality RCTs resulted in a reduction of the OR to 1.28 (95% CI 1.28 to 2.33) and 1.75 (95% CI 1.24 to 2. 47) respectively, with more precision and less heterogeneity.

p. 22:

Pooled data from seven studies (five RCTs, three of high quality, and two CCTs with 1222 participants) showed that those receiving home palliative care had statistically significantly higher odds of dying at home than those receiving usual care (OR 2.21, 95% CI 1.31 to 3.71; Z = 2.98, P value = 0.003; Chi^2 = 20.57, degrees of freedom (df) = 6, P value = 0.002; I^2 = 71%). The study population control risk was of 307 home deaths per 1000 deaths; based on this ACR of 0.307, the NNTB was 5 (95% CI 3 to 14), meaning that for one additional patient to die at home five more would need to receive home palliative care as opposed to usual care. Assuming a medium cancer home death rate population ACR of 0.278 (i.e. 278 home deaths per 1000 cancer deaths), the NNTB was 6 (95% CI 3 to 15). This means that for one additional cancer patient to die at home in a population where there are 278 home deaths per 1000 cancer deaths, six more would need to receive home palliative care. NNTB estimates ranged from 9 patients (95% CI 5 to 16) when applied to a low home death rate population such as the one observed in Norway (128 home deaths per 1000 cancer deaths) to 5 patients (95% CI 3 to 13) when applied to a high home death rate population such as the one observed in the Netherlands (454 home deaths per 1000 cancer deaths).

2012 Submission: The studies are qualitative and observational using administrative data, consequently there are limitations to the quality of the data.

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) <u>across studies</u> 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*)

2012 ASCO PCO (p. 884):

Seven published randomized trials demonstrate the feasibility of providing various components of PC alongside

usual oncology care. There is, however, a dearth of data evaluating the integration of modern PC practices into standard oncology care, especially in concert with ongoing antitumor therapy. Overall, the addition of PC interventions to standard oncology care delivered via different models to patients with cancer provided evidence of benefit.

2013 Cochrane Review (p. 22):

The study population control risk was of 307 home deaths per 1000 deaths; based on this ACR of 0.307, the NNTB was 5 (95% CI 3 to 14), meaning that for one additional patient to die at home five more would need to receive home palliative care as opposed to usual care. Assuming a medium cancer home death rate population ACR of 0.278 (i.e. 278 home deaths per 1000 cancer deaths), the NNTB was 6 (95% CI 3 to 15). This means that for one additional cancer patient to die at home in a population where there are 278 home deaths per 1000 cancer deaths, six more would need to receive home palliative care. NNTB estimates ranged from 9 patients (95% CI 5 to 16) when applied to a low home death rate population such as the one observed in Norway (128 home deaths per 1000 cancer deaths) to 5 patients (95% CI 3 to 13) when applied to a high home death rate population such as the one observed in the Netherlands (454 home deaths per 1000 cancer deaths).

2012 Submission: All studies have shown similar results. As per Ho TH, Barbera L, Saskin R, Lu H, Neville BA, Earle CC.Trends in the Aggressiveness of End-of-Life Cancer Care in the Universal Health Care System of Ontario, Canada. J Clin Oncol April 20, 2011 vol. 29 no. 12 1587-1591, although rates in Canada were lower, trends were similar over time in a comparison with U.S. Medicare patients.

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

2012 ASCO PCO (p. 884-885):

No harm to any patient was observed in any trial, even with discussions of EOL planning, such as hospice and ADs. Two of five trials measuring change in symptoms, two of five studies measuring QOL, two of three studies measuring patient/caregiver satisfaction, and one of three studies measuring survival found statistically significant improvements with PC. Three of six studies measuring mood, two of five studies measuring resource use, and one of four studies measuring outcomes of advance care planning found statistically significant differences, and one outcome of borderline significance was also found in each of these three areas, Therefore, most trials showed benefits ranging from equal to improved overall survival, reduced depression, improved caregiver and/or patient QOL, and overall lower resource use and cost because EOL hospitalizations were avoided.

2013 Cochrane Review: Discussion of harms was not addressed.

2012 Submission: Less use of ICU near death can result in better quality of life (death) as well as resource savings.

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.

No relevant studies have been conducted and published since the systematic reviews.

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.

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 0213_Evidence_Form_3.16.16.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)* Studies suggest that over time, cancer care is becoming more aggressive near the end of life. Intensive care unit (ICU) admissions in the last 30 days of life are deemed as "aggressive care" and often used as an indicator of lower quality of care (Barbera, 2015). A higher quality of life has been predicted in patients who avoid aggressive measures such as ICU stays in the last week of life (Zhang, 2012). Furthermore, a longitudinal population-based study found patients who enrolled in hospice (long-or short-term) vs. those who did not receive hospice services had a reduced likelihood of being admitted to an ICU in the last 30 days of life by approximately 75% (Kao, 2015). ICU admissions, particularly those that result in a patient dying in the ICU, are more likely to result in physical and emotional distress as well as a less positive death experience (Wright, 2010). Despite limited evidence of improved patient outcomes, nearly 25% of Medicare expenditures are spent of intensive care in the final month of life (Wright, 2010). A reduction in health care expenditures can be achieved by reduced utilization of hospital services including ICU stays and a greater focus on palliative care and hospice services (Langton, 2014). Zhang B, Nilsson ME, Prigerson HG. Factors important to patients´quality of life at the end of life. Arch Intern Med 2012;172:1133-1142.Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3806298/ Wright AA, Keating NL, Balboni TA, et al. Place of death: correlations with quality of life of patients with cancer and predictors of bereaved caregivers' mental health. J Clin Oncol 2010; 28:4457–4464. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2988637/ Langton JM, Blanch B, Drew AK, et al. Retrospective studies of end of-life resource utilization and costs in cancer care using health administrative data: a systematic review. Palliat Med 2014;28:1167-1196. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24866758. Kao YH, Chiang JK. Effect of hospice care on quality indicators of end-of-life care among patients with liver cancer: a national longitudinal population based study in Taiwan 2000-2011. BMC Palliat Care 2015: 14:39. Available at:

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4545784/#CR5

Barbera L, Seow H, et al. Quality of end-of-life cancer care in Canada: a retrospective four-province study using administrative health care data. Curr Oncol 2015 Oct: 22(5): 341-355. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4608400/

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.*
Performance data on this measure was obtained from two independent integrated healthcare delivery systems in the United States. Both systems are located in the South.

The first integrated healthcare delivery system manually abstracted data from their EMR using the sampling methodology from ASCO's Quality Oncology Practice Initiative (QOPI®) Registry (a minimum of 40 cases twice each year):

Integrated Healthcare delivery System #1
ICU visits in the last 30 days of life

	Fall 2011	Spring 2012	Fall 2012	Spring 2013
Overall Performance %	20.00	25.00	40.00	37.00

The second integrated healthcare delivery system´s scores were derived from death notification data in a tumor registry combined with electronic clinical data for patients who died from cancer between June 2013 and May 2015:

Based on reported deaths 6/1/2013-5/31/2015 (2 years rolling)

Integrated Healthcare Delivery System #2
Admission to the ICU in the last 30 days

Numerator	1342
Denominator	14988
% of Total	8.95%
Mean (2 year)	9.02%
Standard Deviation (2 year)	1.22%
Minimum (2 year)	6.86%
Maximum (2 year)	11.27%

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.*
Performance data on this measure was obtained from two integrated healthcare delivery systems in the United States. Both systems are located in the South.

The first integrated healthcare delivery system manually abstracted data from their EMR using the sampling methodology from ASCO's Quality Oncology Practice Initiative (QOPI®) Registry (a minimum of 40 cases twice each year):

ICU visits in the last 30 days of life

	Fall 2011	Spring 2012	Fall 2012	Spring 2013

Female	5.88	16.67	31.25	41.67
Male	30.43	31.82	45.83	29.41
Hispanic	23.08	25.00	29.41	40.91
White	18.75	31.25	45.45	21.43
Black	0	0	66.67	60.00
Other	33.33	0	0	0

These scores were derived from death notification data in a tumor registry combined with electronic clinical data for patients who died from cancer between June 2013 and May 2015:

Based on reported deaths 6/1/2013-5/31/2015 (2 years rolling)
Admission to the ICU in the last 30 days

N Female	6858
N Male	8130
% of Total Female	8.30%
% of Total Male	9.51%
Medicare numerator	478
Medicare denominator	6249
% Medicare	7.65%
Non-Medicare numerator	864
Non-Medicare denominator	8739
% Non-Medicare	9.89%

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
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.**

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

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

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.)*
No webpage available

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 have been made since last endorsement

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 who died from cancer and were admitted to the ICU in the last 30 days of life

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.)*
30 days before death

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.
MEDPAR only:
did not include SNF claims
did not include pediatric, psychiatric, burn or trauma ICUs (MEDPAR variable increind ne 3,4,7,8)
- variable in MEDPAR called incrdays, which is number of ICU days per visit
- used hospital admission date variable (admitdate) and then checked if incrdays was >0 for admissions occurring in the last 30 days before death

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Patients who died from cancer

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
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)*
Claims:Patients in the death registry with cancer as their cause of death. In the cited analyses by the measure submitter, this is a field in the cancer registry or denominator file not requiring specific codes. This may be different in other administrative data sets.

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)*
Not applicable

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)*
Not applicable

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)*
Not applicable

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*)
Not applicable

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 = Lower 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.)*
Performance is calculated as:

1. Identify those patients that meet the denominator criteria defined in the measure.

2. Subtract those patients with a denominator exclusion from the denominator if applicable. Note: this measure does not have exclusions.

3. From the patients who qualify for the denominator (after any exclusions are removed), identify those who meet the numerator criteria.

4. Calculation: Numerator/Denominator-Denominator Exclusions

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.
Not applicable

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.
Not applicable

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
This measure is specified with defined criteria and data elements. If a patient record does not include one or more of these components for the initial patient population or denominator, then patients are not considered eligible for the measure and not included.

If data to determine whether a patient should be considered for the numerator or exclusions is missing, then the numerator or exclusions not considered to be met and the practice will not get credit for meeting performance for that patient.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Electronic Clinical Data : Registry

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.
Not applicable

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)*
Clinician : Group/Practice

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic, Hospice, 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.)*
Not applicable

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
0213_MeasureTesting_MSF5.0_Data_Update.doc

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.**

Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), 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)*

ALL data elements are in defined fields in electronic claims

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.**

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.**

No feasibility assessment **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.

The measure and its specifications have been in place for several years and ASCO continues to monitor and ensure that the measure and its specifications are up to date for widespread use.

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)*.

Not applicable

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 0213	**NQF Project:** Cancer Project

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)***:**

The measure was developed using the Medicare claims of all continuously-enrolled patients who died of cancer after having been diagnosed in a SEER region between 1991 and 1996.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)***:**

Evaluation was carried out on 150 consecutive patients treated for advanced cancer at Dana-Farber Cancer Institute and Brigham and Women´s Hospital in Boston. The percent accuracy of death ascertainment for includion into this cohort is unknown but is likely high as the cancer registry regularly uses the death index for ascertainment. Ascertainment would be expected to be highly specific. Hospital billing claims were obtained and analyzed and the accuracy was compared to detailed medical record review.

2a2.3 Testing Results *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**

Sensitivity 0.87, Specificity 0.97, where sensitivity = # true positives (both claims and charts)/(# true positives + # false negatives, i.e., not in claims but present in charts) and specificity = # true negatives/(# true negatives + false positives, i.e., present in claims but not in charts).

2b. VALIDITY. Validity, Testing, including all Threats to Validity: 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:**

They are identical

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)***:**

Evaluation was carried out on 150 consecutive patients treated for advanced cancer at Dana-Farber Cancer Institute and Brigham and Women´s Hospital in Boston. Claims were obtained and analyzed and the accuracy was compared to detailed medical record review.

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

Face validity was determined by focus groups and structured interviews with end-of-life cancer patients and bereaved caregivers, and then vetted by an expert panel of cancer providers. The idea that cancer patients with terminal illness do not benefit from heroic attempts at life prolongation, and that such attempts often reflect a lack of honest discussion with patients, is summed up by the statement of one of the oncologists on the expert panel who said 'for most of our patients, a trip to the ICU is a kind of failure'. The percent agreement between claims and medical record review was calculated.

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)***:**

The measure was 95% accurate (percent true positives + true negatives).

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)***:**

N/A

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

N/A

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)***:**

N/A

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

Traditionally, ASCO believed that risk adjustment or risk stratification was not critical for these palliative care measures. The measures are used for comparison among similar providers and there is no expectation that performance will be 0%.

ASCO is exploring whether risk adjustment is appropriate and if so how it could be accomplished for our outcome measures. ASCO's palliative care measures would be included in that assessment. While several studies indicate potential variables for consideration, additional time is needed to properly define a statistical risk model method and variables, including sociodemographic (SDS) factors now that they may be considered.

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)***:**

N/A

2b4.4 **If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment:** No risk adjustment or risk stratification is necessary because the measure is intended to be used for comparison among similar providers; unless there is a reason to believe that one providers' patients have significantly different risks than others, it will not affect relative comparisons. Since, however, comorbidity risks coule increase the likelihood of experiencing this process of care, stratification or adjustment as described above can be considered.

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)***:**

We used the Medicare claims of all 28,777 continuously-enrolled patients who died of cancer after having been diagnosed in a SEER region between 1991 and 1996. This was an analysis of SEER-Medicare linked data obtained from NCI (http://healthservices.cancer.gov/seermedicare/).

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

Benchmarks were established to identify the outlying 10th decile of practice: The proportion of patients experiencing each process of care in each Health Care Service Area (HCSA) was computed and ranked from best (least aggressive) to worst. A new cohort was created by sequentially adding HCSAs in order starting with the least aggressive until they contained at least 10% of the original cohort and the proportion experiencing each process of care was then recalculated to arrive at the 'Achievable Benchmark of Care). More detail on this, as well as a reference for the Achievable Benchmark of Care method can be found in our publication: Earle CC, Neville BA, Landrum ME, Souza JE, Weeks JC, Block SD, Grunfeld E, Ayanian JZ. Evaluating claims-based indicators of the intensity of end-of-life cancer care. Int J Qual Health Care. 2005;17(6):505-9.

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)*:

A benchmark target of < 4% of patients being admitted to the ICU in the last 30 days of life corresponds to that achieved by the highest performing regions in the country.

Integrated Health System #2 (2 year range)	Mean	St. Dev.	Min	Max
	9.02	1.22	6.86	11.27

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)***:**

Administrative claims and chart review, as described above: 77 entities (HCSAs), 215,484 patients, between 1991 and 2000.

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

We have also assessed the stability of these measures over time by examining the stability of relative aggressive care over time. If the relative aggressiveness of a provider or organization´s practice appeared to change from year to year, then these measures might not be assessing a stable property of practice. To investigate this, we used hierarchical regression models to estimate regional variation in both levels and trends of each measure. We used as our geographic unit of analysis the Health Care Service Area (HCSA). HCSAs are groupings of Metropolitan Statistical Areas defined by the Centers for Medicare & Medicaid Services (CMS) based on observed patient flow patterns in Medicare for tertiary care. As such, each HCSA can be considered to be a self-contained regional health system with a related group of providers. We ranked each region according to the model-estimated rate of each indicator and computed the correlation among relative ranks of each region during the 10-year study period.

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)***:**

We observed significant variation both in levels of aggressive care and in trends in aggressiveness over time, but generally stability of regional practice patterns: Year to year correlation on this measure was 0.97, and over a 5 year span was 0.84. This provides supportive evidence of the reliability of these measures.

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:

N/A

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

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)
Payment Program	Quality Improvement (Internal to the specific organization) Not applicable Multiple Integrated Delivery Systems

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

Multiple integrated delivery systems: In use in multiple integrated delivery systems across the United States for quality improvement purposes. Because it is internal to the specific organization, we are unable to provide any additional information.

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?*)

We are continuously seeking opportunities to advocate for expanded use of this measure in government or other programs, including those intended for accountability or public reporting. For example, this measure was recently selected for inclusion in a Medical Oncology Core Measure Set supported by America's Health Insurance Plans and CMS. See section 4a.3. below for additional details.

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 has also been included in America´s Health Insurance Plans Medical Oncology Core Measure Set. The purpose of this program is to reduce variability in measure selection, specifications and implementation. The measures will be implemented nationally by private health plans using a phased-in approach. Contracts between physicians and private payers are individually negotiated and therefore come up for renewal at different points in time depending on the duration of the contract. It is anticipated that private payers will implement these core sets of measures as and when contracts come up for renewal or if existing contracts allow modification of the performance measure set. CMS is also working to align measures across public programs. They intend to include, for broad input, the agreed upon draft measure sets in the Physician Fee Schedule and other proposed rules. For measures that are not currently in CMS programs, CMS would go through the annual pre-rulemaking and rulemaking processes to solicit stakeholder and public input. Depending on public response, these measures will be included in a timeframe determined by the Agency.

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**

The performance rates show variation with no trend of improvement. There are differences across each measurement period, but given the limited data available conclusions about the significance of these variations cannot be determined.

These rates indicate the opportunity for continued performance improvement.

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.**

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.**
There have been no reports of unintended consequences with this measure.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

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

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

5a. Harmonization

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.)

Appendix

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

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** American Society of Clinical Oncology
Co.2 **Point of Contact:** Tayyaba, Shehzadi, Tayyaba.Shehzadi@asco.org, 571-483-1673-
Co.3 **Measure Developer if different from Measure Steward:** American Society of Clinical Oncology
Co.4 **Point of Contact:** Tayyaba, Shehzadi, Tayyaba.Shehzadi@asco.org, 571-483-1673-

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.
ASCO Palliative Measures Development Panel
The panel is responsible for reviewing evidence and maintaining measures

Tracey Evans, MD (Chair)
University of Pennsylvania

Craig Earle, MD, FASCO (Co-Chair)
Institute for Clinical Evaluative Science

Katherine Ast, MSW, LCSW
American Academy of Hospice and Palliative Medicine

Amy Berman

The John A. Hartford Foundation

Kathleen Bickel, MD, MPhil
White River Junction VA Medical Center

Eduardo Bruera, MD
The University of Texas MD Anderson Cancer Center

Sydney Dy, MD
Johns Hopkins

Esme Finlay, MD
University of New Mexico Cancer Research and Treatment Center

Arif Kamal, MD, MHS, FAAHPM
Duke University

Kristen McNiff, MPH
Dana-Farber Cancer Institute

Michael Neuss, MD, FASCO
Vanderbilt Ingram Cancer Center

John Sprandio, MD
Consultant in Med Onc and Hem Inc

Holley Stallings, RN
Norton Cancer Institute

Jamie Von Roenn, MD, FASCO
American Society of Clinical Oncology

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2005

Ad.3 **Month and Year of most recent revision:** 11, 2015

Ad.4 **What is your frequency for review/update of this measure?** q3years

Ad.5 **When is the next scheduled review/update for this measure?** 12, 2017

Ad.7 **Disclaimers:** These clinical indicators and quality measures are not intended to and should never supplant independent physician judgment with respect to particular patients or clinical situations. Patient care is always subject to the independent physician judgment with respect to particular patients or clinical situations. Patient care is always subject to the independent professional judgment of the treating physician.

Accordingly, QOPI participants' adherence to quality measures contained in this research report is strictly voluntary and discretionary, with the ultimate determination regarding their application to be made by the treating physician in his or her professional judgment and in light of each patient's individual circumstances. ASCO does not endorse the QOPI® measures as guidelines for standards of practice or 'best practices.'

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

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 #: 0215
Measure Title: Proportion of patients who died from cancer not admitted to hospice
Measure Steward: American Society of Clinical Oncology
Brief Description of Measure: Proportion of patients who died from cancer not admitted to hospice
Developer Rationale: Although the use of hospice and other palliative care services at the end of life has increased, many patients are enrolled in hospice less than 3 weeks before their death, which limits the benefit they may gain from these services. By potentially improving quality of life (QOL), cost of care, and even survival in patients with metastatic cancer, palliative care has increasing relevance for the care of patients with cancer (Smith, 2012). The rate of patients who do not have a hospice referral prior to death continues to be higher than desired with one study reporting that more than 30% of patients were not referred and of those patients, only 7% had a documented discussion on the option of palliative care (O'Connor, 2015). Patients who were enrolled in hospice experienced increased survival times along with a reduction in resource use such as aggressive end of life care and hospital admissions; benefits that increased the longer patients were enrollment in hospice (Lee, 2015; Langton, 2014). In addition, Medicare patients were less likely to enroll in hospice in the last 30 days of life than Medicare patients with only 51% of Medicaid patients enrolled versus 64% of Medicare patients (Guadagnolo, 2015).

Citations

Smith, T. J., S. Temin, et al. (2012). "American Society of Clinical Oncology provisional clinical opinion: the integration of palliative care into standard oncology care." J Clin Oncol 30(8): 880-887.

O´Connor, T. L., N. Ngamphaiboon, et al. (2015). "Hospice utilization and end-of-life care in metastatic breast cancer patients at a comprehensive cancer center." J Palliat Med 18(1): 50-55.

Lee, Y. J., J. H. Yang, et al. (2015). "Association between the duration of palliative care service and survival in terminal cancer patients." Support Care Cancer 23(4): 1057-1062.

Langton, J. M., B. Blanch, et al. (2014). "Retrospective studies of end-of-life resource utilization and costs in cancer care using health administrative data: a systematic review." Palliat Med 28(10): 1167-1196.

Guadagnolo, B. A., K. P. Liao, et al. (2015). "Variation in Intensity and Costs of Care by Payer and Race for Patients Dying of Cancer in Texas: An Analysis of Registry-linked Medicaid, Medicare, and Dually Eligible Claims Data." Med Care 53(7): 591-598.

Numerator Statement: Proportion of patients not enrolled in hospice
Denominator Statement: Patients who died from cancer.
Denominator Exclusions: None

Measure Type: Process
Data Source: Administrative claims, Electronic Clinical Data : Registry
Level of Analysis: Clinician : Group/Practice

IF Endorsement Maintenance – Original Endorsement Date: Aug 10, 2009 **Most Recent Endorsement Date:** Aug 09, 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**

Summary of prior review in 2012

- The developer cited studies indicating 1) admission to hospice does not have a detrimental effect on survival among elderly patients with lung cancer and 2) spending more than 3 days in hospice before death increases the likelihood of bereaved family members reporting a) higher quality of end-of-life care, b) no unmet need for help with anxiety or depression, and c) death in the decedent's died in preferred location. The developer also cited an expert consensus paper on the identification of potential indicators of quality of end of life cancer care using administrative data.
- In the previous evaluation of the measures, the Committee agreed that the developer provided good evidence to support that hospice referral would reflect increased quality of care. Members also noted that poor performance on the measure would indicate that providers may be failing to have direct conversations with patients about the futility of further treatment and the benefits of hospice care.

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 provides a diagram of the relationship of this process of care (enrollment in palliative care/hospice) to better patient outcomes (positive death experience, improved patient and caregiver/family satisfaction, reduced resource utilization and costs, and fewer in-hospital deaths).
- The developers reference the following as evidence of these relationships:
  - A 2012 provisional clinical opinion from the American Society of Clinical Oncology outlined evidence-based recommendations from an expert panel on the incorporation of palliative care services into standard oncology care. This panel reviewed 7 RCTs and found there is a need for more research on palliative and oncology care integration. Members noted that several of the available studies found benefits to integration in terms of improved survival, reduced depression, improved caregiver and/or patient quality of life, and overall lower resource use and cost.
  - A 2013 Cochrane Collaborative systematic review of 7 controlled studies examined the effect home-based palliative care on outcomes for adults with advanced illness and their caregivers. Results indicated that home-based palliative care significantly increased the likelihood of the individual dying at home.

- Individual articles by Smith (2012), Lee (2015), Langton (2014), and Guadagnolo (2015).

Exception to evidence – N/A

Guidance from the Evidence Algorithm
Process measure based on systematic review (Box 3) → QQC presented (Box 4) → Quantity: high; Quality: high; Consistency high (Box 5) → High (Box 5a)

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? Do you agree that evidence pertaining to palliative at-home services be used to support the relationship between hospice admissions and outcomes?*
- *Do you know of additional empirical evidence linking hospice admission to positive death experience, improved patient and caregiver/family satisfaction, reduced resource utilization and costs, and fewer in-hospital deaths?*
- *Although the guidelines have been updated/new studies have been provided, the underlying evidence presented appears to be the directionally the same/stronger since the last NQF endorsement evaluation. 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: ☒ High ☐ Moderate ☐ Low ☐ Insufficient
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 from the ASCO Quality Oncology Practice Initiative registry (QOPI) for 2013-2015 were provided, as follows. It is unclear whether the data represent all patients in the registry or a sample.

	2013	2014	2015
Number of practices	180	172	222
Number of charts	4,959	5,035	7,258
Total Measure Population (%)	42.67	44.39	42.64
Mean	41.44	42.6	42.53
Minimum	0	0	0
Maximum	100	100	100
Standard Deviation	21.13	21.62	20.9
10th percentile	67.38	71.79	71.42
25th percentile	55.55	55.54	55
Median	40	41.67	41.42
75th percentile	26.57	28.03	28.30
90th percentile	14.49	14.29	16.67
95th percentile	5.28	6.25	8.33

Disparities

- Although patient-level disparities data from the QOPI registry were provided, practice-level data are needed.

Questions for the Committee:

- *What is meant by "Total Measure Population"? Are there enough practices and patients represented to make a determination about opportunity for improvement?*
- *Is there a gap in care that warrants a national performance measure?*

- *Does the Committee have evidence that there are practice-level racial or other disparities in hospice admission for cancer patients? Besides gender and race, does the Committee have evidence of other disparities in hospice admission for cancer patients?*

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.
*The developer provided updated evidence. They have provided strong evidence by citing many articles. Transitioning from palliative care to hospice is easier than going from treatment into hospice

1b.
*Performance data from QOPI was presented. It is not clear if the data represents all patients or just a sample. Disparity data was presented from the registry but practice-level data is needed.

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 sources for this measure include administrative claims and registry (i.e., the Quality Oncology Practice Initiative (QOPI®) registry).

Specifications:

- The measure is specified at clinician group/practice level of analysis.
- A lower score on the measure is indicative of better quality.
- Care settings for this measure include clinician offices or clinics, hospices, and hospitals/acute care facilities.
- The ASCO Quality Oncology Practice Initiative (QOPI®) registry is populated by practices that abstract data from a random sample of their patients twice a year. The abstract tool utilized by these practices was not provided.
- The numerator includes those without claims in the Medicare Hospice file OR those ware are listed as not enrolled in hospice within the QOPI® registry. It is not clear how hospice admission is determined for those who are not receiving hospice under the Medicare Hospice benefit when the measure is calculated using administrative claims.
- The denominator (patients who died from cancer) includes those listed within the QOPI® registry as those who died as a consequence of his/her cancer or cancer treatment. It is unclear how the denominator would be identified through claims data. The developers briefly mention use of the "death registry" but do not explain how this is used in conjunction with claims data.
- There are no exclusions for this measure.
- A calculation algorithm is provided.
- This measure is not risk-adjusted.

Questions for the Committee :

- *Is use of death registry data required if the measure is calculated from claims data? If so, what is the timeliness of this dataset? If the death registry is not used, can the measure denominator be identified reliably through administrative data?*
- *Are all the data elements clearly defined?*

- *Is the 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:

- For registry data, data element reliability testing was conducted by comparing QOPI data to data re-abstracted from medical records.
- For claims data, ***data element validity testing*** was conducted by comparing claims data to the full medical record. NQF guidance indicates that if data element validity is demonstrated, additional data element reliability testing is not required.

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:

- For registry data:
  - In 2008, data from the QOPI registry was compared to data that were re-abstracted from medical records by QOPI nurse abstractors. The testing sample included 264 records from 44 sites.
  - **If** the QOPI nurse abstractors who did the re-abstraction are considered expert abstractors—and their re-abstracted data are considered the gold standard—then this testing also would meet the requirements for data element validity testing.
- For claims data, see method(s) of validity testing, below.

Results of reliability testing:

- For registry data:
  - Developers report a single kappa value of 0.679.
  - 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 0.679 means that the raters agreed 67.9% of the time over and above what would be expected by chance alone. According to the Landis and Koch classification, this represents "substantial" agreement.
  - NQF guidance indicates that data element testing should be conducted for all critical data elements, although at minimum, results about the numerator, denominator, and exclusions should be provided. For this measure, only a single kappa value is reported, and thus it is unclear whether it applies to the numerator or denominator data element. It likely applies to the numerator, as the developer presumably sampled only cancer decedents from the registry.
- For claims data, see results of validity testing, below.

Guidance from the Reliability Algorithm

For registry data: Precise specifications (Box 1) → empirical reliability testing conducted for registry data (Box 2) → testing at score level not conducted (Box 4) → data element testing conducted with appropriate method, but need to verify testing was for the numerator (Box 9) → Moderate

For claims data: Fairly precise specifications, except for denominator (Box 1) → no empirical reliability testing conducted for claims data (Box 2) →empirical validity testing conducted (Box 3) → FROM VALIDITY ALGORITHM: data element testing for numerator (Box 10) → Accuracy of the denominator data element not demonstrated for claims data, but assume okay if data derived from national death registry (Box 11) → Moderate

Questions for the Committee:

- *Is death from cancer be accurately recorded in the QOPI registry?*
- *Can claims adequately identify cancer deaths? If not, is use of death registry data required?*
- *No updated testing information is presented. The prior testing demonstrated moderate reliability. Does the Committee think there is a need to 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**

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:

- For claims, data element validity testing was done by comparing claims data to the full medical record. This is an appropriate method of data element validation.
- For registry data, data element reliability testing was conducted by comparing QOPI data to data re-abstracted from medical records. This would be considered an appropriate method of data element validation testing **only if** the QOPI nurse abstractors who did the re-abstraction are considered expert abstractors and thus their re-abstracted data considered the gold standard.

Describe any updates to validity testing

- No updated testing was provided.

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 (from previous evaluation):

- For administrative claims data:
  - Claims data for 150 consecutive patients treated for advanced cancer at Boston's Dana-Farber Cancer Institute and Brigham and Women´s Hospital were compared to data from the full medical record. Numerator data elements were compared ***but not the denominator data element***. Dates for the data examined in testing were not provided.

- Face validity assessment:
  - The developer states that face validity was assessed by conducting focus group and structured interviews with end-of-life cancer patients and bereaved caregivers, and then vetted by an expert panel of cancer providers. According to NQF guidance, the face validity of the measure score as a quality indicator may be adequate if it explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. There is not enough information provided to know if the face validity assessment conforms to NQF's requirements.

Validity testing results (from previous evaluation):

- For administrative claims data:
  - Sensitivity=0.24; Specificity=0.96
  - *Sensitivity* measures the proportion of actual positives that are correctly identified as such. A sensitivity value of 0.24 reflects the accuracy of identifying hospice admissions in the claims data when it is present in the medical data (the authoritative source). The developers state a reason for this low sensitivity may be that medical records often lack documentation of a hospice admission.
  - *Specificity* measures the proportion of actual negatives that are correctly identified as such. A specificity value of 0.96 reflects the accuracy of the absence of a hospice admission absence in the claims data when this is not recorded in the medical record data (the authoritative source).
- Face validity assessment:
  - The developer did not provide results from the face validity assessment.

Questions for the Committee:

- *The developer has not provided any new validity testing, even though there is no testing for the denominator element (cancer deaths) for claims data. Does the developer expect the denominator to be identified via a death registry? If so, should these data be assumed valid? If not, is there evidence that claims adequately identify cancer deaths?*
- *Is the test sample adequate to generalize for widespread implementation?*
- *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:

- There are no exclusions for this measure.

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

- The developer states that "*No risk adjustment or risk stratification is necessary because a) the measure is intended to be used for comparison among similar providers; unless there is a reason to believe that one providers' patients have significantly different risks than others, it will not affect relative comparisons. Moreover, the presence of comorbidity should not appreciably affect hospice use at the end of life. If anything it would increase it thereby making performance appear more favorable*."
- No empirical analysis is provided to support lack of adjustment and nothing in the specifications requires comparison of results across similar providers. Note that process measures generally are not risk-adjusted.

Questions for the Committee:

- *Do you agree with the developer's rationale that risk-adjustment or stratification is not needed for this measure?*

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

- Using SEER-Medicare linked data (n=28,777) for 1991-1996, a benchmark <45% of patients not enrolled in hospice prior to death was achieved by the highest performing regions in the country.
- The developers provide performance scores for 2013-2015.

Year	Mean	St. Dev.	Min	10th	25th	50th	75th	90th	Max
2013	41.44	21.13	0	67.38	55.55	40	26.57	14.50	100
2014	42.60	21.62	0	71.79	55.54	41.66	28.03	14.29	100
2015	42.53	20.90	0	71.43	55	41.42	28.30	16.66	100

- The developers conclude "The range of performance across practices suggests there's clinically meaningful variation across physicians' performance."

Question for the Committee:

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

2b6. Comparability of data sources/methods:

- The developer does not provide analysis to assure comparability of results between administrative claims and registry data.

2b7. Missing Data

- The developer describes how missing data are handled in S.22. No information regarding extent of missing data in claims or registry data is provided.

Guidance from the Validity Algorithm:
Specifications consistent with evidence (Box 1) → potential threats to validity somewhat assessed (Box2) →empirical validity testing (Box 3) → data element testing for numerator (Box 10) → Accuracy of the denominator data element not demonstrated for claims data, although may not be needed if required to be obtained from a death registry (Box 11) → Insufficient (or Low, due to poor sensitivity)

Preliminary rating for validity: ☐ **High** ☐ **Moderate** ☐ **Low** ☒ **Insufficient**
Rationale: Additional analysis to assess threats to validity may be needed (risk-adjustment/comparability of data sources). For claims data, testing for the denominator data element for claims was not conducted, but may not be needed if data required to be obtained from a death registry. Sensitivity for numerator is low. For registry data, results cannot be considered demonstration of validity unless the QOPI nurse re-abstraction is considered the gold standard and results are presented for numerator and denominator separately.

Committee pre-evaluation comments

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

2a, 2b.
Reliability – Specifications
* Samples are administrative claims and registries. The random sample is taken twice a year. Use of the data registry is not defined. A calculation algorithm is presented.

Validity – Specifications
* Specifications are consistent with the evidence.

Reliability – Testing
* No updated testing was provided.

Validity Testing
* No updated testing was provided.

Threats to Validity
* The developer does not provide analysis that ensures comparability between claims and registry data. Are the nurse abstractors the gold standard if not needs other demonstration of 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.

- All data elements are in defined fields in electronic clinical data, either in the ASCO Quality Oncology Practice Initiative registry, in electronic administrative claims, or a death registry. Most data elements are routinely generated during care delivery.
- Cancer decedents and hospice recipients are not necessarily easy to identify in administrative claims. As many as 16% of hospice users do not use Medicare to pay for hospice care (http://www.nhpco.org/sites/default/files/public/Statistics_Research/2012_Facts_Figures.pdf).
- No feasibility concerns were raised by the Cancer Steering Committee during the NQF Maintenance of Endorsement review in 2012.

Questions for the Committee:
- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*
- *Is use of a death registry required? If not, how will the measure denominator be identified using claims data?*

Preliminary rating for feasibility: ☐ **High** ☐ **Moderate** ☒ **Low** ☐ **Insufficient**
Rationale: Low for the claims version, as additional data (e.g., from death registry records) likely are needed if the measure is calculated via administrative claims; also, there is no information about how to obtain information regarding hospice admission for those not using the Medicare Hospice Benefit. Feasibility for the registry-based version is high.

Committee pre-evaluation comments
Criteria 3: Feasibility

* The developer does not provide analysis that ensures comparability between claims and registry data. It is not indicated how the hospice population that does not use Medicare will be captured.

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:
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 measure is used the Quality Oncology Practice Initiative (QOPI), a practice-based quality improvement and benchmarking program. While participation in the registry seems to have declined in recent years, there has been a modest increase in the number of practices reporting on this measure.
- The measure is part of America's Health Insurance Plans (AHIP) Medical Oncology Core Measure Set. The AHIP effort is a collaboration of both public and private stakeholders to identify measures that are meaningful to

patients, consumers, and physicians and to reduce variability in measure selection, collection burden, and cost. Payers involved in the collaboration have committed to using for reporting as soon as feasible. By virtue of being included in the AHIP measure set, CMS will consider this measure for inclusion in Medicare quality programs.

Improvement results: While the number of practices reporting to QOPI has increased between 2013-2015, the average performance has not changed appreciably.

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

Potential harms: No potential harms have been reported by the developer.

Feedback: Public comments during the 2012 maintenance of endorsement evaluation by the Cancer Steering Committee supported the measure. The Committee agreed that results from this measure could indicate a need for more hospice facilities or a need for greater physician and patient education around using this resource, leading to improved patient-centered quality of care.

***Questions for the Committee**:*

- *NQF endorsement criteria recommend public reporting of measure results within six years of initial endorsement as a way to maximize transparency. Does the measure have a viable path to use in accountability applications, including 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**

Committee pre-evaluation comments

Criteria 4: Usability and Use

* The measure is not being publicly reported but it is part of AHIP and will be included in the Medicare quality programs. The number of practices has increased but the average performance has not. The developers did not list any potential harms but a public comment cautioned that providers may be dis-incentivized to refer patients to hospice if they are actively dying. The benefits of the measure outweigh any potential unintended consequences. The measure has the potential to indicate a need for more hospice facilities and education of both providers and patients.

Criterion 5: Related and Competing Measures

Related measures:

0210: Proportion of patients who died from cancer receiving chemotherapy in the last 14 days of life
0211: Proportion of patients who died from cancer with more than one emergency department visit in the last 30 days of life
0213: Proportion of patients who died from cancer admitted to the ICU in the last 30 days of life
0216: Proportion of patients who died from cancer admitted to hospice for less than 3 days

Harmonization: N/A

Pre-meeting public and member comments

Name: Michele Galioto, RN, MSN

Organization: Oncology Nursing Society

Comment: Please clarify enrollment in hospice vs. hospice/palliative care services with hospice enrollment as the specific for the measure (specifications may confuse data extraction).

The intent of the measure is not clear - is it that all should be enrolled in hospice at the end of life?

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 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.

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*)**:** 0215

Measure Title: Proportion of patients who died from cancer not admitted to hospice

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): Click here to name the PRO

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: Hospice enrollment

☐ 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.

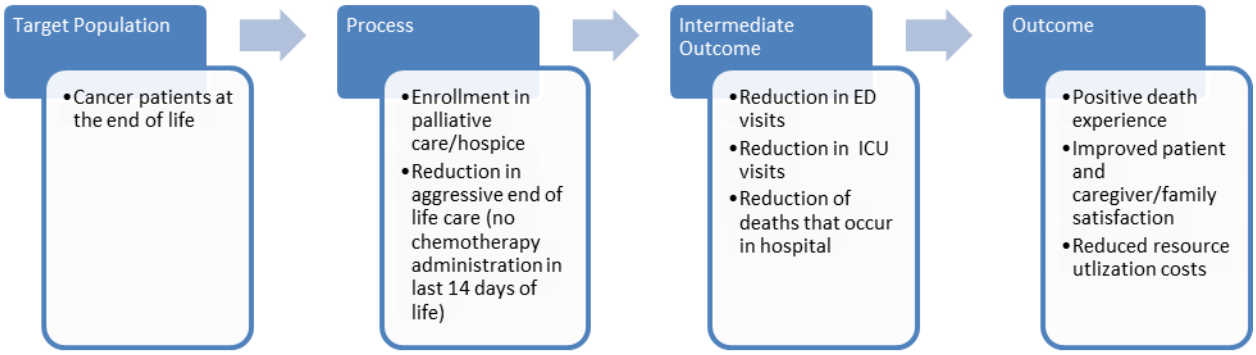
1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

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.

Use from



Although the use of hospice and other palliative care services at the end of life has increased, many patients are enrolled in hospice less than 3 weeks before their death, which limits the benefit they may gain from these services. By potentially improving quality of life (QOL), cost of care, and even survival in patients with metastatic cancer, palliative care has increasing relevance for the care of patients with cancer (Smith, 2012). The rate of patients who do not have a hospice referral prior to death continues to be higher than desired with one study reporting that more than 30% of patients were not referred and of those patients, only 7% had a documented discussion on the option of palliative care (O'Connor, 2015). Patients who were enrolled in hospice experienced increased survival times along with a reduction in resource use such as aggressive end of life care and hospital admissions; benefits that increased the longer patients were enrollment in hospice (Lee, 2015; Langton, 2014). In addition, Medicare patients were less likely to enroll in hospice in the last 30 days of life than Medicare patients with only 51% of Medicaid patients enrolled versus 64% of Medicare patients (Guadagnolo, 2015).

Citations

Smith, T. J., S. Temin, et al. (2012). "American Society of Clinical Oncology provisional clinical opinion: the integration of palliative care into standard oncology care." J Clin Oncol **30**(8): 880-887.

O'Connor, T. L., N. Ngamphaiboon, et al. (2015). "Hospice utilization and end-of-life care in metastatic breast cancer patients at a comprehensive cancer center." J Palliat Med **18**(1): 50-55.

Lee, Y. J., J. H. Yang, et al. (2015). "Association between the duration of palliative care service and survival in terminal cancer patients." Support Care Cancer **23**(4): 1057-1062.

Langton, J. M., B. Blanch, et al. (2014). "Retrospective studies of end-of-life resource utilization and costs in cancer care using health administrative data: a systematic review." Palliat Med **28**(10): 1167-1196.

Guadagnolo, B. A., K. P. Liao, et al. (2015). "Variation in Intensity and Costs of Care by Payer and Race for Patients Dying of Cancer in Texas: An Analysis of Registry-linked Medicaid, Medicare, and Dually Eligible Claims Data." Med Care **53**(7): 591-598.

2012 Submission:

A structural feature: regional availability of hospice, has been shown to correlate with a composite measure of the aggressiveness of cancer care near the end of life that contains this measure. Mostly it is a process measure indicating a possible inadequate focus on palliation and supportive care, that can affect quality of life.

The Process-Outcome link is that admission to hospice, which focuses on improving quality of life, is not associated with a detriment to survival (Saito AM, Landrum MB, Neville BA, et al. Hospice care and survival among elderly patients with lung cancer. J Palliat.Med. 14:929-39, 2011).

In the NIH-funded Cancer Care Outcomes Research and Surveillance Consortium, bereaved family members of 706 lung or colorectal cancer patients rated the quality of end-of-life care their loved one had received. Adjusted for age, sex, marital status, income, education, stage, comorbidity, health system type, census region, and the respondent´s relationship to the patient, respondents were significantly more likely to rate the end-of-life care to have been very good or excellent (86.1 vs 75.4%), with no unmet need for help with anxiety or depression (86.4 vs 78.1%), and that they died in their preferred location (77.3 vs 56.3%) if they were admitted to hospice and had spent more than 3 days in hospice before death (Landrum MB et al, under review).

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***

2012 Submission: Selected individual studies (rather than entire body of evidence)

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*):

Smith TJ, Temin S, Alesi ER, et al. American Society of Clinical Oncology Provisional Clinical Opinion: The Integration of Palliative Care into Standard Oncology Care. J Clin Oncol 2012;30:880-887. Available at: http://www.instituteforquality.org/asco-provisional-clinical-opinion-integration-palliative-care-standard-oncology-care.

Gomes, B., N. Calanzani, et al. (2013). "Effectiveness and cost-effectiveness of home palliative care services for adults with advanced illness and their caregivers." Cochrane Database Syst Rev **6**: CD007760 Available at: http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD007760.pub2/pdf.

2012 Submission: The underlying evidence was obtained by expert consensus, as described in Earle CC, Park ER, Lai B, Weeks JC, Ayanian JZ, Block S. Identifying potential indicators of the quality of end of life cancer care from administrative data. J Clin Oncol. 2003;21(6):1133-8. The panel consisted of oncologists, nurses, palliative care specialists, etc, and used a modified Delphi process to evaluate measures.

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?

A 2012 American Society of Clinical Oncology (ASCO) Provisional Clinical Opinion (PCO) addresses the integration of palliative care (PC) services into standard oncology care at the time a person is diagnosed with metastatic cancer and/or high symptom burden.

A 2013 Cochrane Review, 'Effectiveness and cost-effectiveness of home palliative care services for adults with advanced illness and their caregivers', evaluated the impact of home palliative care services on outcomes for adults with advanced illness or their family caregivers, or both. The aim of the review was to quantify the effect of home palliative care services on a patients' odds of dying at home, examine the clinical effectiveness of home palliative care services on other outcomes such as symptom control, quality of life, caregiver distress and satisfaction with care, and comparing resource use and costs associated with these services.

2012 Submission: The argument is made that because providers cannot predict the future, measures based on decedent cohorts are unfair. However, as described above in 1a.a, the idea is for the measure to be seen as an overall indication of practice style and/or available palliative resources. An individual patient experiencing this process of care has not necessarily received poor quality care. If explanations other than practice style and resource availability, such as unusually

poor prognostic ability on the part of the provider or unexpected toxic deaths (whether unavoidable, from overly aggressive treatment, or poor patient selection) are enough to influence the overall aggregate rates, it is still justifiable to consider it a 'red flag' that should prompt examination of the care provided.

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

2012 ASCO PCO (p. 881):

The American Society of Clinical Oncology (ASCO) has established a rigorous, evidence-based approach—the provisional clinical opinion (PCO)—to offer a rapid response to emerging data in clinical oncology. The PCO is intended to offer timely clinical direction to ASCO's oncologists after publication or presentation of potentially practice- changing data from major studies.

The PCO may serve in some cases as interim direction to the membership pending the development or updating of an ASCO clinical practice guideline. As such, the evidence is not graded in a PCO and is a result of expert consensus. A clinical guideline on palliative care integration with recommendations and the associated grading is under development,

2013 Cochrane Review (p. 12):

Two independent reviewers assessed all included studies for methodological quality using the standard criteria developed by the Cochrane EPOC Review Group for RCTs/CCTs, CBAs and ITSs. The checklist for RCTs/CCTs contains seven items qualified as done, unclear and not done for concealment of allocation, follow-up of professionals, follow up of patients or episodes of care, blinded assessment of primary outcome(s), baseline assessment, reliable primary outcome measure(s) and protection against contamination. Blinding and reliability of all outcomes were also assessed.

Each criterion was scored zero (not done), 0.5 (not clear or when scores varied across outcomes) and one (done). Total scores for RCTs/ CCTs ranged from zero to six; studies with a score of 3.5 or above were considered of high quality. Integration of the results of the quality assessment in data analysis was done in addition to meta-analyses with sensitivity analyses including only high quality studies.

2012 Submission: The underlying evidence was obtained by expert consensus, as described in Earle CC, Park ER, Lai B, Weeks JC, Ayanian JZ, Block S. Identifying potential indicators of the quality of end of life cancer care from administrative data. J Clin Oncol. 2003;21(6):1133-8. The panel consisted of oncologists, nurses, palliative care specialists, etc, and used a modified Delphi process to evaluate measures.

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

See 1a.7.2 for this information.

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

2012 ASCO PCO: 2004-2012

2013 Cochrane Review: 1950 – November 2012

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*)

2012 ASCO PCO: 7 randomized controlled trials

2013 Cochrane Review: 5 randomized controlled trials and 2 controlled clinical trials

2012 Submission: 5 and the studies are observational and use administrative data, consequently there are limitations to the quality of the data.

1a.7.6. What is the overall quality of evidence <u>across studies</u> 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*)

2012 ASCO PCO:

This PCO did not provide an assessment of the overall quality of evidence across the studies. This analysis will be completed during the development of the upcoming clinical guideline.

2013 Cochrane Review :

p. 3: The direction of the effect was consistent across all studies but did not reach statistical significance in 3; ORs ranged from 1. 36 (95% CI 0.80 to 2.31) to 2. 86 (95% CI 0.78 to 10.53) Sensitivity analyses showed that exclusion of the 2 CCTs (both of Swedish hospital-based services with a pooled OR 3.44, 95% CI 0.60 to 19.57) and inclusion of only high quality RCTs resulted in a reduction of the OR to 1.28 (95% CI 1.28 to 2.33) and 1.75 (95% CI 1.24 to 2. 47) respectively, with more precision and less heterogeneity.

p. 22:

Pooled data from seven studies (five RCTs, three of high quality, and two CCTs with 1222 participants) showed that those receiving home palliative care had statistically significantly higher odds of dying at home than those receiving usual care (OR 2.21, 95% CI 1.31 to 3.71; Z = 2.98, P value = 0.003; Chi^2 = 20.57, degrees of freedom (df) = 6, P value = 0.002; I^2 = 71%). The study population control risk was of 307 home deaths per 1000 deaths; based on this ACR of 0.307, the NNTB was 5 (95% CI 3 to 14), meaning that for one additional patient to die at home five more would need to receive home palliative care as opposed to usual care. Assuming a medium cancer home death rate population ACR of 0.278 (i.e. 278 home deaths per 1000 cancer deaths), the NNTB was 6 (95% CI 3 to 15). This means that for one additional cancer patient to die at home in a population where there are 278 home deaths per 1000 cancer deaths, six more would need to receive home palliative care. NNTB estimates ranged from 9 patients (95% CI 5 to 16) when applied to a low home death rate population

such as the one observed in Norway (128 home deaths per 1000 cancer deaths) to 5 patients (95% CI 3 to 13) when applied to a high home death rate population such as the one observed in the Netherlands (454 home deaths per 1000 cancer deaths).

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) <u>across studies</u> 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*)

2012 ASCO PCO (p. 884):

Seven published randomized trials demonstrate the feasibility of providing various components of PC alongside usual oncology care. There is, however, a dearth of data evaluating the integration of modern PC practices into standard oncology care, especially in concert with ongoing antitumor therapy. Overall, the addition of PC interventions to standard oncology care delivered via different models to patients with cancer provided evidence of benefit.

2013 Cochran Review (p. 22):

The study population control risk was of 307 home deaths per 1000 deaths; based on this ACR of 0.307, the NNTB was 5 (95% CI 3 to 14), meaning that for one additional patient to die at home five more would need to receive home palliative care as opposed to usual care. Assuming a medium cancer home death rate population ACR of 0.278 (i.e. 278 home deaths per 1000 cancer deaths), the NNTB was 6 (95% CI 3 to 15). This means that for one additional cancer patient to die at home in a population where there are 278 home deaths per 1000 cancer deaths, six more would need to receive home palliative care. NNTB estimates ranged from 9 patients (95% CI 5 to 16) when applied to a low home death rate population such as the one observed in Norway (128 home deaths per 1000 cancer deaths) to 5 patients (95% CI 3 to 13) when applied to a high home death rate population such as the one observed in the Netherlands (454 home deaths per 1000 cancer deaths).

2012 Submission:

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

2012 ASCO PCO (p. 884-885):

No harm to any patient was observed in any trial, even with discussions of EOL planning, such as hospice and ADs. Two of five trials measuring change in symptoms, two of five studies measuring QOL, two of three studies measuring patient/caregiver satisfaction, and one of three studies measuring survival found statistically significant improvements with PC. Three of six studies measuring mood, two of five studies measuring resource use, and one of four studies measuring outcomes of advance care planning found statistically significant differences, and one outcome of borderline significance was also found in each of these three areas, Therefore, most trials showed benefits ranging from equal to improved overall survival, reduced depression, improved caregiver and/or patient QOL, and overall lower resource use and cost because EOL hospitalizations were avoided.

2013 Cochrane Review: Discussion of harms was not addressed.

2012 Submission: There is no known harm to hospice enrollment.

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.

No relevant studies have been conducted and published since the systematic reviews.

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.

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
0215_Evidence_Form_3.15.16.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)*
Although the use of hospice and other palliative care services at the end of life has increased, many patients are enrolled in hospice less than 3 weeks before their death, which limits the benefit they may gain from these services. By potentially improving quality of life (QOL), cost of care, and even survival in patients with metastatic cancer, palliative care has increasing relevance for the care of patients with cancer (Smith, 2012). The rate of patients who do not have a hospice referral prior to death continues to be higher than desired with one study reporting that more than 30% of patients were not referred and of those patients, only 7% had a documented discussion on the option of palliative care (O'Connor, 2015). Patients who were enrolled in hospice experienced increased survival times along with a reduction in resource use such as aggressive end of life care and hospital admissions; benefits that increased the longer patients were enrollment in hospice (Lee, 2015; Langton, 2014). In addition, Medicare patients were less likely to enroll in hospice in the last 30 days of life than Medicare patients with only 51% of Medicaid patients enrolled versus 64% of Medicare patients (Guadagnolo, 2015).

Citations

Smith, T. J., S. Temin, et al. (2012). "American Society of Clinical Oncology provisional clinical opinion: the integration of palliative care into standard oncology care." J Clin Oncol 30(8): 880-887.

O´Connor, T. L., N. Ngamphaiboon, et al. (2015). "Hospice utilization and end-of-life care in metastatic breast cancer patients at a comprehensive cancer center." J Palliat Med 18(1): 50-55.

Lee, Y. J., J. H. Yang, et al. (2015). "Association between the duration of palliative care service and survival in terminal cancer patients." Support Care Cancer 23(4): 1057-1062.

Langton, J. M., B. Blanch, et al. (2014). "Retrospective studies of end-of-life resource utilization and costs in cancer care using health administrative data: a systematic review." Palliat Med 28(10): 1167-1196.

Guadagnolo, B. A., K. P. Liao, et al. (2015). "Variation in Intensity and Costs of Care by Payer and Race for Patients Dying of Cancer in Texas: An Analysis of Registry-linked Medicaid, Medicare, and Dually Eligible Claims Data." Med Care 53(7): 591-598.

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.*
This data was produced from the QOPI® registry and data was abstracted for a sample of patients seen with the data collection period. Data is reported at the practice level.

In 2013, 180 practices reported on 4959 charts.

In 2014, 172 practices reported on 5035 charts.

In 2015, 222 practices reported on 7258 charts.

	2013	2014	2015
Total Measure Population (%)	42.67	44.39	42.64
Mean	41.44	42.6	42.53
Min	0	0	0
Max	100	100	100
Standard Deviation	21.13	21.62	20.9
Percentiles			
10	67.38	71.79	71.42
25	55.55	55.54	55
50	40	41.67	41.42
75	26.57	28.03	28.30
90	14.49	14.29	16.67
95	5.28	6.25	8.33

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.

This data was produced from the QOPI® registry and data was abstracted for a sample of patients seen with the data collection period. Data is reported at the chart level as practice level disparity data is currently not calculated.

In 2013, 180 practices reported on 4959 charts.

In 2014, 172 practices reported on 5035 charts.

In 2015, 222 practices reported on 7258 charts.

	2013	2014	2015
Total Measure Population	42.67	44.39	42.64
Female	40.67	42.66	40.61
Male	44.43	45.89	44.42
Hispanic	43.88	44.38	46.61
White	41.66	43.43	41.33
Black	48.81	50.32	47.33
Other	41.59	53.17	44.80

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.

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

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.**

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

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

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.)*
No webpage available

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 have been made since the last endorsement

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.
Proportion of patients not enrolled in hospice

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.)*
None

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.
Claims: Those without claims in Medicare HOSPICE file. No codes used.

Registry: Hospice Enrollment = No

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Patients who died from cancer.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
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)*
Claims: Patients in the death registry with cancer as their cause of death. In the cited analyses by the measure submitter, this is a field in the cancer registry or denominator file not requiring specific codes. This may be different in other administrative data sets.

Registry: Deceased = Yes, patient is deceased as a consequence of his/her cancer or cancer treatment

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)*
Not applicable

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)*
Not applicable

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)*
Not applicable

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 = Lower 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.)*
Performance is calculated as:

1. Identify those patients that meet the denominator criteria defined in the measure.

2. Subtract those patients with a denominator exclusion from the denominator. Note: This measure does not have exclusions.

3. From the patients who qualify for the denominator (after any exclusions are removed), identify those who meet the numerator criteria.

4. Calculation: Numerator/Denominator-Denominator Exclusions

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.
Practices that participate in the QOPI registry abstract and upload a random sample of patients twice a year. Practices identify patients who had a diagnosis date in the two years and two office visits in the last six months before the abstraction data period start date. The minimum sample size for each data abstraction period is 40 cases.

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.
Not applicable

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
This measure is specified with defined criteria and data elements. If a patient record does not include one or more of these components for the initial patient population or denominator, then patients are not considered eligible for the measure and not included.

If data to determine whether a patient should be considered for the numerator or exclusions is missing, then the numerator or exclusions not considered to be met and the practice will not get credit for meeting performance for that patient.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Electronic Clinical Data : Registry

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.
ASCO Quality Oncology Practice Initiative (QOPI®)

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)*
Clinician : Group/Practice

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic, Hospice, 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.)*
Not applicable

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
0215_MeasureTesting_MSF5.0_Data_Update.doc

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 0215	**NQF Project:** Cancer Project

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)***:**

The measure was developed using the Medicare claims of all continuously-enrolled patients who died of cancer after having been diagnosed in a SEER region between 1991 and 1996.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)***:**

Evaluation was carried out on 150 consecutive patients treated for advanced cancer at Dana-Farber Cancer Institute and Brigham and Women´s Hospital in Boston. The percent accuracy of death ascertainment for includion into this cohort is unknown but is likely high as the cancer registry regularly uses the death index for ascertainment. Ascertainment would be expected to be highly specific. Hospital billing claims were obtained and analyzed and the accuracy was compared to detailed medical record review.

2a2.3 Testing Results *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**

Sensitivity 0.24 (medical records often do not have documentation of hospice admission. Claims are actually more accurate), Specificity 0.96, where sensitivity = # true positives (both claims and charts)/(# true positives + # false negatives, i.e., not in claims but present in charts) and specificity = # true negatives/(# true negatives + false positives, i.e., present in claims but not in charts).

2b. VALIDITY. Validity, Testing, including all Threats to Validity: 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:**

They are identical

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)***:**

1) Evaluation was carried out on 150 consecutive patients treated for advanced cancer at Dana-Farber Cancer Institute and Brigham and Women´s Hospital in Boston. Claims were obtained and analyzed and the accuracy was compared to detailed medical record review.

2) In QOPI nurse abstractors did a re-abstraction of 264 medical records at 44 sites in 2008.

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

1) Face validity was determined by focus groups and structured interviews with end-of-life cancer patients and bereaved caregivers, and then vetted by an expert panel of cancer providers. The percent agreement between claims and medical record review was calculated.

2) Inter-rater reliability was calculated using Kappa statistics

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)***:**

1) The measure was 88% accurate (percent true positives + true negatives) in the Boston cohort.

2) The Kappa in the QOPI validation study was 0.679

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)***:**

N/A

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

N/A

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)***:**

N/A

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

N/A

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)***:**

N/A

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment: No risk adjustment or risk stratification is necessary because a) the measure is intended to be used for comparison among similar providers; unless there is a reason to believe that one providers' patients have significantly different risks than others, it will not affect relative comparisons. Moreover, tThe presence of comorbidity should not appreciably affect hospice use at the end of life. If anything it would increase it thereby making performance appear more favorable.

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)***:**

used the Medicare claims of all 28,777 continuously-enrolled patients who died of cancer after having been diagnosed in a SEER region between 1991 and 1996. This was an analysis of SEER-Medicare linked data obtained from NCI (http://healthservices.cancer.gov/seermedicare/).

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

Benchmarks were established to identify the outlying 10th decile of practice: The proportion of patients experiencing each process of care in each Health Care Service Area (HCSA) was computed and ranked from best (least aggressive) to worst. A new cohort was created by sequentially adding HCSAs in order starting with the least aggressive until they contained at least 10% of the original cohort and the proportion experiencing each process of care was then recalculated to arrive at the 'Achievable Benchmark of Care). More detail on this, as well as a reference for the Achievable Benchmark of Care method can be found in our publication: Earle CC, Neville BA, Landrum ME, Souza JE, Weeks JC, Block SD, Grunfeld E, Ayanian JZ. Evaluating claims-based indicators of the intensity of end-of-life cancer care. Int J Qual Health Care. 2005;17(6):505-9.

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)*:

A benchmark target of < 45% of patients not enrolled in hospice prior to death corresponds to that achieved by the highest performing regions in the country..

Year	Mean	St. Dev.	Min	10th	25th	50th	75th	90th	Max
2013	41.44	21.13	0	67.38	55.55	40	26.57	14.50	100
2014	42.60	21.62	0	71.79	55.54	41.66	28.03	14.29	100
2015	42.53	20.90	0	71.43	55	41.42	28.30	16.66	100

The range of performance across practices suggests there's clinically meaningful variation across physicians' performance.

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

Comparability of measure results using administrative claims versus registry data has not been analyzed as of March 2016.

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)***:**

Administrative claims and chart review, as described above: 77 entities (HCSAs), 215,484 patients, between 1991 and 2000.

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

We have also assessed the stability of these measures over time by examining the stability of relative aggressive care over time. If the relative aggressiveness of a provider or organization´s practice appeared to change from year to year, then these measures might not be assessing a stable property of practice. To investigate this, we used hierarchical regression models to estimate regional variation in both levels and trends of each measure. We used as our geographic unit of analysis the Health Care Service Area (HCSA). HCSAs are groupings of Metropolitan Statistical Areas defined by the Centers for Medicare & Medicaid Services (CMS) based on observed patient flow patterns in Medicare for tertiary care. As such, each HCSA can be considered to be a self-contained regional health system with a related group of providers. We ranked each region according to the model-estimated rate of each indicator and computed the correlation among relative ranks of each region during the 10-year study period.

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)***:**

We observed significant variation both in levels of aggressive care and in trends in aggressiveness over time, but generally stability of regional practice patterns: Year to year correlation on this measure was 0.98, and over a 5 year span was 0.85.

This provides supportive evidence of the reliability of these measures

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:

N/A

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.**

Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), 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)*

ALL data elements are in defined fields in electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)

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.**

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.**

No feasibility assessment **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.

The measure and its specifications have been in place for several years and ASCO continues to monitor and ensure that the measure and its specifications are up-to-date for widespread use.

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)*.

Not applicable

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)
Payment Program	Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Quality Oncology Practice Initiative (QOPI®) http://www.instituteforquality.org/qopi/manual-qopi-measures

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

Quality Oncology Practice Initiative:
In 2002, the American Society of Clinical Oncology established the Quality Oncology Practice Initiative (QOPI®). QOPI® is a practice-based quality assessment and improvement program designed to foster a culture of self-examination and improvement in oncology. Collection rounds are offered twice per year, in spring and fall, for an eight week period. QOPI® continues to be a successful program in the United States and 12 other countries, with 441, 313, 361 and 256 unique practices participating in Fall 2013, Spring 2014, Spring 2015 and Fall 2015 respectively.

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?*)
QOPI® does not publicly report measure performance by practice; however, it does allow practices to benchmark against multiple external organizations for quality improvement purposes.

We are continuously seeking opportunities to advocate for expanded use of this measure in government or other programs, including those intended for accountability or public reporting.

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 has also been included in America´s Health Insurance Plans Medical Oncology Core Measure Set. The purpose of this program is to reduce variability in measure selection, specifications and implementation. The measures will be implemented nationally by private health plans using a phased-in approach. Contracts between physicians and private payers are individually negotiated and therefore come up for renewal at different points in time depending on the duration of the contract. It is anticipated that private payers will implement these core sets of measures as and when contracts come up for renewal or if existing contracts allow modification of the performance measure set. CMS is also working to align measures across public programs. They intend to include, for broad input, the agreed upon draft measure sets in the Physician Fee Schedule and other proposed rules. For measures that are not currently in CMS programs, CMS would go through the annual pre-rulemaking and rulemaking processes to solicit stakeholder and public input. Depending on public response, these measures will be included in a timeframe determined by the Agency.

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**

The performance rates show variation with no trend of improvement.

These rates indicate the opportunity for continued performance improvement.

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.**

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.**

There have been no reports of unintended consequences with this measure.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

No

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

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

5a. Harmonization

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.)

Appendix

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

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** American Society of Clinical Oncology
Co.2 **Point of Contact:** Tayyaba, Shehzadi, Tayyaba.Shehzadi@asco.org, 571-483-1673-
Co.3 **Measure Developer if different from Measure Steward:** American Society of Clinical Oncology
Co.4 **Point of Contact:** Tayyaba, Shehzadi, Tayyaba.Shehzadi@asco.org, 571-483-1673-

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.
ASCO´s Palliative Measure Development Panel
The panel is responsible for reviewing evidence and maintaining measures

Tracey Evans, MD (Chair)
University of Pennsylvania

Craig Earle, MD, FASCO (Co-Chair)
Institute for Clinical Evaluative Science

Katherine Ast, MSW, LCSW
American Academy of Hospice and Palliative Medicine

Amy Berman
The John A. Hartford Foundation

Kathleen Bickel, MD, MPhil
White River Junction VA Medical Center

Eduardo Bruera, MD
The University of Texas MD Anderson Cancer Center

Sydney Dy, MD
Johns Hopkins

Esme Finlay, MD
University of New Mexico Cancer Research and Treatment Center

Arif Kamal, MD, MHS, FAAHPM
Duke University

Kristen McNiff, MPH
Dana-Farber Cancer Institute

Michael Neuss, MD, FASCO
Vanderbilt Ingram Cancer Center

John Sprandio, MD
Consultant in Med Onc and Hem Inc

Holley Stallings, RN
Norton Cancer Institute

Jamie Von Roenn, MD, FASCO
American Society of Clinical Oncology

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2005

Ad.3 **Month and Year of most recent revision:** 11, 2015

Ad.4 **What is your frequency for review/update of this measure?** q3years

Ad.5 **When is the next scheduled review/update for this measure?** 12, 2017

Ad.7 **Disclaimers:** These clinical indicators and quality measures are not intended to and should never supplant independent physician judgment with respect to particular patients or clinical situations. Patient care is always subject to the independent physician judgment with respect to particular patients or clinical situations. Patient care is always subject to the independent professional judgment of the treating physician.

Accordingly, QOPI participants' adherence to quality measures contained in this research report is strictly voluntary and discretionary, with the ultimate determination regarding their application to be made by the treating physician in his or her professional judgment and in light of each patient's individual circumstances. ASCO does not endorse the QOPI® measures as guidelines for standards of practice or 'best practices.'

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

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 #: 0216
Measure Title: Proportion of patients who died from cancer admitted to hospice for less than 3 days
Measure Steward: American Society of Clinical Oncology
Brief Description of Measure: Proportion of patients who died from cancer, and admitted to hospice and spent less than 3 days there
Developer Rationale: Although the use of hospice and other palliative care services at the end of life has increased, many patients are enrolled in hospice for 3 days or less before their death, which limits the benefit they may gain from these services. One recent retrospective study of more than 64,000 patients with cancer who were admitted to hospice found that over 16% of those patients were only enrolled in the last three days of life or less (O'Connor, 2014). The rate of patients who do not have a hospice referral prior to death continues to be higher than desired with one study reporting that more than 30% of patients were not referred and of those patients, only 7% had a documented discussion on the option of palliative care (O'Connor, 2015). Patients enrolled in hospice experience increased survival times along with a reduction in resource use such as aggressive end of life care and hospital admissions; benefits that increased the longer patients are enrolled in hospice (Lee, 2015; Langton, 2014).

Langton, J. M., B. Blanch, et al. (2014). "Retrospective studies of end-of-life resource utilization and costs in cancer care using health administrative data: a systematic review." Palliat Med 28(10): 1167-1196.

Lee, Y. J., J. H. Yang, et al. (2015). "Association between the duration of palliative care service and survival in terminal cancer patients." Support Care Cancer 23(4): 1057-1062.

O´Connor, T. L., N. Ngamphaiboon, et al. (2015). "Hospice utilization and end-of-life care in metastatic breast cancer patients at a comprehensive cancer center." J Palliat Med 18(1): 50-55.

Numerator Statement: Patients who died from cancer and spent fewer than three days in hospice.
Denominator Statement: Patients who died from cancer who were admitted to hospice
Denominator Exclusions: None

Measure Type: Intermediate Clinical Outcome
Data Source: Administrative claims, Electronic Clinical Data : Registry
Level of Analysis: Clinician : Group/Practice

IF Endorsement Maintenance – Original Endorsement Date: Oct 01, 2007 **Most Recent Endorsement Date:** Aug 09, 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**

Summary of prior review in 2012

- The developer cited studies indicating 1) admission to hospice does not have a detrimental effect on survival among elderly patients with lung cancer and 2) spending more than 3 days in hospice before death increases the likelihood of bereaved family members reporting a) higher quality of end-of-life care, b) no unmet need for help with anxiety or depression, and c) death in the decedent's died in preferred location. The developer also cited an expert consensus paper on the identification of potential indicators of quality of end of life cancer care using administrative data.
- In the previous evaluation of the measure, the Committee agreed that the developer provided good evidence to support that hospice referral would be linked with increased quality of care and poor performance on the measure would indicate that providers may be failing to have direct conversations with patients about the futility of further treatment and the benefits of hospice care.

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 provides a diagram of the relationship of this process of care (enrollment in palliative care/hospice) to better patient outcomes (positive death experience, improved patient and caregiver/family satisfaction, reduced resource utilization and costs).
- The developer cited additional resources for evidence of these relationships:
  - A 2012 provisional clinical opinion from the American Society of Clinical Oncology outlined evidence-based recommendations from an expert panel on the incorporation of palliative care services into standard oncology care. This panel reviewed 7 RCTs and found there is a need for more research on palliative and oncology care integration. Members noted that several of the available studies found benefits to integration in terms of improved survival, reduced depression, improved caregiver and/or patient quality of life, and overall lower resource use and cost.
  - A 2013 Cochrane Collaborative systematic review of 7 controlled studies examined the effect home-based palliative care on outcomes for adults with advanced illness and their caregivers. Results indicated that home-based palliative care significantly increased the likelihood of the individual dying at home.
  - Individual articles by Langton (2014), Lee (2015), and O'Connor (2015).

Exception to evidence: NA

Guidance from the Evidence Algorithm
Process measure based on systematic review (Box 3) → QQC presented (Box 4) → Quantity: high; Quality: high; Consistency high (Box 5) → High (Box 5a)

Questions for the Committee:

- *What is the relationship of this measure to patient outcomes?*
- *What is the evidence for the 3 day timeframe?*
- *How strong is the evidence for this relationship?*
- *Is the evidence directly applicable to the process of care being measured? Do you agree that evidence pertaining to palliative at-home services be used to support the relationship between hospice admissions and outcomes?*
- *Although the guidelines have been updated/new studies have been provided, the underlying evidence presented appears to be the directionally the same/stronger since the last NQF endorsement evaluation. 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: ☒ High ☐ Moderate ☐ Low ☐ Insufficient
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 from the ASCO Quality Oncology Practice Initiative registry (QOPI) for 2013-2015 were provided. It is unclear whether the data represent all patients in the registry or a sample.

	2013	2014	2015
Number of practices	178	170	222
Number of charts	4,951	5,021	7,239
Total Patient Population (%)	10.91	16.79	16.95
Mean	16.63	18.22	17.86
Minimum	0	0	0
Maximum	100	100	100
Standard Deviation	16.46	17.60	14.50
10^{th} percentile	0	0	0
25^{th} percentile	4.76	7.14	7.14
Median	12.97	14.64	15.38
75^{th} percentile	25.00	23.81	25.81
90^{th} percentile	36.36	38.28	33.33
95^{th} percentile	50	50	46.67

Disparities

- Although patient-level disparities data from the QOPI registry were provided, practice-level data are needed.

Questions for the Committee:

- *What is meant by "Total Patient Population"? Are there enough practices and patients represented to make a determination about opportunity for improvement?*
- *Does it appear that performance has gotten worse over time?*
- *Is there a gap in care that warrants a national performance measure?*
- *Does the Committee have evidence that there are practice-level racial or other disparities in patients who died*

from cancer and were admitted to hospice for less than 3 days?

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.

*The developer provided updated evidence for this measure along with a diagram of the relationship of process of care and better patient outcomes. The developer cited several sources including 3 individual articles, a 2013 Cochrane review and a 2012 opinion from ASCO that supported this relationship. Given the additional evidence I recommend a high score and ask the committee to agree to not vote on Evidence.

1b.

* It is unclear what is meant by total patient population in the performance data. It is also unclear if the data represent all patients in the registry or just a sample. It is difficult to ascertain if performance has gotten worse over time.Patient level disparities data was presented but it is needs to be provided at the practice level. I recommend a moderate score.

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 sources for this measure include administrative claims and registry (i.e., the Quality Oncology Practice Initiative (QOPI®) registry).

Specifications:

- The measure is specified at clinician-group/practice level of analysis.
- A lower score on the measure is indicative of better quality.
- Care settings for this measure include clinician offices or clinics, hospices, or hospitals/acute care facilities.
- The ASCO Quality Oncology Practice Initiative (QOPI®) registry is populated by practices that abstract data from a random sample of their patients twice a year. The abstract tool utilized by these practice was not provided.
- The numerator can be calculated one of two ways:
  - With claims Medicare Hospice file: (death date) - (hospice admission date) ≤ 3 days
  - With QOPI registry data: (date of death) – (hospice admission date) ≤ 3 days
- The denominator (patients who died from cancer) includes those listed within the QOPI® registry as those who died as a consequence of his/her cancer or cancer treatment. It is unclear how the denominator would be identified through claims data. The developers briefly mention use of the "death registry" but do not explain how this is used in conjunction with claims data.
- There are no exclusions to this measure.
- A calculation algorithm is provided.
- This measure is not risk-adjusted.

Questions for the Committee :

- *Is use of death registry data required if the measure is calculated from claims data? If so, what is the timeliness of this dataset? If the death registry is not used, can the measure denominator be identified reliably through administrative data?*
- *How are hospice patients in the denominator identified if decedent did not use the Medicare Hospice benefit?*

- *Are all data elements clearly identified?*
- *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:

- For registry data, data element reliability testing was conducted by comparing QOPI data to data re-abstracted from medical records.
- For claims data, ***data element validity testing*** was conducted by comparing claims data to the full medical record. NQF guidance indicates that if data element validity is demonstrated, additional data element reliability testing is not required.

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:

- For registry data:
  - In 2008, data from the QOPI registry was compared to data that were re-abstracted from medical records by QOPI nurse abstractors. The testing sample included 264 records from 44 sites.
  - **If** the QOPI nurse abstractors who did the re-abstraction are considered expert abstractors—and their re-abstracted data are considered the gold standard—then this testing also would meet the requirements for data element validity testing.
- For claims data, see method(s) of validity testing, below.

Results of reliability testing:

- For registry data:
  - Developers report a single kappa value of 0.551.
  - 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 0.551 means that the raters agreed 55.1% of the time over and above what would be expected by chance alone. According to the Landis and Koch classification, this represents "moderate" agreement.
  - NQF guidance indicates that data element testing should be conducted for all critical data elements, although at minimum, results about the numerator, denominator, and exclusions should be provided. For this measure, only a single kappa value is reported, and thus it is unclear whether applies to the numerator or denominator data element. It likely applies to the numerator, as the developer presumably sampled only cancer decedents from the registry.
- For claims data, see results of validity testing, below.

Guidance from the Reliability Algorithm

For registry data: Precise specifications (Box 1) → empirical reliability testing conducted for registry data (Box 2) → testing at score level not conducted (Box 4) → data element testing conducted with appropriate method, but need to verify testing was for the numerator (Box 9) → Moderate

For claims data: Fairly precise specifications, except for denominator (Box 1) → no empirical reliability testing conducted for claims data (Box 2) →empirical validity testing conducted (Box 3) → FROM VALIDITY ALGORITHM: data element testing for numerator (Box 10) → Accuracy of the denominator data element not demonstrated for claims data, but assume okay if data derived from national death registry (Box 11) → Moderate

Questions for the Committee:

- *Is death from cancer be accurately recorded in the QOPI registry?*
- *Can claims adequately identify cancer deaths? If not, is use of death registry data required?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*
- *No updated testing information is presented. The prior testing demonstrated moderate reliability. Does the Committee think there is a need to 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**

- During the 2012 evaluation, the Cancer Steering Committee questioned why three days was selected as the numerator. The developer noted that three days is the minimum lowest bar but that; seven days may be a better indicator of quality of care. They also noted that data are more easily obtained with the three day threshold than the seven day threshold.

Questions for the Committee:

- *Are the specifications consistent with the evidence? Is there evidence for the 3 day timeframe?*
- *Is the 3-day threshold an appropriate threshold for this measure?*

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:

- For claims, data element validity testing was done by comparing claims data to the full medical record. This is an appropriate method of data element validation.
- For registry data, data element reliability testing was conducted by comparing QOPI data to data re-abstracted from medical records. This would be considered an appropriate method of data element validation testing **only if** the QOPI nurse abstractors who did the re-abstraction are considered expert abstractors and thus their re-abstracted data considered the gold standard.

Describe any updates to validity testing

- No updated testing provided.

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 (from previous evaluation):

- For administrative claims data:
  - Claims data for 150 consecutive patients treated for advanced cancer at Boston's Dana-Farber Cancer Institute and Brigham and Women´s Hospital were compared to data from the full medical record. Numerator data elements were compared ***but not the denominator data element***. Dates for the data examined in testing were not provided.
- Face validity assessment:
  - The developer states that face validity was assessed by conducting focus group and structured interviews with end-of-life cancer patients and bereaved caregivers, and then vetted by an expert panel of cancer providers. According to NQF guidance, the face validity of the measure score as a quality indicator may be adequate if it explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. There is not enough information provided to know if the face validity assessment conforms to NQF's requirements.

Validity testing results:

- For administrative claims data:
  - Sensitivity=0.97; Specificity=1.00
  - *Sensitivity* measures the proportion of actual positives that are correctly identified as such. A sensitivity value of 0.97 reflects the accuracy of identifying a hospice stay of <3 days in the claims data when it is present in the medical record data (the authoritative source).
  - *Specificity* measures the proportion of actual negatives that are correctly identified as such. A specificity value of 1.00 reflects the accuracy of the absence of a hospice stay of <3 days in the claims data when this is not recorded in the medical record data (the authoritative source).
- Face validity assessment:
  - The developer does not provide results from the face validity assessment.

Questions for the Committee:

- *The developer has not provided any new validity testing, even though there is no testing for the denominator element (cancer deaths) for claims data. Does the developer expect the denominator to be identified via a death registry? If so, should these data be assumed valid? If not, is there evidence that claims adequately identify cancer deaths?*
- *Is the test sample adequate to generalize for widespread implementation?*
- *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:

- There are no exclusions for this measure.

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

Conceptual rationale for SDS factors included ? ☐ **Yes** ☒ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary:

- In their earlier submission for the 2012 evaluation, the developer states that risk-adjustment
- In the current submission, the developer states, "*Traditionally, ASCO believed that risk adjustment or risk stratification was not critical for these palliative care measures. The measures are used for comparison among*

similar providers and there is no expectation that performance will be 0%. ASCO is exploring whether risk adjustment is appropriate and if so how it could be accomplished for our outcome measures. ASCO's palliative care measures would be included in that assessment. While several studies indicate potential variables for consideration, additional time is needed to properly define a statistical risk model method and variables, including sociodemographic (SDS) factors now that they may be considered."

Questions for the Committee:

- *Do you believe that risk-adjustment or stratification is needed for this measure?*

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

- Using SEER-Medicare linked data (n=28,777) for 1991-1996, a benchmark <8% of patients enrolled in hospice in the last 3 days of life was achieved by the highest performing regions in the country.
- The developers provide performance scores for 2013-2015 and suggest that the variation across practices indicates clinically meaningful variation across physicians' performance.

Question for the Committee:

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

2b6. Comparability of data sources/methods:

- The developer describes analysis of claims data to explore stability of performance across time and across Health Care Service Area regions. However, these analysis do not address whether results from claims vs. registry are comparable.

2b7. Missing Data

- The developer describes how missing data are handled in S.22.

Guidance from the Validity Algorithm:
Specifications consistent with evidence (Box 1) → potential threats to validity somewhat assessed (Box2) → empirical validity testing (Box 3) → data element testing for numerator (Box 10) → Accuracy of the denominator data element not demonstrated for claims data, although may not be needed if required to be obtained from a death registry (Box 11) → Insufficient (or Moderate, if Committee decides no need for risk-adjustment)

Preliminary rating for validity: ☐ **High** ☐ **Moderate** ☐ **Low** ☒ **Insufficient**
Rationale: Additional analysis to assess threats to validity may be needed (risk-adjustment/comparability of data sources). For claims data, testing for the denominator data element for claims was not conducted, but may not be needed if data required to be obtained from a death registry. For registry data, results cannot be considered demonstration of validity unless the QOPI nurse re-abstraction is considered the gold standard and results are presented for numerator and denominator separately.

Committee pre-evaluation comments

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

2a, 2b.
Reliability – Specifications
*A calculation algorithm is provided. The abstract tool used in the QOPI was not provided. It is unclear how the denominator will be generated in the claims data. A reference is made to a death registry but it is not explained how it will be used. It is unclear how patients in the denominator who used hospice outside of Medicare are identified. I recommend a score of moderate.

Validity – Specifications
* The question is raised about using 3-day as opposed to 7-day as the appropriate threshold for enrollment in hospice. The developers stated that a 3-day threshold is the minimum lowest bar and easier to obtain. They also noted that it

was a better indicator of quality of care.

Reliability – Testing
* No updated testing information is presented. There is a question as to the ability of the QOPI registry to accurately record deaths due to cancer or cancer treatment and if this is consistent with the information in the claims data. Does the committee think we need to vote on reliability if no new information is presented. The previous score was moderate.

Validity Testing
* For registry data a QOPI nurse abstractor was used the question is whether the abstractor is considered an expert. For administrative claims data the numerators were compared but not the denominators.The sensitivity was 0.97 and specificity was 1.00 for administrative claims data. Face validity assessment was done utilizing a focus group but the results were not provided. It is difficult do ascertain if the measure is an indicator of quality.

Threats to Validity
* It is not clear if there is a need for risk adjustment. It is not clear if claims data and registry data are comparable. It needs to be determined if the nurse re-abstractor is the gold standard. I would rate it low.

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.

- All data elements are in defined fields in electronic clinical data, either in the ASCO Quality Oncology Practice Initiative registry, administrative claims, or a death registry. Most data elements are routinely generated during care delivery.
- Cancer decedents and hospice recipients are not necessarily easy to identify in administrative claims.
- No feasibility concerns were raised by the Cancer Steering Committee during the NQF Maintenance of Endorsement review in 2012.

Questions for the Committee:
- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*
- *Is use of a death registry required? If not, how will the measure denominator be identified using claims data?*

Preliminary rating for feasibility: ☐ **High** ☐ **Moderate** ☒ **Low** ☐ **Insufficient**
Rationale: Low for the claims version, as additional data (e.g., from death registry records) likely are needed if the measure is calculated via administrative claims; also, there is no information about how to obtain information regarding hospice admission for those not using the Medicare Hospice Benefit. Feasibility for the registry-based version is high.

Committee pre-evaluation comments
Criteria 3: Feasibility

* Most data elements are routinely generated during care of delivery. The issue of claims version and the possible need for death registry data. It is not clear how non-Medicare hospice patients will be tracked. Rating is Low.

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
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 measure is used the Quality Oncology Practice Initiative (QOPI), a practice-based quality improvement program, and can be submitted to CMS to meet PQRS requirements. In Fall 2015, 36 practices and 3,124 patient charts were submitted to PQRS through QOPI. Practice-level data (for which this measure is specified) will be publicly reported on Physician Compare in 2017.
- The measure is part of America's Health Insurance Plans (AHIP) Medical Oncology Core Measure Set. The AHIP effort is a collaboration of both public and private stakeholders to identify measures that are meaningful to patients, consumers, and physicians and to reduce variability in measure selection, collection burden, and cost. Payers involved in the collaboration have committed to using for reporting as soon as feasible. By virtue of being included in the AHIP measure set, CMS will consider this measure for inclusion in Medicare quality programs.

Improvement results: While the number of practices reporting to QOPI has increased between 2013-2015, the average performance has not changed appreciably.

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

Potential harms: The submission does not list any potential unintended consequences. A public comment on the final MAP 2016 recommendations cautioned that providers may be dis-incentivized to refer patients to hospice care if they are actively dying.

Feedback: NQF#0216 was supported for inclusion in the Merit-Based Incentive Program by the Measure Applications Partnership (MAP) in 2016. MAP, supported by public comments, requested the Standing Committee re-evaluate the a longer timeframe for this measure. In the 2012 review for maintenance of endorsement, the Cancer Steering Committee underlined the importance of NQF# 0216, citing the potential to indicate a need for more hospice facilities, or a need to educate both patients and clinicians about palliative care.

***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**

Committee pre-evaluation comments

Criteria 4: Usability and Use

* The measure is not being publicly reported but it is part of AHIP and will be included in the Medicare quality programs. The number of practices has increased but the average performance has not. The developers did not list any potential harms but a public comment cautioned that providers may be dis-incentivized to refer patients to hospice if they are actively dying. The benefits of the measure outweigh any potential unintended consequences. The measure has the potential to indicate a need for more hospice facilities and education of both providers and patients. Rating is high.

Criterion 5: Related and Competing Measures

Related measures:

0210: Proportion of patients who died from cancer receiving chemotherapy in the last 14 days of life

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

0213: Proportion of patients who died from cancer admitted to the ICU in the last 30 days of life
0215: Proportion of patients who died from cancer not admitted to hospice

Harmonization: N/A

Pre-meeting public and member comments

Name: Michele Galioto, RN, MSN
Organization: Oncology Nursing Society
Comment: Should the measure include immunotherapy? Does “chemotherapy” include hormonal and biotherapy? It may be more inclusive to refer to all as "antineoplastic therapy."

Should the measure include radiation therapy as well?

Does “death from cancer” include all death within 14 days? Death may be the result of infection, accident (e.g., fall), bleeding, etc. which could be tied to cancer or cancer treatment. Death attributed to side effects of therapy may be indistinguishable from cancer deaths. Is the intent that death occurs within a timeframe of receiving chemotherapy?

Please clarify enrollment in hospice vs. hospice/palliative care services with hospice enrollment as the specific for the measure (specifications may confuse data extraction).

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 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*)**:** 0216

Measure Title: Proportion of patients who died from cancer admitted to hospice for less than 3 days

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): Click here to name the PRO

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

☒ Intermediate clinical outcome (*e.g., lab value*): Patients admitted to hospice for less than 3 days

☐ Process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

2012 Submission: Selected individual studies (rather than entire body of evidence)

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

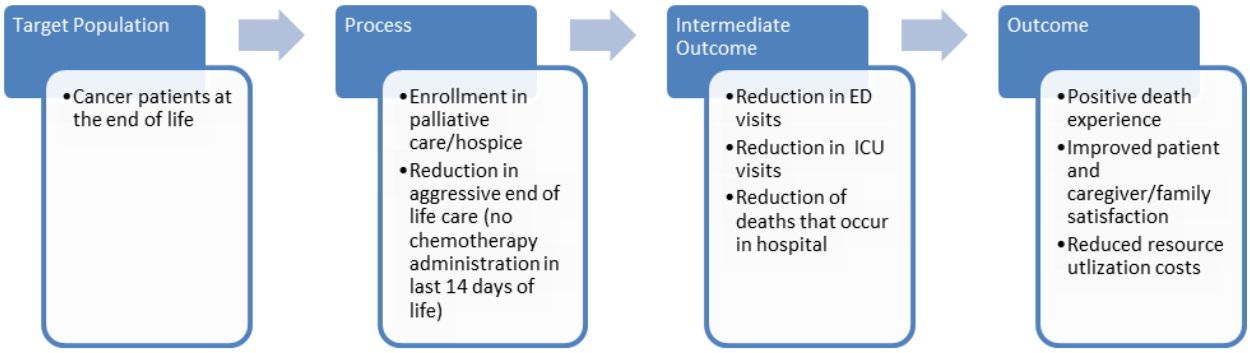
1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

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*).

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.



Although the use of hospice and other palliative care services at the end of life has increased, many patients are enrolled in hospice for 3 days or less before their death, which limits the benefit they may gain from these services. One recent retrospective study of more than 64,000 patients with cancer who were admitted to hospice found that over 16% of those patients were only enrolled in the last three days of life or less (O'Connor, 2014). The rate of patients who do not have a hospice referral prior to death continues to be higher than desired with one study reporting that more than 30% of patients were not referred and of those patients, only 7% had a documented discussion on the option of palliative care (O'Connor, 2015). Patients enrolled in hospice experience increased survival times along with a reduction in resource use such as aggressive end of life care and hospital admissions; benefits that increased the longer patients are enrolled in hospice (Lee, 2015; Langton, 2014).

Langton, J. M., B. Blanch, et al. (2014). "Retrospective studies of end-of-life resource utilization and costs in cancer care using health administrative data: a systematic review." Palliat Med **28**(10): 1167-1196.

Lee, Y. J., J. H. Yang, et al. (2015). "Association between the duration of palliative care service and survival in terminal cancer patients." Support Care Cancer **23**(4): 1057-1062.

O'Connor, T. L., N. Ngamphaiboon, et al. (2015). "Hospice utilization and end-of-life care in metastatic breast cancer patients at a comprehensive cancer center." J Palliat Med **18**(1): 50-55.

2012 Submission: A structural feature: regional availability of hospice, has been shown to correlate with a composite measure of the aggressiveness of cancer care near the end of life that contains this measure. Mostly it is a process measure indicating a possible inadequate focus on palliation and supportive care, that can affect quality of life.

The Process-Outcome link is that early admission to hospice, which focuses on improving quality of life, is not associated with a detriment to survival (Saito AM, Landrum MB, Neville BA, et al. Hospice care and survival among elderly patients with lung cancer. J Palliat.Med. 14:929-39, 2011).

In the NIH-funded Cancer Care Outcomes Research and Surveillance Consortium, bereaved family members of 706 lung or colorectal cancer patients rated the quality of end-of-life care their loved one had received. Adjusted for age, sex, marital status, income, education, stage, comorbidity, health system type, census region, and the respondent´s relationship to the

patient, respondents were significantly more likely to rate the end-of-life care to have been very good or excellent (86.1 vs 75.4%), with no unmet need for help with anxiety or depression (86.4 vs 78.1%), and that they died in their preferred location (77.3 vs 56.3%) if they had spent more than 3 days in hospice before death (Landrum MB et al, under review).

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***

2012 Submission: Selected individual studies (rather than entire body of evidence)

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*):

Smith TJ, Temin S, Alesi ER, et al. American Society of Clinical Oncology Provisional Clinical Opinion: The Integration of Palliative Care into Standard Oncology Care. J Clin Oncol 2012;30:880-887. Available at: http://www.instituteforquality.org/asco-provisional-clinical-opinion-integration-palliative-care-standard-oncology-care.

Gomes, B., N. Calanzani, et al. (2013). "Effectiveness and cost-effectiveness of home palliative care services for adults with advanced illness and their caregivers." Cochrane Database Syst Rev **6**: CD007760 Available at: http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD007760.pub2/pdf.

2012 Submission: The underlying evidence was obtained by expert consensus, as described in Earle CC, Park ER, Lai B, Weeks JC, Ayanian JZ, Block S. Identifying potential indicators of the quality of end of life cancer care from administrative data. J Clin Oncol. 2003;21(6):1133-8. The panel consisted of oncologists, nurses, palliative care specialists, etc, and used a modified Delphi process to evaluate measures.

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?

A 2012 American Society of Clinical Oncology (ASCO) Provisional Clinical Opinion (PCO) addresses the integration of palliative care (PC) services into standard oncology care at the time a person is diagnosed with metastatic cancer and/or high symptom burden.

A 2013 Cochrane Review, 'Effectiveness and cost-effectiveness of home palliative care services for adults with advanced illness and their caregivers', evaluated the impact of home palliative care services on outcomes for adults with advanced illness or their family caregivers, or both. The aim of the review was to quantify the effect of home palliative care services on a patients' odds of dying at home, examine the clinical effectiveness of home palliative care services on other outcomes such as symptom control, quality of life, caregiver distress and satisfaction with care, and comparing resource use and costs associated with these services.

2012 Submission: The argument is made that because providers cannot predict the future, measures based on decedent cohorts are unfair. However, as described above in 1a.a, the idea is for the measure to be seen as an overall indication of practice style and/or available palliative resources. An individual patient experiencing this process of care has not necessarily received poor quality care. If explanations other than practice style and resource availability, such as unusually poor prognostic ability on the part of the provider or unexpected toxic deaths (whether unavoidable, from overly aggressive treatment, or poor patient selection) are enough to influence the overall aggregate rates, it is still justifiable to consider it a 'red flag' that should prompt examination of the care provided.

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

2012 ASCO PCO (p. 881):

The American Society of Clinical Oncology (ASCO) has established a rigorous, evidence-based approach—the provisional clinical opinion (PCO)—to offer a rapid response to emerging data in clinical oncology. The PCO is intended to offer timely clinical direction to ASCO's oncologists after publication or presentation of potentially practice- changing data from major studies.

The PCO may serve in some cases as interim direction to the membership pending the development or updating of an ASCO clinical practice guideline. As such, the evidence is not graded in a PCO and is a result of expert consensus. A clinical guideline on palliative care integration with recommendations and the associated grading is under development.

2013 Cochrane Review (p. 12):

Two independent reviewers assessed all included studies for methodological quality using the standard criteria developed by the Cochrane EPOC Review Group for RCTs/CCTs, CBAs and ITSs. The checklist for RCTs/CCTs contains seven items qualified as done, unclear and not done for concealment of allocation, follow-up of professionals, follow up of patients or episodes of care, blinded assessment of primary outcome(s), baseline assessment, reliable primary outcome measure(s) and protection against contamination. Blinding and reliability of all outcomes were also assessed.

Each criterion was scored zero (not done), 0.5 (not clear or when scores varied across outcomes) and one (done). Total scores for RCTs/ CCTs ranged from zero to six; studies with a score of 3.5 or above were considered of high quality. Integration of the results of the quality assessment in data analysis was done in addition to meta-analyses with sensitivity analyses including only high quality studies.

2012 Submission: All studies have shown similar results.

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

See 1a.7.2 for this information.

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

2012 ASCO PCO: 2004-2012

2013 Cochrane Review: 1950 – November 2012

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*)

2012 ASCO PCO: 7 randomized controlled trials

2013 Cochrane Review: 5 randomized controlled trials and 2 controlled clinical trials

1a.7.6. What is the overall quality of evidence <u>across studies</u> 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*)

2012 ASCO PCO:

This PCO did not provide an assessment of the overall quality of evidence across the studies. This analysis will be completed during the development of the upcoming clinical guideline.

2013 Cochrane Review :

p. 3: The direction of the effect was consistent across all studies but did not reach statistical significance in 3; ORs ranged from 1. 36 (95% CI 0.80 to 2.31) to 2. 86 (95% CI 0.78 to 10.53) Sensitivity analyses showed that exclusion of the 2 CCTs (both of Swedish hospital-based services with a pooled OR 3.44, 95% CI 0.60 to 19.57) and inclusion of only high quality RCTs resulted in a reduction of the OR to 1.28 (95% CI 1.28 to 2.33) and 1.75 (95% CI 1.24 to 2. 47) respectively, with more precision and less heterogeneity.

p. 22:

Pooled data from seven studies (five RCTs, three of high quality, and two CCTs with 1222 participants) showed that those receiving home palliative care had statistically significantly higher odds of dying at home than those receiving usual care (OR 2.21, 95% CI 1.31 to 3.71; Z = 2.98, P value = 0.003; Chi^2 = 20.57, degrees of freedom (df) = 6, P value = 0.002; I^2 = 71%). The study population control risk was of 307 home deaths per 1000 deaths; based on this ACR of 0.307, the NNTB was 5 (95% CI 3 to 14), meaning that for one additional patient to die at home five more would need to receive home palliative care as opposed to usual care. Assuming a medium cancer home death rate population ACR of 0.278 (i.e. 278 home deaths per 1000 cancer deaths), the NNTB was 6 (95% CI 3 to 15). This means that for one additional cancer patient to die at home in a population where there are 278 home deaths per 1000 cancer deaths, six more would need to receive home palliative care. NNTB estimates ranged from 9 patients (95% CI 5 to 16) when applied to a low home death rate population such as the one observed in Norway (128 home deaths per 1000 cancer deaths) to 5 patients (95% CI 3 to 13) when applied to a high home death rate population such as the one observed in the Netherlands (454 home deaths per 1000 cancer deaths).

2012 Submission: The studies are observational and use administrative data, consequently there are limitations to the quality of the data. Still, the sensitivity in claims is 0.97, specificity 1.00, accuracy 0.97 and variability 2.39 (95% CI 1.99-2.95)

REF: 3. Earle CC, Neville BA, Landrum ME, Souza JE, Weeks JC, Block SD, Grunfeld E, Ayanian JZ. Evaluating claims-based indicators of the intensity of end-of-life cancer care. Int J Qual Health Care. 2005;17(6):505-9.

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) <u>across studies</u> 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*)

2012 ASCO PCO (p. 884):

Seven published randomized trials demonstrate the feasibility of providing various components of PC alongside usual oncology care. There is, however, a dearth of data evaluating the integration of modern PC practices into standard oncology care, especially in concert with ongoing antitumor therapy. Overall, the addition of PC interventions to standard oncology care delivered via different models to patients with cancer provided evidence

of benefit.

2013 Cochrane Review (p. 22):

The study population control risk was of 307 home deaths per 1000 deaths; based on this ACR of 0.307, the NNTB was 5 (95% CI 3 to 14), meaning that for one additional patient to die at home five more would need to receive home palliative care as opposed to usual care. Assuming a medium cancer home death rate population ACR of 0.278 (i.e. 278 home deaths per 1000 cancer deaths), the NNTB was 6 (95% CI 3 to 15). This means that for one additional cancer patient to die at home in a population where there are 278 home deaths per 1000 cancer deaths, six more would need to receive home palliative care. NNTB estimates ranged from 9 patients (95% CI 5 to 16) when applied to a low home death rate population such as the one observed in Norway (128 home deaths per 1000 cancer deaths) to 5 patients (95% CI 3 to 13) when applied to a high home death rate population such as the one observed in the Netherlands (454 home deaths per 1000 cancer deaths).

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

2012 ASCO PCO (p. 884-885):

No harm to any patient was observed in any trial, even with discussions of EOL planning, such as hospice and ADs. Two of five trials measuring change in symptoms, two of five studies measuring QOL, two of three studies measuring patient/caregiver satisfaction, and one of three studies measuring survival found statistically significant improvements with PC. Three of six studies measuring mood, two of five studies measuring resource use, and one of four studies measuring outcomes of advance care planning found statistically significant differences, and one outcome of borderline significance was also found in each of these three areas, Therefore, most trials showed benefits ranging from equal to improved overall survival, reduced depression, improved caregiver and/or patient QOL, and overall lower resource use and cost because EOL hospitalizations were avoided.

2013 Cochrane Review: Discussion of harms was not addressed.

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.

No relevant studies have been conducted and published since the systematic reviews.

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

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 #: 0216
De.2. **Measure Title:** Proportion of patients who died from cancer admitted to hospice for less than 3 days
Co.1.1. **Measure Steward:** American Society of Clinical Oncology
De.3. **Brief Description of Measure:** Proportion of patients who died from cancer, and admitted to hospice and spent less than 3 days there
1b.1. **Developer Rationale:** Although the use of hospice and other palliative care services at the end of life has increased, many patients are enrolled in hospice for 3 days or less before their death, which limits the benefit they may gain from these services. One recent retrospective study of more than 64,000 patients with cancer who were admitted to hospice found that over 16% of those patients were only enrolled in the last three days of life or less (O'Connor, 2014). The rate of patients who do not have a hospice referral prior to death continues to be higher than desired with one study reporting that more than 30% of patients were not referred and of those patients, only 7% had a documented discussion on the option of palliative care (O'Connor, 2015). Patients enrolled in hospice experience increased survival times along with a reduction in resource use such as aggressive end of life care and hospital admissions; benefits that increased the longer patients are enrolled in hospice (Lee, 2015; Langton, 2014).

Langton, J. M., B. Blanch, et al. (2014). "Retrospective studies of end-of-life resource utilization and costs in cancer care using health administrative data: a systematic review." Palliat Med 28(10): 1167-1196.

Lee, Y. J., J. H. Yang, et al. (2015). "Association between the duration of palliative care service and survival in terminal cancer patients." Support Care Cancer 23(4): 1057-1062.

O´Connor, T. L., N. Ngamphaiboon, et al. (2015). "Hospice utilization and end-of-life care in metastatic breast cancer patients at a comprehensive cancer center." J Palliat Med 18(1): 50-55.

S.4. **Numerator Statement:** Patients who died from cancer and spent fewer than three days in hospice.
S.7. **Denominator Statement:** Patients who died from cancer who were admitted to hospice
S.10. **Denominator Exclusions:** None

De.1. **Measure Type:** Intermediate Clinical Outcome
S.23. **Data Source:** Administrative claims, Electronic Clinical Data : Registry
S.26. **Level of Analysis:** Clinician : Group/Practice

IF Endorsement Maintenance – Original Endorsement Date: Oct 01, 2007 **Most Recent Endorsement Date:** Aug 09, 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
0216_Evidence_Form_3.16.16.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)*
Although the use of hospice and other palliative care services at the end of life has increased, many patients are enrolled in hospice for 3 days or less before their death, which limits the benefit they may gain from these services. One recent retrospective study of more than 64,000 patients with cancer who were admitted to hospice found that over 16% of those patients were only enrolled in the last three days of life or less (O'Connor, 2014). The rate of patients who do not have a hospice referral prior to death continues to be higher than desired with one study reporting that more than 30% of patients were not referred and of those patients, only 7% had a documented discussion on the option of palliative care (O'Connor, 2015). Patients enrolled in hospice experience increased survival times along with a reduction in resource use such as aggressive end of life care and hospital admissions; benefits that increased the longer patients are enrolled in hospice (Lee, 2015; Langton, 2014).

Langton, J. M., B. Blanch, et al. (2014). "Retrospective studies of end-of-life resource utilization and costs in cancer care using health administrative data: a systematic review." Palliat Med 28(10): 1167-1196.

Lee, Y. J., J. H. Yang, et al. (2015). "Association between the duration of palliative care service and survival in terminal cancer patients." Support Care Cancer 23(4): 1057-1062.

O´Connor, T. L., N. Ngamphaiboon, et al. (2015). "Hospice utilization and end-of-life care in metastatic breast cancer patients at a comprehensive cancer center." J Palliat Med 18(1): 50-55.

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.*
This data was produced from the QOPI® registry and data was abstracted for a sample of patients seen with the data collection period. Data is reported at the practice level.

In 2013, 178 practices reported on 4951 charts.

In 2014, 170 practices reported on 5021 charts.

In 2015, 222 practices reported on 7239 charts.

	2013	2014	2015
Total Patient Population (%)	10.91	16.79	16.95
Mean	16.63	18.22	17.86
Minimum	0	0	0
Maximum	100	100	100
Standard Deviation	16.46	17.60	14.50
Percentiles			
10	0	0	0
25	4.76	7.14	7.14
50	12.97	14.64	15.38
75	25.00	23.81	25.81
90	36.36	38.28	33.33

95	50	50.00	46.67

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.*
This data was produced from the QOPI® registry and data was abstracted for a sample of patients seen with the data collection period. Data is calculated at the chart level as demographics are not currently calculated at the practice level.

In 2013, 178 practices reported on 4951 charts.

In 2014, 170 practices reported on 5021 charts.

In 2015, 222 practices reported on 7239 charts.

	2013	2014	2015
Total Measure Population	10.91	16.79	16.95
Female	10.56	14.86	16.38
Male	11.24	18.57	17.48
Hispanic	10.91	25.84	19.83
White	12.02	16.18	17.08
Black	7.89	19.66	21.33
Other	10.61	13.79	13.87

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

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.

1c.4. Citations for data demonstrating high priority provided in 1a.3

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

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.)*
No webpage available

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 have been made since last endorsement

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 who died from cancer and spent fewer than three days in hospice.

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.)*
Last 3 days of life

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.
Claims: Medicare HOSPICE file only:
Subtracted hospice admission date (admndate) from death date variable to get hospice length of stay.

Registry:
Date of Death – Hospice Enrollment Date </= 3 days

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Patients who died from cancer who were admitted to hospice

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
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)*
Claims: Patients in the death registry with cancer as their cause of death who also appear in the Medicare hospice file. In the cited analyses by the measure submitter, this is a field in the cancer registry or denominator file not requiring specific codes. This may be different in other administrative data sets.

Registry:
Deceased = Yes, patient is deceased as a consequence of his/her cancer or cancer treatment
AND
Hospice Enrollment = Yes

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)*
Not applicable

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)*
Not applicable

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)*
Not applicable

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 = Lower 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.)*

Performance is calculated as:
1. Identify those patients that meet the denominator criteria defined in the measure.
2. Subtract those patients with a denominator exclusion from the denominator. Note: this measure does not have any denominator exclusions
3. From the patients who qualify for the denominator (after any exclusions are removed), identify those who meet the numerator criteria.
4. Calculation: Numerator/Denominator-Denominator Exclusions

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.

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.
Not applicable

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
This measure is specified with defined criteria and data elements. If a patient record does not include one or more of these components for the initial patient population or denominator, then patients are not considered eligible for the measure and not included.

If data to determine whether a patient should be considered for the numerator or exclusions is missing, then the numerator or exclusions not considered to be met and the practice will not get credit for meeting performance for that patient.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Electronic Clinical Data : Registry

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.
ASCO Quality Oncology Practice Initiative (QOPI®)

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)*
Clinician : Group/Practice

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic, Hospice, 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.)*
Not applicable

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
0216_MeasureTesting_MSF5.0_Data_Update-635937462215054041.doc

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 0216	**NQF Project:** Cancer Project

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)***:**

The measure was developed using the Medicare claims of all continuously-enrolled patients who died of cancer after having been diagnosed in a SEER region between 1991 and 1996.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)***:**

Evaluation was carried out on 150 consecutive patients treated for advanced cancer at Dana-Farber Cancer Institute and Brigham and Women´s Hospital in Boston. The percent accuracy of death ascertainment for includion into this cohort is unknown but is likely high as the cancer registry regularly uses the death index for ascertainment. Ascertainment would be expected to be highly specific. Hospital billing claims were obtained and analyzed and the accuracy was compared to detailed medical record review.

2a2.3 Testing Results *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**

Sensitivity 0.97, Specificity 1.00 , where sensitivity = # true positives (both claims and charts)/(# true positives + # false negatives, i.e., not in claims but present in charts) and specificity = # true negatives/(# true negatives + false positives, i.e., present in claims but not in charts).

2b. VALIDITY. Validity, Testing, including all Threats to Validity: 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:**

They are identical

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)***:**

1) Evaluation was carried out on 150 consecutive patients treated for advanced cancer at Dana-Farber Cancer Institute and Brigham and Women´s Hospital in Boston. Claims were obtained and analyzed and the accuracy was compared to detailed medical record review.

2) In QOPI nurse abstractors did a re-abstraction of 264 medical records at 44 sites in 2008.

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

1) Face validity was determined by focus groups and structured interviews with end-of-life cancer patients and bereaved caregivers, and then vetted by an expert panel of cancer providers. The percent agreement between claims and medical record review was calculated.

2) Inter-rater reliability was calculated using Kappa statistics

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)***:**

1) This measure was 97% accurate (percent true positives + true negatives).

2) The Kappa in the QOPI validation study was 0.551

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)***:**

N/A

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

N/A

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)*:

N/A

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

Traditionally, ASCO believed that risk adjustment or risk stratification was not critical for these palliative care measures. The measures are used for comparison among similar providers and there is no expectation that performance will be 0%.

ASCO is exploring whether risk adjustment is appropriate and if so how it could be accomplished for our outcome measures. ASCO's palliative care measures would be included in that assessment. While several studies indicate potential variables for consideration, additional time is needed to properly define a statistical risk model method and variables, including sociodemographic (SDS) factors now that they may be considered.

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)*:

N/A

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment: No risk adjustment or risk stratification is necessary because the measure is intended to be used for comparison among similar providers; unless there is a reason to believe that one providers' patients have significantly different risks than others, it will not affect relative comparisons. Moreover, tThe presence of comorbidity should not appreciably affect hospice use at the end of life. If anything it would increase it thereby making performance appear more favorable.

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)*:

We used the Medicare claims of all 28,777 continuously-enrolled patients who died of cancer after having been diagnosed in a SEER region between 1991 and 1996. This was an analysis of SEER-Medicare linked data obtained from NCI (http://healthservices.cancer.gov/seermedicare/).

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

Benchmarks were established to identify the outlying 10th decile of practice: The proportion of patients experiencing each process of care in each Health Care Service Area (HCSA) was computed and ranked from best (least aggressive) to worst. A new cohort was created by sequentially adding HCSAs in order starting with the least aggressive until they contained at least 10% of the original cohort and the proportion experiencing each process of care was then recalculated to arrive at the 'Achievable Benchmark of Care). More detail on this, as well as a reference for the Achievable Benchmark of Care method can be found in our publication: Earle CC, Neville BA, Landrum ME, Souza JE, Weeks JC, Block SD, Grunfeld E, Ayanian JZ. Evaluating claims-based indicators of the intensity of end-of-life cancer care. Int J Qual Health Care. 2005;17(6):505-9.

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)*:

A benchmark target of < 8% of patients enrolled in hospice within only the last 3 days of life corresponds to that achieved by the highest performing regions in the country.

Year	Mean	St. Dev.	Min	10th	25th	50th	75th	90th	Max
2013	11.47	11.87	0	0	3.35	9.88	15.81	24.26	100
2014	12.92	12.58	0	0	5.80	11.45	17.07	21.88	100
2015	13.16	11.50	0	0	6.45	11.95	16.67	23.08	100

The range of performance suggests there's clinically meaningful variation across physicians' 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)***:**

Administrative claims and chart review, as described above: 77 entities (HCSAs), 215,484 patients, between 1991 and 2000.

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

We have also assessed the stability of these measures over time by examining the stability of relative aggressive care over time. If the relative aggressiveness of a provider or organization´s practice appeared to change from year to year, then these measures might not be assessing a stable property of practice. To investigate this, we used hierarchical regression models to estimate regional variation in both levels and trends of each measure. We used as our geographic unit of analysis the Health Care Service Area (HCSA). HCSAs are groupings of Metropolitan Statistical Areas defined by the Centers for Medicare & Medicaid Services (CMS) based on observed patient flow patterns in Medicare for tertiary care. As such, each HCSA can be considered to be a self-contained regional health system with a related group of providers. We ranked each region according to the model-estimated rate of each indicator and computed the correlation among relative ranks of each region during the 10-year study period.

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):

We observed significant variation both in levels of aggressive care and in trends in aggressiveness over time, but generally stability of regional practice patterns: Year to year correlation on this measure was 0.98, and over a 5 year span was 0.79. This provides supportive evidence of the reliability of these measures.

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:

N/A

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.**

Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), 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)*

ALL data elements are in defined fields in electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)

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.**

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.**

No feasibility assessment **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.

The measure and its specifications have been in place for several years and ASCO continues to monitor and ensure that the measure and its specifications are up-to-date for widespread use.

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)*.

Not applicable

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)
Public Reporting	Payment Program CMS Physician Quality Reporting Program Qualified Clinical Data Registry http://www.instituteforquality.org/qopi/pqrs-measures-0 Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Quality Oncology Practice Initiative (QOPI®) http://www.instituteforquality.org/qopi/manual-qopi-measures

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

Quality Oncology Practice Initiative:
In 2002, the American Society of Clinical Oncology established the Quality Oncology Practice Initiative (QOPI®). QOPI® is a practice-based quality assessment and improvement program designed to foster a culture of self-examination and improvement in oncology. Collection rounds are offered twice per year, in spring and fall, for an eight week period. QOPI® continues to be a successful program in the United States and 12 other countries, with 441, 313, 361 and 256 unique practices participating in Fall 2013, Spring 2014, Spring 2015 and Fall 2015 respectively.

PQRS Qualified Clinical Data Registry:
In addition to the current use for quality improvement with benchmarking in the QOPI® registry, this measure has been reported to CMS by the registry as a Qualified Clinical Data Registry. QOPI® is approved to be a Qualified Clinical Data Registry and this measure is included in the list of measures that can be reported by participating practices. QOPI® was deemed as a QCDR to report to PQRS in 2015 and 2016. Eligible professionals will be considered to have satisfactorily participated in PQRS if they submit quality measures data or results to CMS via a qualified clinical data registry. In Fall 2015, 36 practices and 3,124 patient charts were submitted to PQRS through QOPI. In late 2016, all 2015 individual-level QCDR data will be available for public reporting on Physician Compare. In 2017, both individual and group-level QCDR measures will be available for public reporting.

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?*)
As described above, this measure is included in the CMS PQRS program. CMS is planning to publicly report QCDR data. This measure

Additionally, although the measure is currently in use, we will continue to seek opportunities to advocate for expanded use of this measure in government or other programs.

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.*)
The measure was recently considered for CMS MIPS program by the NQF Measures Application Partnership MAP and received a recommendation of support. With the MAP recommendation, it can now be considered for inclusion in the next proposed rule for this program with the earliest implementation of 2017.

This measure has also been included in Americas Health Insurance Plans Medical Oncology Core Measure Set. The purpose of this program is to reduce variability in measure selection, specifications and implementation. The measures will be implemented

nationally by private health plans using a phased-in approach. Contracts between physicians and private payers are individually negotiated and therefore come up for renewal at different points in time depending on the duration of the contract. It is anticipated that private payers will implement these core sets of measures as and when contracts come up for renewal or if existing contracts allow modification of the performance measure set. CMS is also working to align measures across public programs. They intend to include, for broad input, the agreed upon draft measure sets in the Physician Fee Schedule and other proposed rules. For measures that are not currently in CMS programs, CMS would go through the annual pre-rulemaking and rulemaking processes to solicit stakeholder and public input. Depending on public response, these measures will be included in a timeframe determined by the Agency.

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**

The performance rates show variation with no trend of improvement. These rates indicate the opportunity for continued performance improvement.

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.**

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.**
There have been no reports of unintended consequences with this measure.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

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

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

5a. Harmonization

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.)

Appendix

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

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** American Society of Clinical Oncology
Co.2 **Point of Contact:** Tayyaba, Shehzadi, Tayyaba.Shehzadi@asco.org, 571-483-1673-
Co.3 **Measure Developer if different from Measure Steward:** American Society of Clinical Oncology
Co.4 **Point of Contact:** Tayyaba, Shehzadi, Tayyaba.Shehzadi@asco.org, 571-483-1673-

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.
ASCO Palliative Measure Development Panel
The panel is responsible for reviewing evidence and maintaining measures

Tracey Evans, MD (Chair)
University of Pennsylvania

Craig Earle, MD, FASCO (Co-Chair)
Institute for Clinical Evaluative Science

Katherine Ast, MSW, LCSW
American Academy of Hospice and Palliative Medicine

Amy Berman
The John A. Hartford Foundation

Kathleen Bickel, MD, MPhil
White River Junction VA Medical Center

Eduardo Bruera, MD
The University of Texas MD Anderson Cancer Center

Sydney Dy, MD
Johns Hopkins

Esme Finlay, MD
University of New Mexico Cancer Research and Treatment Center

Arif Kamal, MD, MHS, FAAHPM
Duke University

Kristen McNiff, MPH
Dana-Farber Cancer Institute

Michael Neuss, MD, FASCO
Vanderbilt Ingram Cancer Center

John Sprandio, MD
Consultant in Med Onc and Hem Inc

Holley Stallings, RN
Norton Cancer Institute

Jamie Von Roenn, MD, FASCO
American Society of Clinical Oncology

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2005
Ad.3 **Month and Year of most recent revision:** 11, 2005
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?** 12, 2017

Ad.7 **Disclaimers:** These clinical indicators and quality measures are not intended to and should never supplant independent physician judgment with respect to particular patients or clinical situations. Patient care is always subject to the independent physician judgment with respect to particular patients or clinical situations. Patient care is always subject to the independent professional judgment of the treating physician.

Accordingly, QOPI participants' adherence to quality measures contained in this research report is strictly voluntary and discretionary, with the ultimate determination regarding their application to be made by the treating physician in his or her professional judgment and in light of each patient's individual circumstances. ASCO does not endorse the QOPI® measures as guidelines for standards of practice or 'best practices.'

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

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 #: 1641
Measure Title: Hospice and Palliative Care – Treatment Preferences
Measure Steward: University of North Carolina-Chapel Hill
Brief Description of Measure: Percentage of patients with chart documentation of preferences for life sustaining treatments.
Developer Rationale: Seriously ill and dying patients who are given the opportunity to express life-sustaining treatment preferences are more likely to receive care consistent with their values, and patient and family satisfaction outcomes improve. Patients and physicians alike hesitate to initiate discussions, while acknowledging their value and desirability. Use of the Treatment Preferences quality measure will improve attention to this important practice, in order to enhance patient autonomy, facilitate patient-centered decision-making, and communicate patient preferences via documentation to other treating providers.

Numerator Statement: Patients whose medical record includes documentation of life sustaining preferences
Denominator Statement: Seriously ill patients enrolled in hospice OR receiving specialty palliative care in an acute hospital setting.
Denominator Exclusions: Patients with length of stay < 1 day in hospice or palliative care

Measure Type: Process
Data Source: Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record
Level of Analysis: Clinician : Group/Practice, 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 provided a rationale for the relationship of this process of care (documenting preferences for life sustaining treatment) and patient outcomes (e.g., patient autonomy and control over treatments; patient and

family satisfaction with care; improved transitions to hospice and palliative care; reduced emotional distress for family).

- The developer cited individual studies and systematic reviews examining the impact of communication quality on patient and family outcomes. The study findings support the effect of high-quality communication in reducing hospice utilization, family distress, and the use of intensive treatments treatment concordant with preferences in seriously and terminally ill patient populations. The systematic reviews do not appear to explicitly examine the relationship between *documentation* of care preferences and patient or family outcomes.
- During its 2012 evaluation of the measure, the Palliative/End-of-Life (EOL) Committee agreed that there is evidence demonstrating a need for a discussion around life-sustaining treatment preferences; they also noted that poor communication is major quality concern for palliative and end-of-life care.

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:

- Although the developer attested that there have been no changes in the evidence since it was endorsed in 2012, they added two new guidelines to the submission.
    - The 2014 Michigan Quality Improvement Consortium guideline relates to advanced care planning for patients with a serious illness or whose death within the next 12 months would not be unexpected. It is based on a number of clinical expert reports and it calls for the incorporation of the patient's preferences and choices into the Treatment Preferences portion of the Advance Directive.
    - The 2013 ICSI guideline on Palliative Care for Adults. The pertinent recommendations from this graded guideline states that "Clinicians should initiate or facilitate advance care planning for all adult patients and their families with regular review as the patient's condition changes [Low Quality Evidence, Strong Recommendation]" and "Clinicians should engage in shared decision-making with the patient and/or their families when establishing or revising goals of care [Low Quality Evidence, Strong Recommendation]".

Exception to evidence – N/A

Guidance from the Evidence Algorithm

Process measure based on systematic review (Box 3) → QQC presented (Box 4) → Quantity: high; Quality: moderate; Consistency: high (Box 5) → Moderate (Box 5b) → Moderate

Questions for the Committee:

- *Is this a measure of documentation of a discussion or a measure of documentation of actual preferences?*
- *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?*
- *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 discussion and vote on Evidence?*

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

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.

- Facility-Level (Hospice) :

- Data from the Hospice Quality Reporting Program (HQRP) for FY15 are provided. The data are from 3,922 hospice organizations and approximately 1.2 million patient stays.
- The developer reported that 53.5% of hospices had perfect scores, and 4.6% of hospices scored below 90%.

Mean	98.0%
Range	0% - 100%
10th percentile	95.0%
25th percentile	98.5%
Median	100%
75th percentile	100%
90th percentile	100%

- Clinician Group/Practice Level of Analysis (Palliative Care):
  - Data specific to palliative care (for clinician group/practice in the hospital setting) are not yet available, although the developer expects these data will become available for NQF review next year.

Disparities

- Facility-level (Hospice):

Subgroup	**Hospice-level**	
	Score	**Wilcoxon-Mann-Whitney test p-value**
Race		
Non-white ≥ national median of 11.9%	98.4%	p=0.005
Non-white < national median of 11.9%	97.6%	
Sex		
Female ≥ national median of 55.2%	97.9%	p=0.70
Female < national median of 55.2%	98.0%	
Medicaid status (proxy for SES)		
Medicaid patients >= national median of 21.5%	97.4%	p<0.001
Medicaid patients < national median of 21.5%	98.5%	
Geographic location		
Rural	98.8%	p<0.001
Urban	97.7%	

- Clinician Group/Practice Level (Palliative Care):
  - Disparities data for palliative care are not available, although the developer expects these data will become available for NQF review next year.

Questions for the Committee:

- *Although there are statistically significant differences in performance between hospices for particular subgroups, the developer notes these may not be clinically meaningful. Is there a gap in hospice care (for documenting treatment preferences) that warrants a national performance measure?*
- *Is the Committee aware of evidence demonstrating a gap in hospital-based palliative care (for documenting treatment preferences) that warrants a national performance measure?*
- *Palliative-specific disparities information is not provided. Are you aware of evidence that disparities exist in this area of healthcare?*

Preliminary rating for opportunity for improvement: ☐ **High** ☐ **Moderate** ☒ **Low** ☐ **Insufficient**
Rationale: Performance for this measure in the hospice setting appears to be "topped out"; while statistically

significant, the differences in performance across different population subgroups likely are not clinically meaningful.

Committee pre-evaluation comments

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

1a.

* Process Measure- Maintenance. Quantity of studies, and quality related moderate of the studies. Specific guideline recommendation for this measure.
* Evidence supports importance of activity.
* This measure documents the percentage of seriously ill patients enrolled in hospice or receiving specialty palliative care in an acute hospital setting whose medical record includes documentation of life sustaining preferences. Technically, this measure documents the documentation of life sustaining treatment preferences rather than the percentage of patients in these settings who have expressed those preferences or been given the opportunity to express them. However, it would seem reasonable to assume that the percentage of patients with documented preferences is a realistic surrogate for the percentage of patients who have expressed those preferences. If one could assume that refusal to state preferences was also documented and thus counted, this measure can be a true surrogate for the percentage of patients offered this opportunity.

To relate this measure directly to the outcomes documented in the research one must also assume that medical record documentation is at least the predominant way treatment preferences are communicated which seems reasonable. Given that caveat, the relationship of the measure to outcomes would be considered strong mainly on the basis of its consistency. The quantity, at least in terms of the number of studies is high with several systematic reviews and individual studies reported. The evidence presented by the developers does not give enough detail to be sure that the quality of those studies meets NQF standards but it would seem that a grade of at least moderate as the developers report would be warranted. As the developers report, the consistency is quite high with virtually all evidence pointing toward the clinical benefits that are important to patients and families and no significant harms reported. The quality as reported is moderate in that a number of the benefits are indirect.

The developer reports that the evidence has not change although they do report two new guidelines. Given that the previously reported evidence was highly consistent and the new guidelines are consistent with that direction, the new evidence does not significantly impact the evidence evaluation. Therefore it is reasonable to state that the evidence has not changed.

* The process of obtaining and recording certain patients' Life Sustaining Treatment preferences is being measured. Whether the preferences were "discussed" is assumed by either the presense of the patient's signature on the POLST form in the medical record or further notes of a discussion with the patient or health care proxy.

Thus, yes, the obtaining of the preference (a process and intended outcome) is directly measured; but, whether a true consultation and discussion occurred with the patient is less clear and directly shown through this measure.

Question: Could the latter be obtained through other evidence such as billing codes for the advance care discussion?

1b.

* Performance data provided. Low preliminary rating for opportunity for improvement.
* not much evidence of a performance gap or differences among subgroups
* The gap in care/opportunities for improvement does need consideration by the committee. In doing that, it should be noted that this measure is approved for two distinct contexts- hospice and specialty palliative care in an acute hospital setting. The developers have reported gap data only for the hospice context as the palliative care data is currently being revised. Thus the evidence in the palliative care context would seem to be insufficient.

It would seem that the measure meets the NQF standards for "topped out" in the hospice context. All the compliance numbers are well over 90% and the median is reported as 100%. In that context, the developers' contention that the disparities reported are not clinically significant since the range is very small and discrete scores are all very high is appropriate. Thus it would seem that the committee should consider this measure in the hospice contest for Inactive Endorsement with Reserve Status. Certainly the number of people impacted by this measure is very high and the importance of this measure to patients and families is also arguably very high. What is not clear is, given how hard wired this measure now is in the hospice system, whether discontinuing the measure would significantly reduce compliance.

The issue of the measure in the specialty palliative care context deserves separate consideration including a review of

the literature on disparities in this area which is beyond this reviewer's expertise and a report from the developers on compliance in this population.
* Yes, performance data on the measure was provided. Because the denominator in the measure only includes seriously ill patients enrolled in hospice or receiving receiving specialty palliative care in an acute hospital setting, the performance on the measure is high and there is much variation and not much room for improvement (low).

More information is needed for the palliative care in the clinician group/practice level of analysis and that is expected for review by NQF next year.

The developer recommends removing an exclusion which would broaden the applicable pool of data --- to delete the requirement that the patient be in hospice at least 7 days.

Question: Should we wait until next year when we have more data on the palliative care portion of 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): The data sources specified in the submission form include EHRs and other electronic clinical data.

Specifications:

- This measure is specified for the clinician group/practice level of analysis in the hospital setting and for the facility level of analysis in the hospice setting. A higher score indicates better quality.
- For hospice settings, the numerator (patient whose medical record includes documentation of life sustaining preferences) and denominator (patients enrolled in hospice) are identified from the Hospice Item Set, a standardized, patient-level dataset used in the Medicare Hospice Quality Reporting Program.
- For palliative care (hospital) settings, data for the numerator and denominator (patients admitted to palliative care) are collected using a structured medical record abstraction tool. For this setting, patients with a length of stay <1 day are excluded from the denominator.
- Documentation using the POLST paradigm with evidence of patient or surrogate involvement, such as co-signature or description of discussion can be counted in this numerator.
- Data are collected at the hospice admission evaluation or initial clinical encounter for palliative care.
- A calculation algorithm is provided. However, this algorithm calls for documentation of a ***discussion*** of preference for life sustaining treatments, rather than documentation of the actual preferences, as is described in the numerator statement.
- The developer indicates that sampling is permissible for the hospital setting. While some basic instructions are given, there is no guidance about the number of patients needed for the sample.

Changes to specifications since previous evaluation:

- After analysis of FY15 hospice data, the developer has changed the specifications so as to no longer exclude hospice stays of less than 7 days.

Questions for the Committee :

- *Is this a measure of documentation of a discussion or a measure of documentation of actual preferences?*
- *Are all the data elements clearly defined?*
- *Is the logic or calculation algorithm clear?*
- *Is it reasonable to exclude palliative care patients with < 1 day length of stay?*
- *Do all hospitals use EHRs or are some still using paper records (note: the clinician group/practice measure for the*

hospital setting specifies use of EHRs only)?

- *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:

- Previous reliability testing included inter-rater reliability testing of hospital data for 20 patients. Reliability testing using hospice data was not previously conducted.

Describe any updates to testing

- Score-level testing for the hospice setting included split-half reliability, signal-to-noise, and stability analyses at the facility level.

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

- Clinician Group/Practice level (Palliative Care):
  - Reliability of the data elements was assessed using data obtained from the medical charts for 20 seriously ill patients without specialty palliative care admitted to an acute care hospital. Inter-rater reliability between two nurse abstractors was calculated. This is an appropriate method for demonstrating data element reliability.

- Facility-level (Hospice):
  - Reliability of the measure score was assessed using FY15 data from the Hospice Quality Reporting Program (n=3,992 hospice organizations and n=1,218,786 patient stays). Two types of reliability testing were conducted: the first via a signal-to-noise analysis and the second using a split-sample (or "test-retest") methodology. Both are appropriate methods for testing reliability. Stability analysis examined changes in provider's performance over time. 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

- Clinician Group/Practice level (Palliative Care)
  - For data element testing:
    - Developers report a single kappa value of 1.00.
    - 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 1.00 means that the raters agreed 100% 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.
    - NQF guidance indicates that data element testing should be conducted for all critical data elements, although at minimum, results about the numerator, denominator, and exclusions should be provided. Only a single kappa value was reported. Although not explicitly stated, we assume that the 20 charts that were abstracted included only seriously ill patients who received specialty palliative care in an acute hospital setting and whose length of stay was at least one day, and thus this kappa value applies to the numerator.

- Facility-level (Hospice):
  - For measure score testing:
    - Split-half analysis ICC=0.91.
      - The ICC reflects the percentage of variance in score results that is due to "true" or real variance between the hospices. A value of 0.7 is often regarded as a minimum acceptable reliability value.
    - Developers report a signal-to-noise ratio of 0.98.
      - A signal-to-noise analysis quantifies the amount of variation in a performance measure that is due to true differences between hospices (i.e., signal) as opposed to random measurement error (i.e., noise). Results will vary based on the amount of variation between the providers and the number of patients treated by each provider. This method results in a reliability statistic that ranges from 0 to 1 for each hospice. A value of 0 indicates that all variation is due to measurement error and a value of 1 indicates that all variation is due to real differences in hospices performance. A value of 0.7 often is regarded as a minimum acceptable reliability

Guidance from the Reliability Algorithm
Facility-level (hospice):
Precise specifications (Box 1) → Empirical reliability testing with measure as specified (Box 2) → Score-level testing (Box 4) → Appropriate method (Box 5) → High certainty that measure results are reliable (Box 6a) → High

Clinician-level (Palliative):
Precise specifications (Box 1) → Empirical reliability testing with measure as specified (Box 2) → Score-level testing not conducted (Box 4) → Data elements tested (Box 8) → Appropriate method for assessing data elements (Box 9) → High level of agreement between raters (Box 10a) →Moderate

Questions for the Committee:
- *Are there any concerns about accurately and consistently identifying the denominator and exclusions in the hospital setting?*
- *Is the test sample adequate to generalize for widespread implementation for palliative care? If not, is current testing sufficient until more data are available (assuming the developer is planning to do additional testing)?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

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**

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:
- In the prior review, the measure was tested for the palliative care (hospital) setting using face and construct validity.

Describe any updates to validity testing

- Additional validity testing of the measure score at the facility level of analysis (for the hospice setting) was conducted using FY15 data from the Hospice Quality Reporting System (HQRP).

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

Validity testing method:

- Facility-level (Hospice):
  - Using FY15 data from the Hospice Quality Reporting System (n=3,992 hospice organizations; n=1,218,786 patient stays), the developer conducted non-parametric Spearman rank correlation analysis between this measure and 5 other hospice quality measures. They hypothesized that agencies should perform similarly on assessment processes at hospice admission and expected the resulting correlations to be high. This is an appropriate method of score-level validation.
- Clinician Group/Practice level (Palliative Care)
  - Face validity was assessed by a group of nursing and physician stakeholders who were asked to comment on the validity, accuracy, and actionability of the measure. According to NQF guidance, the face validity of the measure score as a quality indicator may be adequate if it explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. It does appear that this testing conforms to NQF's requirements for face validity.
  - Construct validity was tested by comparing measure results for seriously ill patients seen in specialty interdisciplinary palliative care consultations (n=102) in one hospital to those who did not receive these services (n=460). Developers did not explain what they expected to find with this analysis and how it would demonstrate that the measure results reflect quality of care.

Validity testing results:

- Facility-level (Hospice):
  - Correlation results were positive and statistically significant, confirming the developer's hypothesis. However, the magnitude of the correlation was lower than they had expected. The developer states reasons for this may be due to clustering of scores, skewed distributions, and low score variability.

- Clinician Group/Practice (Palliative Care):
  - Face validity results from the stakeholder group indicated broad endorsement of the face validity of the measure.
  - Construct validity results found patients who received specialty palliative care were more likely to have documentation of their preferences for or against receiving life-sustaining treatments (91% vs 59%, $p<0.001$).

Questions for the Committee:

- *Are the test samples adequate to generalize for widespread implementation?*
- *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 examined the effect of removing the previous length of stay <7 days exclusion criterion for hospice patients. Using FY15 data from the HQRP, the developers found that the removal of the criterion increased the average size of the denominator per hospice organization and had little effect on the distribution

of the measure scores. The developers stated these finds supported the removal of the length of stay criterion as an exclusion criterion from the measure.
- The developer did not include information on exclusion analyses for palliative care patients.

Questions for the Committee:
- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the 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)*:

- Meaningful differences among hospices were examined using FY15 data from the Hospice Quality Reporting Program. The developers examined 95% confidence intervals to determine the proportion of hospices with results significantly different from the national hospice-level mean. Results indicated 33.2% of hospices had a score that was significantly different from the national mean. Hospices were more likely to report scores above the national mean than below the national mean (23.3% vs 9.9%, respectively).
- Developers compared measure score results for patients in one acute care hospital who received palliative care consults to those who did not receive palliative care consults. However, this analysis does not speak to whether the measure results reflect meaningful differences between clinician groups at different acute care hospitals.

Question for the Committee:
- *Does this measure identify meaningful differences about quality?*

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

2b7. Missing Data
- For hospices, the rate of missingness is low (0.01% - 0.02% at the patient-level). More than 95% of hospices had no missing information for three key data elements from the Hospice Item set that make up this measure.
- The developer does not provide any information on missing data specific to palliative care.

Guidance from the Validity Algorithm
Facility-level (Hospice)
Specifications consistent with evidence (Box 1) → potential threats to validity assessed (Box2) → empirical validity testing (Box 3) → validity testing of the measure score (Box 6) → method appropriate (Box7) → moderate certainty (Box 8b) → Moderate

Clinician group/practice level (Palliative Care)
Specifications consistent with evidence (Box 1) → potential threats to validity only somewhat assessed (Box2) → empirical validity testing (Box 3) → validity testing of the measure score (Box 6) → method not well described (Box7) → face validity systematically assessed (Box 4) → results indicate substantial agreement as to validity (Box 5) → Moderate (assuming no concerns around exclusions, missing data, or meaningful differences).

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

Committee pre-evaluation comments
Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a, 2b.
Reliability – Specifications
* Data elements clearly defined. The measure documentation notes this algorithm calls for documentation of a discussion of preference for life sustaining treatments, rather than documentation of the actual preferences, as is described in the numerator statement.
* difference between documentation of discussion vs documentation of preferences
* The data elements and the algorithm are well described. The one exception is that the denominator for the palliative care data seems to be defined somewhat differently in different places in the report. Usually, it is defined as "seriously

ill patients who received specialty palliative care" but in validity section the sample seems to be described as patients with a "palliative care consultation" which does not necessarily imply that those patients are receiving palliative care.
* The calculation algorithim calls for a “discussion of preferences” for life sustaining treatments but the clinical records used to calculate performance do not necessarily demonstrate that a discussion took place, when it took place, and whether it was revisited.

Question: would other documentation besides a POLST form and other entries in the medical record be included in the data consistently?

Validity – Specifications
* Specifications consistent with the evidence.
* as above, may be a concern that evidence presented is for evidence of preferences vs discussion of preferences
* The specifications do seem to be consistent with the evidence and the description
* The measure underwent two types of reliability testing: clinician group/practice level (palliative care) and facility-level (hospice). The clinician group/practice testing indicates a perfect agreement between two raters/abstractors that is not explained by chance alone. It is important to keep in mind that the developers only provided a single kappa value of 1.00. The results of the facility-level testing indicates a high rating. The split-half analysis ICC is 0.91, which is greater than the minimum acceptable reliability value. It is reported that the measure has a signal-to-noise ratio of 0.98, which is also greater than the minimum acceptable reliability.

Reliability – Testing
* Single kappa value of 1.0 for the 20 hospital charts used. Algorithms: high for facility level and moderate for clinician model.
* sufficient evidence of reliability
* Both hospice and palliative care data were tested. There does not seem to be any data element testing on the denominator for the palliative care data. The reliability data in general are well done and consistent with NQF standards. The data sample for palliative care does appear small (20). Even though there was perfect interrater reliability, the developers should address why they consider this sufficient.

The developers are proposing to remove the LOS requirement for hospice patients. Given that their results seem to indicate that this increases the denominator without seemingly changing the overall result, this request seems reasonable.
* Yes, the measure was validity tested with an adequate scope to generalize for widespread implementation and with an appropriate method. The measure was tested on both the facility-level and the clinician group/practice-level. On the facility-level, the correlation results were positive and statistically significant. The results of the clinician group/practice testing indicate broad endorsement of the face validity of the measure. In addition, the validity results found that patients who received specialty palliative care were more likely to have documentation of their preferences for or against receiving life-sustaining treatments.

However, as one commenter said, that is like looking for a dollar bill in the light ---- given that they are enrolled in hospice or specialty palliative care, the measure is limited to the group most likely to have documents their care preferences.

Question: does the denominator need to be broader to be meaningful?

Validity Testing
* Validity testing with results to generalize for widespread implementation.
* sufficient evidence of validity
* The hospice data was tested for score level validation and those correlations were positive. The palliative care data was tested for face and construct validity. Both of these results were positive.
* Yes, the measure was validity tested with an adequate scope to generalize for widespread implementation and with an appropriate method. The measure was tested on both the facility-level and the clinician group/practice-level. On the facility-level, the correlation results were positive and statistically significant. The results of the clinician group/practice testing indicate broad endorsement of the face validity of the measure. In addition, the validity results found that patients who received specialty palliative care were more likely to have documentation of their preferences for or against receiving life-sustaining treatments.

However, as one commenter said, that is like looking for a dollar bill in the light ---- given that they are enrolled in

hospice or specialty palliative care, the measure is limited to the group most likely to have documents their care preferences.

Question: does the denominator need to be broader to be meaningful?

Threats to Validity
* The developers stated these finds supported the removal of the length of stay criterion as an exclusion criterion from the measure. I am in agreement with this.
Overall, we found that only a very small number of admission records for the eligible stays did not include data, i.e. coded as dash, on treatment preferences items F2000B, F2100B, and F2200B. And a vast majority of hospices did not have any missing data for these items. These results indicate that missing data for these items should not have a negative impact on the validity of the QM.
* no -- measure changed since original to reduce excluded population with no threat to validity
* There was no discussion of missing data or exclusion or disparities analysis for the palliative care patients and at least some of the testing seems to have been done in only one hospital thus limiting conclusions about generalizability and conclusions about meaningful differences. The elimination of the LOS requirement for hospice seems appropriate as stated above. There do not appear to be any groups inappropriately excluded. Again, overall, the hospice data seems much stronger than the palliative care data especially based on sample size and testing of validity although both seem to rate as moderate.
* Note that the developers intend to eliminate the previous <7 days length of stay in hospice exclusion. The elimination of the exclusion will increase the average size of the denominator per hospice organization and have little effect of the distribution of measure scores. The change makes sense because the reason for the exclusion was that for patients in hospice less than seven days it was thought that it would be difficult to meet the various care benchmarks required, but obtaining the information on what a patient's preferences are should happen before the patient is enrolled in hospice.

No risk adjustment is applied to this measure.

For facility level hospice, this measure may be topped out. The mean is 98%, the median is 100%.
For palliative care at the clinician/group practice level, the data is not available yet but will be next year.

Missing data does not constitute a threat to the validity of the measure re hospice. The rate of missingness is low (0.01%-0.02%). Over 95% of hospices had no missing information for three key data elements from the Hospice Item set that make up this measure.

For palliative care, the developer did not provide information on missing data.

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 are routinely collected and are incorporated in hospice providers' electronic clinical documentation system (i.e., electronic health record (EHR)). Data collection for all measures is achieved through use of the Hospice Item Set (HIS). It is unclear if the HIS is automatically populated by the EHR.
- The developer does not provide any information on potential or actual implementation challenges for palliative care.
- The developer reports cost estimates are not available. According to the Guidance Manual for Completion of the Hospice Item Set, completing an entire response for any one admission takes 20 minutes (though seven performance measures, of which this measure is only one, are collected as part of this process).
- In the 2012 endorsement evaluation, the Palliative and End-of-Life Care Steering Committee noted that a substantial data collection effort may be needed to abstract data if electronic data are not available. They also expressed concern that the documentation may not be standardized, making it somewhat challenging to extract reliably.

Questions for the Committee:

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

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

Committee pre-evaluation comments

Criteria 3: Feasibility

* No concerns about the feasibility.
* for electronic data systems, no barriers to collection
* The developers report that all data elements are available digitally in discrete fields. The missing data analysis for the hospice data suggests that there is very little missing data. This analysis would suggest that this data collection strategy could be put into effect in the hospital setting. There is no such analysis for the palliative care data. There is no discussion of how this data is used in care delivery.
* The required data elements are routinely generated and used during care delivery. The required data elements are available in electronic health records or other electronic sources. There are concerns that the documentation may not be standardized, which may make it somewhat challenging to extract reliably.

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

Publicly reported? ☐ **Yes** ☒ **No**

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

OR

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

Accountability program details:

- This measure is included in the Hospice Quality Reporting Program (HQRP). Medicare-certified hospices are required to submit a Hospice Item Set (HIS)-Admission record and an HIS-Discharge record for each patient admission on or after July 1st, 2015. In FY2015, 3,992 hospice organizations provided measure data for 1,218,786 patient stays. The distribution of reporting hospice agencies by geographic area is provided.

Improvement results:

- Longitudinal data for this measure 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 2016, the Measures Application Partnership Post-Acute Care and Long-Term Care workgroup

supported the continued development of a composite measure of Hospice Item Set measures, which includes this measure.

- In February 2013, the Measure Application Partnership supported the measure for inclusion in the Hospice Quality Reporting Program and the Physician Quality Reporting System (PQRS) to address a core concept (i.e., pain assessment) not addressed in the programs' measure sets. Public comments from American Academy of Hospice and Palliative Medicine (AAHMP), Center to Advance Palliative Care (CAPC), and National Coalition for Hospice and Palliative Care (NCHPC) supported the inclusion of this measure in the PQRS. CAPC and NCHPC suggested that the measure should be expanded beyond hospice patients. The measure is not currently included in the PQRS program, but it is included in the Hospice Quality Reporting Program.
- In 2014, the AAHMP and the Hospice and Palliative Nurses Association published a list of "Top Ten Measures That Matter". This measure is Measure #8 of this list .

***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**

NOTE: Usability and Use is low for the hospice setting if in fact there is no opportunity for improvement.

Committee pre-evaluation comments

Criteria 4: Usability and Use

* No current use of public reporting for measure. Accountability program: • This measure is included in the Hospice Quality Reporting Program (HQRP). Medicare-certified hospices are required to submit a Hospice Item Set (HIS)-Admission record and an HIS-Discharge record for each patient admission on or after July 1st, 2015. In FY2015, 3,992 hospice organizations provided measure data for 1,218,786 patient stays. The distribution of reporting hospice agencies by geographic area is provided
* in hospice, more uniformly collected
* There are two potentially competing measures.

0326: Advance Care Plan

1626: 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

It would seem that these measures could be harmonized with the current measure if so desired.

The developers report that they do not yet have sufficient longitudinal data to report on improvement of use. The developers report no unexpected findings or unintended consequences.

The measure is included on the top 10 Measure What Matters list by the AAHPM and HPNA. This measure does clearly have the potential to further ability to meet patients' needs and provide the care which patients want has been shown to have significant positive consequences for overall quality of care. It would be anticipated that the discussions resulting from this measure would raise patient satisfaction. Given that the literature suggests that most patients want to have this discussion and do not find it a burden, it would seem that the benefits of this measure outweigh any potential unintended consequences.

* The measure is not currently publicly reported; however, it is used in an accountability program. It is included in the Hospice Quality Reporting Program (HQRP). The developer did not report any 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.

Question: whether the data captured is too limited and should include rate of discussions and documentation of patient's preferences for life sustaining treatments for all patients in hospital or long term care setting from whom it would be likely that they would die within one year --- rather than limiting it to patients enrolled in hospice or a specialty palliative care in-patient program.

This measure should/could be harmonized with 0326 and 1626.

Criterion 5: Related and Competing Measures

Competing measures

- 0326: Advance Care Plan *[individual and clinician group/practice-level measure in various settings including hospital and hospice]*
- 1626: 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 *[facility/health plan/integrated delivery system-level process measure in hospital setting*]

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: Palliative care should be initiated with diagnosis and treatment. Measure specifications seem to focus on the benefits of early communication of treatment preferences, therefore recommending that the measure focus on communication of this within a time frame following diagnosis and less about timing related to enrollment in hospice

Without these changes, recommend making the measure specific to hospice care, not palliative care or both in the same measure. Palliative care is not restricted to inpatient treatment.

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.

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 1641	**NQF Project:** Palliative Care and End-of-Life Care

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)***:** The Treatment Preferences quality measure addresses a key process -- eliciting and documenting patient treatment preferences -- with evidence linking it to outcomes of patient autonomy and control over treatments, patient and family satisfaction with care, improved transitions to hospice and palliative care, and reduced emotional distress for surviving family. There is broad legal and ethical consensus that the treatment of seriously ill and dying patients should be guided by their values and preferences regarding life-sustaining treatments.(1) Failure to elicit and communicate these preferences can result in the intermediate outcome of treatment which is contradictory to patients' and families' values, in turn decreasing patient and family satisfaction. If patients die without adequate opportunity for treatment guided by their own preferences, families report markedly greater emotional distress following the death. An early systematic review found evidence for poor quality communication, but limited evidence for its relationship to outcomes.(2) However, an updated systematic review by the same investigators found moderate evidence to support multicomponent interventions to increase advance directives and for care planning through engaging values, with improved rates of hospice use, reduced ICU days, and enhanced quality of patient-provider communication.(3) More targeted trials that enhance the frequency and quality of communication have positive effects, including treatment consistent with preferences, reduced family distress, improved comprehension, and decreased the use of intensive treatments without adverse effects on mortality.(4,5) In addition to this direct evidence, some indirect evidence supports the link between enhanced communication about treatment preferences in palliative care interventions, and improved patient and family outcomes. One systematic review of specialized palliative care, covering heterogeneous complex clinical interventions which include communication of treatment preferences, found a small number of interventions

resulted in improved quality of life and family satisfaction with care, but concluded that future trials need improved methodologic rigor.(6) Several subsequent palliative care clinical trials and 2 observational studies have added evidence that these complex interventions, which include enhanced clinical communication about treatment preferences as a key component, are associated with enhanced attention to patient autonomy, improved satisfaction with care, less high cost life-sustaining treatment use, and these benefits accrue without adverse effects on mortality. (7,8,9,10,11,12)

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

Clinical Practice Guideline

Selected individual studies (rather than entire body of evidence)

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)*:

The central topic of communication of patient care preferences is support by direct and indirect evidence. Study populations are typically hospitalized or in intensive care, and have varied diagnoses in advanced or incurable stages. The strongest direct evidence for improved outcomes comes from systematic communication interventions (ref Scheunemann), and the outcomes most clearly impacted are treatment choices, use of life-sustaining treatment, and satisfaction with care. Although discussion of treatment choices and patient autonomy are clearly elements of hospice care, the research evidence base generally does not not address hospice patients, whose broader treatment preferences must be addressed in advance of hospice care in order to elect that option.

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)*: 1. Systematic review (Lorenz, 2004) including 4 earlier systematic reviews, 21 individual trials, and 22 observational cohort studies.

2. Systematic review (Lorenz, 2008) including 9 systematic reviews, 32 individual intervention studies
3. Systematic review (Zimmerman, 2008) including 22 clinical trials of specialized palliative care interventions
4. Systematic review (Parker, 2007) including 46 studies
5. Systematic review (Scheunemann, 2011) including 21 studies of 16 distinct intervention trials
6. Individual additional studies - 6 (Gade, Temel, Casarett, Engelhardt, Bakitas, Wright)

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):* The overall quality of this body of evidence is moderate. The strongest and most direct evidence comes from a small number of rigorously designed randomized trials testing communication interventions with seriously ill hospitalized patients and their families (ref Scheunemann). In addition, several

recent well-designed clinical trials (ref Temel, Gade, Bakitas) have tested complex palliative care interventions which add meaningful but more indirect evidence, since they include communication of patient treatment preferences as a key component. These clinical trials have been adequately powered.

1c.7 Consistency of Results <u>across Studies</u> *(Summarize the consistency of the magnitude and direction of the effect):* Studies of communication interventions, including more complex palliative care service delivery, have consistently shown improvements in treatment consistent with preferences, reduced family distress, improved comprehension, and decreased the use of intensive treatments without adverse effects on mortality. Results have been consistent, and effect sizes modest but clinically as well as statistically significant.

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

The net benefit from interventions to enhance communication of patient treatment preferences is positive. Several studies have shown no adverse effect on mortality, and one study of a complex palliative care intervention (ref Temel) has shown survival benefit. Careful review of communication studies for adverse emotional effects have found no evidence of significant harms. (ref Emanuel)

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: Not graded

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)*:

OTHER GUIDELINES:

1. Harle I, Johnston J, MacKay J et al. Advance Care Planning with Cancer Patients: Guideline Recommendations. Toronto (ON): Cancer Care Ontario (CCO); 2008 Jan 28, 37 p. http://www.guideline.gov/content.aspx?id=12499; viewed May 2011.

SYSTEMATIC REVIEWS:

2. Lorenz KA, Lynn J, Dy SM et al. Evidence for improving palliative care at the end of life: a systematic review. Ann Intern Med 2008; 148:147-159.

3. Lorenz KA, Lynn J et al. End-of-life care and outcomes. AHRQ Publication No. 05-E004-2, December 2004.

4. Parker SM, Clayton JM, Hancock K et al. A systematic review of prognostic / end of life communication with adults in the advanced stages of a life-limiting illness: patient / caregiver preferences for the content, style and timing of information. J Pain Symptom Manage 2007; 34:81-93

5. Scheunemann LP, McDevitt M, Carson SS, Hanson LC. Randomized, controlled trials of interventions to improve communication in intensive care: a systematic review. Chest 2011; 139:543-554.

6. Zimmerman C, Riechelmann R, Krzyzanowska M et al. Effectiveness of specialized palliative care: a systematic review. JAMA 2008; 299:1698-1709.

7. Michigan Quality Improvement Consortium. Advance care planning. Southfield (MI): Michigan Quality Improvement Consortium; 2014 Jan. 1 p.

ADDITIONAL INDIVIDUAL STUDIES

7. Bakitas M, Lyons KD, Hegel MT et al. Effects of a palliative care intervention on clinical outcomes in patients with advanced cancer: the Project ENABLE II randomized controlled trial. JAMA 2009; 302:741-749.

8. Casarett D, Pickard A, Bailey FA et al. Do palliative consultations improve patient outcomes? J Am Geriatr Soc 2008; 56:593-599.

9. Temel JS, Greer JA, Muzikansky A et al. Early palliative care for patients with metastatic non-small-cell lung cancer. N Engl J Med 2010; 363:733-742.

10. Gade G, Venohr I, Conner D et al. Impact of an inpatient palliative care team: a randomized controlled trial. J Palliat Med 2008; 11:180-190.

11. Engelhardt JB, McClive-Reed KP, Toseland RW et al. Effects of a program for coordinated care of advanced illness on patients, surrogates, and healthcare costs: a randomized trial. Am J Manag Care 2006; 12:93-100.

12. Wright AA, Zhang B, Ray A et al. Associations between end-of-life discussions, patient mental health, medical care near death, and caregiver bereavement adjustment. JAMA 2008; 300:1665-1673.

13. Emanuel EJ, Faircloth DL, Wolfe P, Emanuel LL. Talking with terminally ill patients and their caregivers about death, dying and bereavement: is it stressful? is it helpful? Arch Intern Med 2004; 164:1999-2004.

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

Health Care Guideline: Palliative Care for Adults

Guideline 9. Ethical and Legal Aspects of Care: Clinicians should initiate or facilitate advance care planning for all adult patients and their families with regular review as the patient's condition changes.

Guideline 10: Develop or Revise Palliative Care Plan and Establish Goals of Care Through the Process of Shared Decision-Making.

Guideline 8.1 The patient´s goals, preferences and choices are respected within the limits of applicable state and federal law, within current accepted standards of medical care, and form the basis for the plan of care.

Criteria:

•The interdisciplinary team includes professionals with knowledge and skill in ethical, legal, and regulatory aspects of medical decision-making.

•The patient or surrogate´s expressed wishes, in collaboration with the family and the interdisciplinary team, form the basis for the care plan.

•The adult patient with decisional capacity determines the level of involvement of the family in decision-making and communication about the care plan.

•Evidence of patient preferences for care is routinely sought and documented in the medical record. Failure to honor these preferences is documented and addressed by the team.

•Among minors with decision-making capacity, the child´s views and preferences for medical care, including assent for treatment, should be documented and given appropriate weight in decision-making. When the child´s wishes differ from those of the adult decision-maker, appropriate professional staff members are available to assist the child.

•The palliative care program promotes advance care planning in order to understand and communicate the patient´s or an appropriate surrogate´s preferences for care across the health care continuum.

•When patients are unable to communicate, the palliative care program seeks to identify advance care directives, evidence of previously expressed wishes, values, and preferences, and the appropriate surrogate decision-makers. The team must advocate the observance of previously expressed wishes of the patient or surrogate when necessary.

•Assistance is provided to surrogate decision-makers on the legal and ethical bases for surrogate decision-making, including honoring the patient´s known preferences, substituted judgment, and best-interest criteria.

1c.17 Clinical Practice Guideline Citation: McCusker M, Ceronsky L, Crone C, Epstein H, Greene B, Halvorson J, Kephart K, Mallen E, Nosan B, Rohr M, Rosenberg E, Ruff R, Schlecht K, Setterlund L. Institute for Clinical Systems Improvement. Palliative Care for Adults. Updated November 2013.p.26

1c.18 National Guideline Clearinghouse or other URL: Citation: National Guideline Clearinghouse http://www.guideline.gov/content.aspx?id=47629

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? Yes

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: Institute for Clinical Systems Improvement (ICSI)

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

1c.22 If other, identify and describe the grading scale with definitions:

1c.23 Grade Assigned to the Recommendation: Low Quality Evidence, Strong Recommendation

1c.24 Rationale for Using this Guideline Over Others: We chose to use the Palliative Care for Adults guideline from the Institute for Clinical Systems Improvement (ICSI) because it is evidenced based, uses GRADE methodology, and is included in the National Guideline Clearinghouse. In addition, the National Consensus Project for Quality Palliative Care was the first United States national guidelines development project for palliative care quality, inclusive of hospice care. This set of guidelines, along with 38 preferred practices, has been rigorously reviewed and endorsed by the National Quality Forum. Although specific investigative groups and specialty organizations have published other guidelines in pain management or hospice or palliative care practice for specific settings or populations, none have been as comprehensive or comprehensively debated , peer reviewed, or NQF endorsed.

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: Moderate **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 #: 1641
De.2. **Measure Title:** Hospice and Palliative Care – Treatment Preferences
Co.1.1. **Measure Steward:** University of North Carolina-Chapel Hill
De.3. **Brief Description of Measure:** Percentage of patients with chart documentation of preferences for life sustaining treatments.
1b.1. **Developer Rationale:** Seriously ill and dying patients who are given the opportunity to express life-sustaining treatment preferences are more likely to receive care consistent with their values, and patient and family satisfaction outcomes improve. Patients and physicians alike hesitate to initiate discussions, while acknowledging their value and desirability. Use of the Treatment Preferences quality measure will improve attention to this important practice, in order to enhance patient autonomy, facilitate patient-centered decision-making, and communicate patient preferences via documentation to other treating providers.

S.4. **Numerator Statement:** Patients whose medical record includes documentation of life sustaining preferences
S.7. **Denominator Statement:** Seriously ill patients enrolled in hospice OR receiving specialty palliative care in an acute hospital setting.
S.10. **Denominator Exclusions:** Patients with length of stay < 1 day in hospice or palliative care

De.1. **Measure Type:** Process
S.23. **Data Source:** Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record
S.26. **Level of Analysis:** Clinician : Group/Practice, 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?** N/A

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
1641_Evidence_MSF5.0_Data.doc,1641_Evidence_3.17.16.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)*
Seriously ill and dying patients who are given the opportunity to express life-sustaining treatment preferences are more likely to receive care consistent with their values, and patient and family satisfaction outcomes improve. Patients and physicians alike hesitate to initiate discussions, while acknowledging their value and desirability. Use of the Treatment Preferences quality measure will improve attention to this important practice, in order to enhance patient autonomy, facilitate patient-centered decision-making, and communicate patient preferences via documentation to other treating providers.

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.*
Hospice: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1641 Hospice and Palliative Care Treatment Preferences.

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.

The mean score for this QM was 98.0% with a range from 0% to 100%, the median was 100%, the interquartile range was 1.5, and the standard deviation was 6.3. For this QM, 53.5% of hospices had perfect scores and 4.6% of hospices scored below 90%.

Scores by decile:
10th percentile 95.0%
25th percentile 98.5%
Median 100%
75th percentile 100%
90th percentile 100%

Palliative Care:

This submission to the Palliative and End-of-Life Care project updates hospice setting data for NQF #s 1634, 1637, 1638, 1639, 1641, 1647. We are currently in the process of updating palliative care data by collecting and analyzing data in multiple non-hospice settings but final analyses are not available for this submission cycle.(1) Data comes from two sources -- a multi-site study of quality of care in palliative care (R18HS022763 Implementing Best Practice in Palliative Care, PI Johnson) and (CMS Health Care Innovation Award: Increasing patient and system value with community based palliative care, PI Bull / Four Seasons Compassion for Life). We anticipate these data will become available for NQF review next year. This data will allow further updates to the evidence base for non-hospice palliative care beyond what is currently submitted.

(1) Kamal AH, Bull J, Ritchie CS, Kutner JS, Hanson LC, Friedman F, Taylor DH Jr; AAHPM Research Committee Writing Group. Adherence to Measuring What Matters measures using point-of-care data collection across settings. J Pain Symptom Manage 2016; 51:497-503.

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.
Poor communication about patient preferences has been identified as a major quality concern in palliative and end-of-life care since an early, comprehensive Institute of Medicine report.(1) The SUPPORT Study found marked discrepancies between patient report of treatment preferences and provider awareness of or use of these preferences to guide treatment.(2) Patients and families prioritize communication with providers and control over treatment choices when faced with serious or life-threatening illness.(3) However, physicians and other providers fail to open the door to these discussions at critical time points in illness progression.(4) A recent systematic review of communication research found a consistent discrepancy between the quality and content of communication providers believed they provided, and the quality and content of communication experienced by seriously ill patients and their families. (5)

1. Field MJ, Cassell CK eds. Approaching Death: Improving Care at the End of Life. Washington, DC: National Academy Press, 1997.
2. SUPPORT Principal Investigators. A controlled trial to improve care for seriously ill hospitalized patients: the Study to Understand Prognosis and Preferences for Outcomes and Risks of Treatments (SUPPORT). JAMA 1995; 274:1591-1598.
3. Steinhauser KE, Christakis NA, Clipp EC et al. Preparing for the end of life: preferences of patients, families, physicians and other care providers. J Pain Symptom Manage 2001; 22:727-737.
4. Gysels M, Richardson A, Higginson I. Communication training for health professionals who care for patients with cancer: a systematic review of effectiveness. Support Care Cancer 2004; 12:692-700.
5. Hancock K, Clayton JM, Parker SM et al. Discrepant perceptions of end-of-life communication: a systematic review. J Pain Symptom Manage 2007; 34: 190-200.

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.*
Hospice:

Racial disparity analysis. We conducted racial and ethnic disparity analyses at both the patient-stay and hospice levels. At the patient-stay level, we compared the percentage of patients with documentation of treatment preferences among different racial and ethnic groups. Patients were grouped by their racial and ethnic identification as follows: white non-Hispanic, black non-Hispanic, other non-Hispanic, or Hispanic. Other non-Hispanic includes patients who identify as American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, or who identify as more than one race. A Chi-square test was performed to determine if there were any statistically significant differences in documentation of treatment preferences between groups. The lowest rate of documentation of treatment preferences was found for patients with racial and ethnic group missing (94.5%), and the highest rate was among patients identifying as White non-Hispanic (98.0%). Differences in the rate of treatment preferences by racial identification were found to be statistically significant ($p<0.001$).

Analyses at the hospice level examined the differences in this measure across two groups: hospices with proportions of non-white patients that are greater than or equal to the national median proportion (11.9%), and hospices with fewer non-white patients than the national median. For this analysis, white non-Hispanic patients were included in the white group, and all other racial and ethnic identifications, as described above, were grouped as non-white. We ran a Wilcoxon-Mann-Whitney test for statistical dependence between group and QM score. The results showed that the QM score was significantly different between the two groups of hospices (98.4% compared to 97.6%, $p=0.005$). Although statistically significant results were found at both the patient and hospice level, actual differences in screening rates do not seem to be clinically substantial.

Gender disparity analysis. We performed both the patient stay- and hospice-level analyses on gender disparity. At the patient-stay level, we compared the percentage of patients with documentation of treatment preferences between female and male patients. A Chi-square test was performed to determine if there were any statistically significant differences in treatment preferences between groups. We found a slightly lower rate of documentation of treatment preferences for female patients (97.6%) than for male patients (97.7%). Differences in the rate of treatment preferences by gender were statistically significant ($p=0.0029$). At the hospice level, we examined the differences in this measure across hospices with proportions of female patients that are greater than or equal to the national median proportion (55.2%). We ran a Wilcoxon-Mann-Whitney test for statistical dependence between group and QM score. The results showed that the QM score was not statistically significantly different between the two groups of hospices split by median proportion of female patients (97.9% compared to 98.0%, $p=0.70$). Although statistically significant results were found at both the patient level, actual differences in documentation rates do not seem to be clinically substantial.

Socioeconomic disparity analysis. We performed socioeconomic disparity analyses at both the patient-stay and hospice levels using the same methods described in previous disparity analyses. Medicaid status was used as a proxy measure of low socioeconomic status. Both Medicaid and non-Medicaid patient groups show a comparable rate of documentation of treatment preferences (97.4%). The highest rate of documentation of treatment preferences was seen for patients with Medicaid status missing (98.4%). Although statistically significant ($p<0.001$), actual differences in documentation rates do not seem to be clinically substantial. At the hospice level, the results showed that the QM score was significantly different between the two groups of hospices split by median proportion of Medicaid patients, 21.5% (97.4% compared to 98.5%, $p<0.001$). The significant findings at the hospice level indicate that hospices with a smaller proportion of Medicaid patients are less likely to provide treatment preferences to patients at admission. A potential caveat of this analysis is that the data used for this analysis may not be a reliable indicator of a patient's Medicaid eligibility status. This data is missing for nearly a quarter of patients. In addition, some of the Medicaid numbers submitted do not appear to be legitimate Medicaid numbers. However, the data used in this analysis is the only national patient-level assessment data that is currently available in hospices. We will update this analysis as more-accurate data sources are available and

accessible.

Rural-urban disparity analysis. We compared the QM score between rural and urban hospices using a Wilcoxon-Mann-Whitney test. The results showed that the QM score was significantly different between rural and urban hospices (98.8% compared to 97.7%, $p<0.001$). Although statistically significant results were found, actual differences in screening rates between rural and urban hospices do not seem to be clinically substantial.

Palliative Care: Disparities data not available.

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.**
In advanced incurable illness, treatment options range from life-sustaining and disease modifying interventions to control of acute exacerbations to hospice care. African Americans with advanced cancer less often access treatment through clinical trials, palliative care for pain management, or hospice.(1,2) They less often prepare advance directives, including Health Care Powers of Attorney that can facilitate family advocacy during illness.(3,4,5,6,7,8) If these choices are fully informed expressions of values, they should be supported. However, African Americans desire more information on treatment options and are less likely to have discussions with their physicians, indicating that communication and information access serve as barriers to optimal care.(9,10,11,12)

1.Smith AK, Earle CC, McCarthy EP. Racial and ethnic differences in end of life care in fee for service Medicare beneficiaries with advanced cancer. J Am Geriatr Soc 2009; 57:153-158.
2. Cintron A, Morrison RS. Pain and ethnicity in the United States: a systematic review. J Pall Med 2006; 9:1454-1473.
3. Hanson LC, Rodgman E. The use of living wills at the end of life: a national study. Arch Intern Med 1996; 156:1018-22.
4. Murphy ST, Palmer JM, Azen S, Frank G, Michel V, Blackhall LJ. Ethnicity and advance care directives. J Law Med Ethics 1996; 24:108-17.
5. Morrison RS, Zayas LH, Mulvihill M, Baskin SA, Meier DE. Barriers to completion of health care proxy forms: a qualitative analysis of ethnic differences. J Clin Ethics 1998; 9:118-26.
6. Tilden VP, Tolle SW, DrachLL, Perrin NA. Out-of-hospital death: advance care planning, decedent symptoms and caregiver burden. JAGS 2004; 52:532-39.
7. Kiely DK, Mitchell SL, Marlow A, Murphy KM, Morris JN. Racial and state differences in the designation of advance directives in nursing home residents. JAGS 2001; 49:1346-52.
8. Hopp FP, Duffy SA. Racial variations in end of life care. J Am Geriatr Soc 2000; 48:658-663.
9. McKinley ED, Garrett JM, Evans AT, Danis M. Differences in end-of-life decision making among black and white ambulatory patients. J Gen Intern Med 1996; 11:651-56.
10. Borum ML, Lynn J, Zhong Z. The effects of patient race on outcomes in seriously ill patients in SUPPORT: an overview of economic impact, medical intervention, and end-of-life decisions. JAGS 2000; 48:S194-S198.
11. Haas JS, Weissman JS, Cleary PD, Goldberg J, Gatsonis Cm, Seage GR, Fowler FJ, Massagli MP, Makadon HJ, Epstein AM. Discussion of preferences for life-sustaining care by persons with AIDS: predictors of failure in patient-physician communication. Arch Intern Med 1993; 153:1241-48.
12. Born W, Greiner KA, Sylvia E, Butler J, Ahluwalia JS. Knowledge, attitudes and beliefs about end-of-life care among inner-city African Americans and Latinos. J Pall Med 2004; 7:247-256.

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**
Affects large numbers, Patient/societal consequences of poor quality, Severity of illness
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.**
The Hospice and Palliative Care - Treatment Preferences measure addresses patient autonomy for patients with high severity of illness and risk of death, including seriously and incurably ill patients enrolled in hospice or hospital-based palliative care. The

National Priorities Partnership has identified palliative and end-of-life care as one of its national priorities. A goal of this priority is to ensure that all patients with life-limiting illness have the right to express preferences that guide use of invasive or life-sustaining forms of treatment.(1) The affected populations are large; in 2014, 1.66 million people with life-limiting illness received hospice care.(2) In 2013, 67% of US hospitals with 50 or more beds had some form of palliative care service, up from 58.5% in 2008, and national trends show steady expansion of these services.(3)

Patients and family caregivers rate control over treatment decisions as a high priority when living with serious and life-limiting illnesses. (4) From a recent systematic review of clinical trials, moderate evidence supports multicomponent interventions to increase advance directives, and "care planning through engaging values, involving skilled facilitators, and focusing on key decision makers." These studies found improved outcomes of patient-physician communication, improved satisfaction with care, and increased hospice enrollment.(5) The more recently published Coping with Cancer Study, a prospective observational study of over 300 patients with advanced cancer, found that communication of patient treatment preferences was associated with use of treatments honoring those preferences and wish lesser use of aggressive, high-cost treatments.(6,7)The Hospice and Palliative Care - Treatment Preferences measure addresses patient autonomy for patients with high severity of illness and risk of death, including seriously and incurably ill patients enrolled in hospice or hospital-based palliative care. The National Priorities Partnership has identified palliative and end-of-life care as one of its national priorities. A goal of this priority is to ensure that all patients with life-limiting illness have the right to express preferences that guide use of invasive or life-sustaining forms of treatment.(1) The affected populations are large; in 2009, 1.56 million people with life-limiting illness received hospice care.(2) In 2008, 58.5% of US hospitals with 50 or more beds had some form of palliative care service, and national trends show steady expansion of these services.(3)
Patients and family caregivers rate control over treatment decisions as a high priority when living with serious and life-limiting illnesses. (4) From a recent systematic review of clinical trials, moderate evidence supports multicomponent interventions to increase advance directives, and "care planning through engaging values, involving skilled facilitators, and focusing on key decision makers." These studies found improved outcomes of patient-physician communication, improved satisfaction with care, and increased hospice enrollment.(5) The more recently published Coping with Cancer Study, a prospective observational study of over 300 patients with advanced cancer, found that communication of patient treatment preferences was associated with use of treatments honoring those preferences and wish lesser use of aggressive, high-cost treatments.(6,7)

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

1. http://www.nationalprioritiespartnership.org/PriorityDetails.aspx?id=608
2. NHPCO Facts and figures: hospice care in America 2015 edition
http://www.nhpco.org/hospice-statistics-research-press-room/facts-hospice-and-palliative-care
3. Dumanovsky T, Augustin R, Rogers M, Lettang K, Morrison RS. The growth of palliative care in U.S. hospitals: a status report. J Pall Med. 2016; 19(1): 8-15
4.Singer PA, Martin DK, Kelner M. Quality end-of-life care: patients´ perspective. JAMA 1999; 281: 163-168.
5. Lorenz KA, Lynn J, Dy SM et al. Evidence for improving palliative care at the end of life: a systematic review. Ann Intern Med 2008: 148:147-159.
6. Wright AA, Mack JW, Kritek PA, Balboni TA, Massaro AF, Matulonis UA, Block SD, Prigerson HG. Influence of patients' preferences and treatment site on cancer patients' end of life care. Cancer. 2010 Oct 1;116(19):4656-63.
7. Wright AA, Zhang B, Ray A et al. Associations between end-of-life discussions, patient mental health, medical care near death, and caregiver bereavement adjustment. JAMA 2008; 300:1665-1673.

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):*

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.)*
PEACE Hospice and Palliative Care Quality Measures: http://www.med.unc.edu/pcare/resources/PEACE-Quality-Measures

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.
We would like to propose the removal of the less than 7 day length of stay (LOS) denominator exclusion for hospice patients.

Background: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1641 - Hospice and Palliative Care Treatment Preferences.

Six of the seven QMs exclude patient stays that are less than 7 days from the measure denominator. When the length of stay (LOS) is too short, hospices may not have enough time to complete all the clinically recommended care processes. Thus, at the time the measures were developed, technical experts recommended that short patient stays be excluded from those measure denominators for assessing quality of care in hospices. However, no national data regarding the implications of the LOS exclusion was available to the Technical Expert Panel (TEP) at that time. Under contract to the Centers for Medicare and Medicaid Services (CMS), the Research Triangle Institute (RTI) performed descriptive analyses on 5 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 1,218,786 patient stays.

Rationale for inclusion of all hospice patients regardless of LOS: At the patient level, approximately 40% of patients were excluded based on the LOS < 7 days exclusion, thus omitting this essential quality of care standard for many seriously ill patients. Further, excluding short-stay patients omits measures of quality of care for a subset of the sickest patients entering hospice care, and those who arguably may be in greatest need of rapid symptom assessment and treatment. In addition, the original rationale for this exclusion -- allowing time for hospice providers to complete this care process -- does not appear to be necessary. Our analyses show that a large portion of treatment preference discussions were performed on day 1 of admission to hospice, demonstrating a normative standard of care includes prompt attention to treatment preferences.

At the hospice level, applying different LOS exclusion criteria had an impact on denominator size. Smaller denominator size at the hospice level may generate less stable and reliable QM scores. Under the LOS < 7 days exclusion, the median number of qualifying stays in the denominator was 94 stays per hospice vs. a median number of 136 stays per hospice with no LOS exclusion. Applying the LOS exclusion resulted in some hospices with no qualifying stays, which excluded 14 hospices from QM score calculations altogether.

Applying or removing the LOS exclusion generally had little impact on the distribution of hospices' QM scores. Under the 7-day LOS exclusion, the mean score was 98.1%, the median score was 100%, and the score for hospices in the 10th percentile distribution was 95.2%. With no LOS exclusions, the mean score was 98.0%, the median score was 100%, and the score for hospices in the 10th percentile distribution was 95.0%. The impact of the different LOS criteria on the distribution of QM scores was consistent across

quarters.

In summary, these new analyses demonstrate that removing the LOS exclusion in QM calculations, thus including all hospice patients in the denominator, is feasible and appropriate for quality of care at the patient level, and enhances completeness and statistical stability of QM reporting at the hospice level.

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 medical record includes documentation of life sustaining preferences

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.)*
N/A

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.
Documentation of life-sustaining treatment preferences should reflect patient self-report; if not available due to patient loss of decisional capacity, discussion with surrogate decision-maker and/or review of advance directive documents are acceptable. The numerator condition is based on the process of eliciting and recording preferences, whether the preference statement is for or against the use of various life-sustaining treatments such as resuscitation, ventilator support, dialysis, or use of intensive care or hospital admission. This item is meant to capture evidence of discussion and communication. Therefore, brief statements about an order written about life-sustaining treatment, such as "Full Code" or "DNR/DNI" do not count in the numerator. Documentation using the POLST paradigm with evidence of patient or surrogate involvement, such as co-signature or description of discussion, is adequate evidence and can be counted in this numerator.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Seriously ill patients enrolled in hospice OR receiving specialty palliative care in an acute hospital setting.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, 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)*
The Treatment Preferences quality measure is intended for patients with serious illness who are enrolled in hospice care OR receive specialty palliative care in an acute hospital setting. Conditions may include, but are not limited to: cancer, heart disease, pulmonary disease, dementia and other progressive neurodegenerative diseases, stroke, HIV/AIDS, and advanced renal or hepatic failure.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Patients with length of stay < 1 day in hospice or palliative care

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)*
Calculation of length of stay; discharge date is identical to date of initial encounter.

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)*
N/A

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*)

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.)*
Chart documentation of life sustaining preferences:
a.Step 1- Identify all patients with serious, life-limiting illness who are enrolled in hospice OR who received specialty palliative care in an acute hospital
b.Step 2- Exclude patients if length of stay is < 1 day.
c.Step 3- Identify patients with documented discussion of preference for life sustaining treatments.

Quality measure = Numerator: Patients with documented discussion in Step 3 / Denominator: Patients in Step 1 – Patients excluded in Step 2

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.
Hospice: The hospice analysis was not based on a sample. It was conducted on the entire hospice population that had admission and discharge records in the specified period of analysis.

Palliative care: consecutive sample of equal numbers of admissions + decedents beginning with a randomly selected date.

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.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Hospice: For the treatment preferences measure, there are three items on the HIS that can include missing data – F2000B, F2100B, and F2200B. If missing, these items are coded as dashes and analyzed as though the process did not happen.

In order to assess how these missing data impact the validity of the treatment preferences measure, we conducted the patient stay- and hospice-level analyses. For the patient stay-level analysis, we calculated the number and percentage of eligible patient stays for which the HIS-Admission records included a dash for each of these three items. The overall rate of missing data ranged from 0.01 percent to 0.02 percent between October 2014 and September 2015. For the hospice-level analysis we calculated each hospice's percent of eligible admissions that included missing data for items F2000B, F2100B, and F2200B. Over 95% of hospices did not have any admissions with missing data for these items.

Overall, we found that only a very small number of admission records for the eligible stays did not include data, i.e. coded as dash, on treatment preferences items F2000B, F2100B, and F2200B. And a vast majority of hospices did not have any missing data for these items. These results indicate that missing data for these items should not have a negative impact on the validity of the QM.

Palliative care: N/A

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record

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.
Hospice: Hospice analysis uses the Hospice Item Set (HIS) as the data source to calculate the quality measure.

Palliative Care: Structured medical record abstraction tool, with separate collection of denominator and numerator data

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 in attached appendix at A.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Group/Practice, Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospice, 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
1641_MeasureTesting_MSF5.0_Data.doc,1641_MeasureTestingAttachment_2.26.16.docx

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 1641	**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)***:**

Hospice: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1641 - Hospice and Palliative Care Treatment Preferences.

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.

Palliative Care: Two research nurse abstractors independently recorded quality measures data on a random subset of 20 seriously ill patients. Abstractors used the pre-defined operational definitions and a structured chart abstraction tool to record numerator and denominator data separately. Patients were a subsample of 460 seriously ill patients without specialty palliative care admitted to an acute care hospital for at least 1 day to four inpatient services. Records eligible for sampling included all seriously ill adult patients admitted to medical and surgical intensive care, medically complex patients aged 65 and older admitted to an Acute Care of the Elderly Unit, and medical oncology patients with Stage IV carcinoma.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)*:

Hospice:

Split-half reliability. Split-half reliability assesses the internal consistency of a QM in different samples by randomly dividing the patient stays within each hospice into two halves, and calculating correlation between the QM scores on the basis of the two randomly divided halves. In this analysis, we conducted a split-half reliability analysis on all facilities with 20 or more patient stays counted in the measure denominator, and used the Interclass Correlation (ICC) coefficients to measure the internal reliability. In general, the ICC coefficient varies between 0 and 1, where an ICC of 0 indicates no reliability and an ICC of 1 indicates perfect reliability.

Signal-to-noise analysis. If a measure is reliable, then true differences in provider performance should explain a substantial proportion of the variance in QM scores. We conducted an analysis of variance (ANOVA) to determine what proportion of total variance in the measure is attributable to differences among providers. A higher proportion indicates better reliability.

Stability analysis. Stability analysis describes the extent to which providers' performance assessed by a QM changes across time. We analyzed the change in facility scores between four consecutive quarters (Q4 2014 and Q1 2015, Q1 and Q2 2015, and Q2 and Q3 2015). The changes in facility scores are reported in standard deviations.

Palliative Care: Inter-rater reliability between the two abstractors was assessed using kappa statistics Two research nurse abstractors independently recorded quality measures data on a random subset of 20 seriously ill patients. Abstractors used the pre-defined operational definitions and a structured chart abstraction tool to record numerator and denominator data separately.

2a2.3 Testing Results *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:

Hospice:

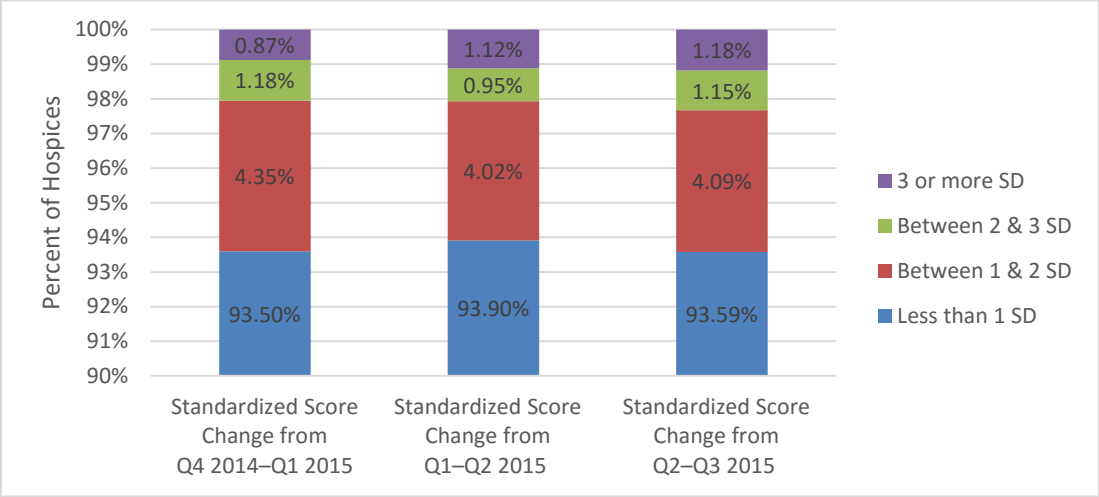
Split-half reliability: In general, the ICC coefficient varies between 0 and 1, where an ICC of 0 indicates no reliability and an ICC of 1 indicates perfect reliability. The ICC coefficient for this measure is 0.91, indicating high internal reliability.

Signal-to-noise analysis: The analysis results found the signal-to-noise ratio to be 0.98, indicating that about 98% of the variance in this measure is because of differences among facilities. This proportion indicates strong reliability for this measure.

Stability analysis: The results of this analysis indicated that facility scores were very stable. Slightly less than 95% of facilities had a change in QM score of less than one standard deviation, indicating high stability of the QM. The number of facilities with a change in QM of less than one standard deviation showed little change across quarters from 93.5 to 93.9 to 93.6, suggesting consistent reliability across time. Less than 5% of facilities

had a change in QM between one and two standard deviations. These results indicate a measure that is generally quite stable. Figure 1 illustrates the change in facility scores between the four consecutive quarters.

Figure 1
Standardized Score Change in QM Score from Q4 2014 to Q3 2015



Palliative Care: Kappa scores range from 0 to 1 with higher scores indicating better agreement. The nurse abstractors achieved prefect (kappa=1.0) inter-rater reliability for this measure. Landis and Koch describe kappa values that range from 0.81 – 0.99 as almost perfect and Fleiss describes kappas over 0.75 as excellent.

Landis, J.R.; Koch, G.G. (1977). "The measurement of observer agreement for categorical data". Biometrics 33 (1): 159–174

Fleiss, J.L. (1981). Statistical methods for rates and proportions (2nd ed.). New York: John Wiley

2b. VALIDITY. Validity, Testing, including all Threats to Validity: 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:**

The measure focus is communication of patient Treatment Preferences. The target populations are hospice patients, and seriously ill hospitalized patients with diverse underlying diagnoses who are at high risk for palliative care clinical needs. This measure focus and target population is consistent with the research evidence base discussed in 1c; however, the research evidence base includes less direct evidence for the hospice population.

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)***:**

Hospice: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate

seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1641 - Hospice and Palliative Care Treatment Preferences.

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.

Palliative Care: The total patient sample size was 562. Chart abstractions were completed for 102 consecutive seriously ill patients with specialty palliative care consultation, and a random sample of 460 seriously ill patients without specialty palliative care admitted to an acute care hospital for at least 1 day to four inpatient services with high proportions of seriously ill patients. Records eligible for sampling included all patients admitted to medical and surgical intensive care, medically complex patients aged 65 and older admitted to a Geriatric Evaluation Unit, and medical oncology patients with Stage IV carcinoma. Because palliative care domains become even more relevant closer to death, patients dying in hospital were over-sampled to ensure a final ratio of 1 decedent to 1 live discharge. Consistent with oversampling of decedent records, 55% of these patients died in hospital.

The age of the patients ranged from 16 to 99 years, with the mean age 61. Patients were predominantly Caucasian (65%), with smaller subgroups who were African American (24%) and Hispanic / Latino (4%) The most common life-limiting diagnoses were infections (37%), cancer (34%), pulmonary (29%), and neurologic diseases (21%).

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

Hospice:

Correlations Providers should perform similarly on QMs that reflect the quality of similar care processes. Thus, a common strategy used to evaluate validity of a QM is to examine the relative performance of the provider on QMs that are conceptually related, either due to some core competency manifested across a stay or for processes that occur in close proximity We conducted nonparametric Spearman rank correlation analysis among all seven HQRP QMs, which address care processes around hospice admission that are clinically recommended or required in the hospice Conditions of Participation. We used nonparametric methods since the data are heavily skewed (most providers perform well, while a small proportion have lower scores). Since all of the seven current QMs measure desirable care assessment processes upon hospice admission, higher positive Spearman correlation coefficients among these QMs will demonstrate higher validity of the QM set. Statistically significant correlations, indicated by p-values of the Spearman correlation coefficients less than 0.05, also support the validity of the QM set.

Missing data. Missing data represent a potential threat to the validity of an HIS item, which in turn may harm the validity of the QM that relies on the item. For the treatment preferences measure, there are three items on the HIS that can include missing data – F2000B, F2100B, and F2200B. In order to assess how these missing data impact the validity of the treatment preferences measure, we conducted the patient stay- and hospice-level analyses.

Palliative Care sample:

Face validity of PEACE quality measures for hospital-based specialty palliative care was addressed using stakeholder review and feedback. Investigators prepared data reports in a summary format with detailed operational definitions, and led a 1-hour discussion with nursing and physician leaders from each service group – MICU, SICU, Acute Care for the Elderly (Geriatrics, Oncology, and Palliative Care. The discussion included feedback of quality measure data, response to questions and critiques, and eliciting stakeholder feedback about the validity and actionability of this data for the care of their patients. Stakeholders were specifically asked to comment on the accuracy of the data as a reflection of current care practices, and their highest priority area for future quality improvement.

Construct validity was tested by comparing the PEACE quality measures for patients seen by specialty interdisciplinary palliative care consultants to those not receiving specialty palliative care services.

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)***:**

Hospice:

Correlations: Table 1 presents the Spearman correlation coefficients. The p-values for all the Spearman correlation coefficients are significant ($p < 0.001$). The significant positive correlations between every pair of QMs indicate that high-performing hospices in one area also provided high-quality care in other areas at hospice admission. Overall, the correlations between the QMs are low to moderate. The clustering of QM scores, skewed distributions, and low variability across all seven QMs may affect the level of correlations between QMs.

Table 1
Correlation of Hospice QMs, Percentile Ranking

Quality measure	NQF #1647 (modified) Beliefs/Values Addressed	NQF #1634 Pain Screening	NQF #1637 Pain Assessment	NQF #1639 Dyspnea Screening	NQF #1638 Dyspnea Treatment	NQF #1617 Bowel Regimen
NQF #1641 Treatment Preferences	0.64***	0.37***	0.21***	0.44***	0.28***	0.25***
NQF #1647 (modified) Beliefs/Values Addressed		0.42***	0.17***	0.43***	0.31***	0.25***
NQF #1634 Pain Screening			0.17***	0.41***	0.28***	0.15***
NQF #1637 Pain Assessment				0.17***	0.08***	0.20***
NQF #1639 Dyspnea Screening					0.30***	0.25***
NQF #1638 Dyspnea Treatment						0.34***

NOTE: The correlation is on the basis of each hospice's percentile ranking on the QMs.

*** indicates significant correlation at $p < 0.001$

Overall, this measure has significant positive correlations with the other QMs, indicating hospices providing

higher-quality care in this area also performed better in other areas at hospice admission. The QM having the strongest correlation with this measure is NQF #1647, (modified) Beliefs/Values Addressed at (ρ = 0.64). We expect the strong correlation between these two QMs because they both address the competency of the hospice to solicit the patient's preferences.

Missing data. For the patient stay-level analysis, we calculated the number and percentage of eligible patient stays for which the HIS-Admission records included a dash for each of these three items. The overall rate of missing data ranged from 0.01 percent to 0.02 percent between October 2014 and September 2015. For the hospice-level analysis we calculated each hospice's percent of eligible admissions that included missing data for items F2000B, F2100B, and F2200B. Over 95% of hospices did not have any admissions with missing data for these items.

Overall, we found that only a very small number of admission records for the eligible stays did not include data, i.e. coded as dash, on treatment preferences items F2000B, F2100B, and F2200B. And a vast majority of hospices did not have any missing data for these items. These results indicate that missing data for these items should not have a negative impact on the validity of the QM.

Palliative Care:

Face Validity: Stakeholder discussions provided broad endorsement of face validity, with some considerations for specific patient populations. Medical oncologists endorsed the face validity of these quality measures, but favored quality measures endorsed by oncology professional organizations.

Construct Validity: Patients who received specialty palliative care were more likely to have documentation of their preferences for or against receiving life-sustaining treatments (91% vs 59%, p>0.001).

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)***:**

Hospice: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1641 - Hospice and Palliative Care Treatment Preferences

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.

Palliative Care:N/A

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

Hospice: Under contract to the Centers for Medicare and Medicaid Services (CMS), the Research Triangle Institute (RTI) performed descriptive analyses on 5 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.

Palliative Care: N/A

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

Hospice: At the patient level, approximately 40% of patients were excluded based on the LOS < 7 days exclusion, thus omitting this essential quality of care standard for many seriously ill patients. Further, excluding short-stay patients omits measures of quality of care for a subset of the sickest patients entering hospice care, and those who arguably may be in greatest need of rapid symptom assessment and treatment. In addition, the original rationale for this exclusion -- allowing time for hospice providers to complete this care process -- does not appear to be necessary. Our analyses show that a large portion of treatment preference discussions were performed on day 1 of admission to hospice, demonstrating a normative standard of care includes prompt attention to treatment preferences.

At the hospice level, applying different LOS exclusion criteria had an impact on denominator size. Smaller denominator size at the hospice level may generate less stable and reliable QM scores. Under the LOS < 7 days exclusion, the median number of qualifying stays in the denominator was 94 stays per hospice vs. a median number of 136 stays per hospice with no LOS exclusion. Applying the LOS exclusion resulted in some hospices with no qualifying stays, which excluded 14 hospices from QM score calculations altogether.

Applying or removing the LOS exclusion generally had little impact on the distribution of hospices' QM scores. Under the 7-day LOS exclusion, the mean score was 98.1%, the median score was 100%, and the score for hospices in the 10th percentile distribution was 95.2%. With no LOS exclusions, the mean score was 98.0%, the median score was 100%, and the score for hospices in the 10th percentile distribution was 95.0%. The impact of the different LOS criteria on the distribution of QM scores was consistent across quarters.

In summary, these new analyses demonstrate that removing the LOS exclusion in QM calculations, thus including all hospice patients in the denominator, is feasible and appropriate for quality of care at the patient level, and enhances completeness and statistical stability of QM reporting at the hospice level.

Palliative Care: N/A

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)*:

N/A

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

N/A

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)***:**

N/A

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

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)***:**

Hospice: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1641 - Hospice and Palliative Care Treatment Preferences.

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.

Palliative Care: The total patient sample size was 562. Chart abstractions were completed for 102 consecutive seriously ill patients with specialty palliative care consultation, and a random sample of 460 seriously ill patients without specialty palliative care admitted to an acute care hospital for at least 1 day to four inpatient services with high proportions of seriously ill patients. Records eligible for sampling included all patients admitted to medical and surgical intensive care, medically complex patients aged 65 and older admitted to a Geriatric Evaluation Unit, and medical oncology patients with Stage IV carcinoma. Because palliative care domains become even more relevant closer to death, patients dying in hospital were over-sampled to ensure a final ratio of 1 decedent to 1 live discharge. Consistent with oversampling of decedent records, 55% of these patients died in hospital.

The age of the patients ranged from 16 to 99 years, with the mean age 61. Patients were predominantly Caucasian (65%), with smaller subgroups who were African American (24%) and Hispanic / Latino (4%) The most common life-limiting diagnoses were infections (37%), cancer (34%), pulmonary (29%), and neurologic diseases (21%).

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

Hospice: Confidence interval analysis. We examined proportions of hospices with the QM scores that are significantly different from the national hospice-level mean. If a high proportion of hospices have a measure score significantly different from the mean, the QM can identify facilities with different levels of performance. For this analysis, statistical significance was determined using 95 percent confidence intervals: a hospice's QM score was significantly different from the national mean if the national mean was not included within the hospice's 95 percent confidence interval. High-performing facilities should have scores that are significantly below average, and low-performing facilities should be significantly above average.

Palliative Care: Construct validity was tested by comparing the PEACE quality measures for patients seen by specialty interdisciplinary palliative care consultants to those not receiving specialty palliative care services. Percentage of patients with and without specialty palliative care for whom the quality measure was met was compared for difference using the chi-square statistic.

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)*:

Hospice:

The mean score for this QM was 98.0% with a range from 0% to 100%, the median was 100%, the interquartile range was 1.5, and the standard deviation was 6.3. For this QM, 53.5% of hospices had perfect scores and 4.6% of hospices scored below 90%.

Scores by decile:
10th percentile 95.0%
25th percentile 98.5%
Median 100%
75th percentile 100%
90th percentile 100%

Across all hospices, 33.2% had a QM score that is significantly different than the national mean. Hospices were more likely to report scores above the national mean than below the national mean (23.3% vs 9.9%, respectively, overall). The QM is able to identify those hospices that are performing well (higher than the national mean) and those that are performing less well (lower than the national mean).

.

Palliative care sample: Seriously ill hospitalized patients who received specialty palliative care were more likely to have documentation of their treatment preferences compared to patients without specialty palliative care. (91% vs 59%, p>0.001).

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)*:

N/A

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

N/A

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)*:

N/A

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)*:

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

Future research with larger sample sizes can be used to test for differential performance by race / ethnicity and by gender.

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)*

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

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.

Missing Data. Missing data represent a potential threat to the validity of an HIS item, which in turn may harm the validity of the QM that relies on the item. For the treatment preferences measure, there are three items on the HIS that can include missing data – F2000B, F2100B, and F2200B. If missing, these items are coded as dashes. In order to assess how these missing data impact the validity of the treatment preferences measure, we conducted the patient stay- and hospice-level analyses. For the patient stay-level analysis, we calculated the number and percentage of eligible patient stays for which the HIS-Admission records included a dash for each of these three items. The overall rate of missing data ranged from 0.01 percent to 0.02 percent between October 2014 and September 2015. For the hospice-level analysis we calculated each hospice's percent of eligible admissions that included missing data for items F2000B, F2100B, and F2200B. Over 95% of hospices did not have any admissions with missing data for these items.

Overall, we found that only a very small number of admission records for the eligible stays did not include data, i.e. coded as dash, on treatment preferences items F2000B, F2100B, and F2200B. And a vast majority of hospices did not have any missing data for these items. These results indicate that missing data for these items should not have a negative impact on the validity of the QM. There was no missing data for the elements needed to calculate this measure for the Palliative Care sample.

Record abstraction does not require collection of unique patient identifiers and thus protects confidentiality. Timing of data collection can be concurrent with admission / initial encounter care, or can be retrospective based on medical record sampling.

Costs have not been formally estimated; medical record abstraction or electronic capture of this type of data will have more modest costs compared to survey data.

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)*.
N/A

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)
Public Reporting	Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Hospice Item Set-Hospice Quality Reporting Program https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Hospice-Quality-Reporting/Hospice-Item-Set-HIS.html Quality Improvement (Internal to the specific organization) Voluntary use by UNC Palliative Care Program http://www.med.unc.edu/pcare/resources/PEACE-Quality-Measures

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

Name of program and sponsor: Hospice: HIS/CMS; Palliative Care: Voluntary use by PC organizations

Purpose- Hospice: Quality reporting for hospice; Palliative Care: Internal quality improvement

Geographic area and number and percentage of accountable entities and patients included – United States and all accountable entities and 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**

a. Due to insufficient longitudinal data, we are currently unable to discuss progress.

b. Analyses encompassed 3,922 hospice organizations and 1,218,786 patient stays across the United States.
Region:
South: 39.3%
West: 25.1%
Midwest: 23.1%
Northeast: 11.3%
Territories: 0.94%
Unknown: 0.25%

Urban/rural status:
Urban: 75.8%
Rural: 23.9%
Unknown: 0.31%

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.**
We only have one year of data to report, which is not enough to show trends over time.

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.**
None

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.)
This measure is part of the NPCRC Key Palliative Measures Bundle. Refer to the NPCRC cover letter and table of bundle measures for description of the selection and harmonization of the Key Palliative Measures Bundle.

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 **Attachment:** Appendix_A.1_NQF_1641.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** University of North Carolina-Chapel Hill
Co.2 **Point of Contact:** Laura, Hanson, lhanson@med.unc.edu, 919-843-4096-
Co.3 **Measure Developer if different from Measure Steward:** University of North Carolina- Chapel Hill
Co.4 **Point of Contact:** Laura, Hanson, lhanson@med.unc.edu, 919-843-4096-

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.
The Carolinas Center for Medical Excellence PEACE Project Technical Expert Panel

The PEACE project team convened a 14-member Technical Expert Panel (TEP) of nationally recognized experts with extensive experience in the following areas: medical or nursing expertise in hospice and palliative care, methods and instrumentation, and quality improvement. Using criteria provided by the CCME study team, TEP members rated each potential quality measure on four criteria: importance, scientific soundness, feasibility and usability.
Mary Ersek, PhD, RN, Research Associate Professor, Swedish Medical Center- Pain Research Department, Seattle, WA
Betty R. Ferrell, PhD, FAAN, Research Scientist, City of Hope National Medical Center, Duarte, CA
Sean Morrison, MD, Mount Sinai Medical Center, NY, NY
Richard Payne, MD, Director, Duke Institute on Care at the End of Life, Duke Divinity School, Durham, NC

Chris Feudtner, MD, PHD, MPH, Children's Hospital of Philadelphia, Philadelphia, PA
Karen Steinhauser, PhD, Research Health Scientists, Center for Health Services Research in Primary Care, Durham VA Medical Center and Duke University, Durham, NC
Joan M. Teno, MD, Professor of Community Health and Medicine, Center for Gerontology and Health Care Research, Brown University, Providence, RI
Melanie Merriman, PhD, MBA, Touchstone Consulting, North Bay Village, FL
Sydney Dy, MD, MSc, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
David Casarett, MA, MD, Assistant Professor, Division of Geriatrics, Institute on Aging and Center for Bioethics, University of Pennsylvania School of Medicine and NHPCO Board of Directors
Judi Lund-Person, Vice President, Division of Quality, National Hospice and Palliative Care Organization, Washington, DC
Jean Kutner, MD, MSPH, Associate Professor, University of Colorado Health Sciences Center, Denver, CO
Lin Simon, Analyst, National Hospice and Palliative Care Organization, Washington, DC
Karen Pace, NAHC

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2010

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

Ad.4 **What is your frequency for review/update of this measure?** 3 years or as requested

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

Ad.6 **Copyright statement:**

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:** This measure is part of the NPCRC Key Palliative Measures Bundle. Refer to the NPCRC cover letter and table of bundle measures for description of the selection and harmonization of the Key Palliative Measures Bundle.

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 #: 1634
Measure Title: Hospice and Palliative Care -- Pain Screening
Measure Steward: University of North Carolina-Chapel Hill
Brief Description of Measure: Percentage of hospice or palliative care patients who were screened for pain during the hospice admission evaluation / palliative care initial encounter.
Developer Rationale: Pain is prevalent and undertreated for many populations of seriously ill patients, including those patients nearing the end of life. Poor screening, assessment and under treatment of pain is more common for patients with serious illness who are also of minority race ethnicity. Use of the Pain Screening and Pain Assessment quality measures will increase reporting and efforts to improve awareness of the presence of pain (screening) and assessment of severity, etiology and effect on function (assessment) which are the two essential first steps required for quality pain management and treatment.

Numerator Statement: Patients who are screened for the presence or absence of pain (and if present, rating of its severity) using a standardized quantitative tool during the admission evaluation for hospice / initial encounter for palliative care.
Denominator Statement: Patients enrolled in hospice OR patients receiving specialty palliative care in an acute hospital setting.
Denominator Exclusions: Patients with length of stay < 1 day in palliative care.

Measure Type: Process
Data Source: Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record
Level of Analysis: Clinician : Group/Practice, 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 provided a rationale for the relationship of this process of care (screening for pain) and patient outcomes (improvements in pain and reductions in the effect of pain on function and quality of life).
- The developer cited selected individual studies, systematic reviews, and clinical practice guidelines to support the effectiveness of medical treatment for pain, the effectiveness of expert pain assessment and specialty care teams to improve pain, and the importance of screening, assessing, and treating pain in seriously and terminally ill patient populations.
    - Of these, the 2005 American Pain Society (APS) guidelines seem to best match the focus on the measure (recommendation states that all patients should be routinely screened for pain, and when it is present, pain intensity should be recorded). The 2005 APS recommendations result from literature reviews, expert experience, and consensus; the evidence was not graded and the evidence for screening was not summarized.
    - The 2008 American College of Physician Clinical Practice Guideline (Qaseem et al., 2008) includes a recommendation regarding assessment ("In patients with serious illness at the end of life, clinicians should regularly assess patients for pain, dyspnea, and depression" Grade: strong recommendation, moderate quality of evidence.) While graded, the evidence summarized was in fact relevant to treatment modalities, not screening or assessment, and was for the most part limited to cancer patients.
- During its 2012 evaluation of the measure, the Palliative/End-of-Life (EOL) Committee agreed that the evidence provided primarily addresses gaps in assessment for pain rather than supporting a link between screening for pain and better patient outcomes. Because a pain assessment is triggered by screening, members of the Committee considered this as additional evidence for the importance of the measure.
- The Committee recommended that this measure and the pain assessment measure (#1637) be paired.

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:

- Although the developer attested that there have been no changes in the evidence since it was endorsed in 2012, they added two new guidelines to the submission.
    - The 2011 British Columbia Medical Services Commission guideline relates to palliative care for patients with incurable cancer or advanced disease; it is based on a systematic review of evidence but the evidence is not graded. This guideline calls for the assessment of pain using the OPQRSTUV mnemonic (onset, provoking, quality, region, severity, treatment, understanding, values).
    - The 2013 ICSI guideline on Palliative Care for Adults. The pertinent recommendation from this graded guideline states that "The physical aspects of the patient's serious illness should be an integral component of the palliative care plan (Low Quality Evidence, Strong Recommendation)". The guideline goes on to say "Each patient should be frequently evaluated for these issues."

Exception to evidence:

- Because the evidence for this measure is mostly tangential to the relationship of pain screening to patient outcomes or based on expert opinion only, it is insufficient to meet NQF's criterion for evidence. However, an exception to the evidence criterion is allowed if the Committee agrees that empirical evidence is not needed to hold providers accountable for the measure and that other outcome measures or evidence-based measures are not available or feasible at this time.

Guidance from the Evidence Algorithm

Process measure but mostly tangential evidence (Box 3) → Evidence not graded (Box 7) → An outcome measure for

pain exists, but is limited to hospice only (Box 10) → Systematic assessment of expert opinion (Box 11) → If Committee agrees it is OK/beneficial to hold providers accountable for performance in the absence of empirical evidence of benefits to patients → rate as INSUFFICIENT WITH EXCEPTION

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?*
- *Are there, or could there be, performance measures of a related health outcome, OR evidence-based intermediate clinical outcomes, intervention/treatment?*
- *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*
- *Does the SC agree that it is acceptable (or beneficial) to hold providers accountable without empirical evidence?*

Preliminary rating for evidence: ☐ **High** ☐ **Moderate** ☐ **Low** ☒ **Insufficient**
Rationale: No evidence links screening for pain to improved patient outcomes.

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.

- Facility-level (Hospice):
  - The developer provided data from the Hospice Quality Reporting Program (HQRP) for FY15. The data were collected from 3,922 hospice organizations and approximately 1.2 million patient stays.
  - The developer reported that 20.1% of hospices had perfect scores and 20.7% of hospices scored below 90%.

Mean	93.5%
Range	0% - 100%
10th percentile	82.6%
25th percentile	91.5%
Median	96.8%
75th percentile	99.4%
90th percentile	100%

- Clinician Group/Practice level (Palliative Care):
  - Data specific to palliative care (for clinician group/practice in the hospital setting) are not yet available , although the developer expects these data will become available for NQF review next year.

Disparities

- Facility-level (Hospice):

Subgroup	**Hospice-level**	
	Score	**Wilcoxon-Mann-Whitney test p-value**
Race		
Non-white ≥ national median of 11.9%	93.7%	p=0.45
Non-white < national median of 11.9%	93.3%	
Sex		
Female ≥ national median of 55.2%	93.3%	p=0.005
Female < national median of 55.2%	93.7%	
Medicaid status (proxy for SES)		

Medicaid patients >= national median of 21.5%	92.1%	p<0.001
Medicaid patients < national median of 21.5%	94.6%	
Geographic location		
Rural	93.2%	0.98
Urban	93.6%	

- Clinician Group/Practice Level of Analysis (Palliative Care):
  - Disparities data for palliative care (for clinician group/practice in the hospital setting) are not available, although the developer expects these data will become available for NQF review next year.

Questions for the Committee:

- *Is there a gap in hospice care (for pain screening) that warrants a national performance measure?*
- *Is the Committee aware of evidence demonstrating a gap in hospital-based palliative care (for pain screening) that warrants a national performance measure?*
- *Palliative-specific disparities information is not provided. Are you aware of evidence that disparities exist in this area of healthcare?*

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.

* Pain is a common symptom in patients enrolled in palliative care and hospice programs, is often both unaddressed or less than adequately addressed by clinicians.
This measure is extracted from the records of patients and is highly relevant to their care. I would consider this a direct measurement.
Initial pain screening is important to both uncover the symptom and lay a foundation for further treatment
* Maintenance measure- Process. Evidence cited to be mostly tangential. However, an exception to the evidence criterion is allowed if the Committee agrees that empirical evidence is not needed to hold providers accountable for the measure and that other outcome measures or evidence-based measures are not available or feasible at this time. I agree with the Committee's previous decision.

1b.

* There is a performance gap and it does demonstrate variability and less than optimal performance in Hospice programs ranging from 83 to 100%. Pain screening is virtually universal for inpatient palliative care patients. Racial disparities exist but are modest but statistically less for Hispanic patients in hospice.
* Performance data provided, rated with moderate opportunity for improvement, which surprises me. Opportunity also related to disparities gap.

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 sources specified in the submission form include EHRs and other electronic clinical data.
Specifications:

- This measure is specified for the clinician group/practice level of analysis in the hospital setting and for the facility level of analysis in the hospice setting. A higher score indicates better quality.
- For the hospice setting, the numerator (patients who are screened for the presence or absence of pain) and

denominator (patients enrolled in hospice) are identified from the Hospice Item Set, a standardized, patient-level dataset used in the Medicare Hospice Quality Reporting Program.
- For palliative care (hospital) setting, data for the numerator and denominator are collected using a structured medical record abstraction tool.
- Screening may be completed using verbal, numeric, visual analog, rating scales designed for use the non-verbal patients, or other standardized tools.
- Data are collected at the hospice admission evaluation or initial clinical encounter for palliative care.
- There is one exclusion criterion for the denominator: patients with a length of stay <1 day in palliative care
- A calculation algorithm is provided.
- The developer indicates that sampling is permissible for the hospital setting. While some basic instructions are given, there is no guidance about the number of patients needed for the sample.

Prior Evaluation
- In the 2012 evaluation of this measure, the Committee expressed concern that the exclusion of patients with length of stay <7 days in hospice or <1 day in palliative care might exclude a significant percentage of patients. Committee members noted a preference for a measure without these exclusions.

Changes to specifications since previous evaluation:
- After analysis of FY15 hospice data, the developer has changed the specifications so as to no longer exclude hospice stays of less than 7 days.

Questions for the Committee :
- *Are all the data elements clearly defined?*
- *Is the logic or calculation algorithm clear?*
- *Is it reasonable to exclude palliative care patients with < 1 day length of stay?*
- *Do all hospitals use EHRs or are some still using paper records (note: the clinician group/practice measure for the hospital setting specifies use of EHRs only)?*
- *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:
- Previous reliability testing included inter-rater reliability testing of hospital data for 20 patients. Reliability testing using hospice data was not previously conducted.

Describe any updates to testing
- Score-level testing for the hospice setting included split-half reliability, signal-to-noise, and stability analyses at the facility level.

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:
- Clinician Group/Practice level (Palliative Care):
  - Reliability of the data elements was assessed using data obtained from the medical charts for 20 seriously ill patients without specialty palliative care admitted to an acute care hospital. Inter-rater reliability between two nurse abstractors was calculated. This is an appropriate method for

demonstrating data element reliability.

- Facility-level (Hospice):
    - Reliability of the measure score was assessed using FY15 data from the Hospice Quality Reporting Program (n=3,992 hospice organizations and n=1,218,786 patient stays). Two types of reliability testing were conducted: the first via a signal-to-noise analysis and the second using a split-sample (or "test-retest") methodology. Both are appropriate methods for testing reliability. Stability analysis examined changes in provider's performance over time. 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:

- Clinician Group/Practice level (Palliative Care):
    - Developers report a single kappa value of 1.0.
    - 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 1.0 means that the raters agreed 100% 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.
    - NQF guidance indicates that data element testing should be conducted for all critical data elements, although at minimum, results about the numerator, denominator, and exclusions should be provided. Only a single kappa value was reported. Although not explicitly stated, we assume that the 20 charts that were abstracted included only seriously ill patients who received specialty palliative care in an acute hospital setting and whose length of stay was at least one day, and thus this ,kappa value applies to the numerator.
- Facility-level (Hospice):
    - Split-half analysis ICC=0.86.
        - The ICC reflects the percentage of variance in score results that is due to "true" or real variance between the hospices. A value of 0.7 is often regarded as a minimum acceptable reliability value.
    - Developers report a signal-to-noise ratio of 0.97.
        - A signal-to-noise analysis quantifies the amount of variation in a performance measure that is due to true differences between hospices (i.e., signal) as opposed to random measurement error (i.e., noise). Results will vary based on the amount of variation between the providers and the number of patients treated by each provider. This method results in a reliability statistic that ranges from 0 to 1 for each hospital. A value of 0 indicates that all variation is due to measurement error and a value of 1 indicates that all variation is due to real differences in hospices performance. A value of 0.7 often is regarded as a minimum acceptable reliability value.

Guidance from the Reliability Algorithm
Facility-level (hospice):
Precise specifications (Box 1) → Empirical reliability testing with measure as specified (Box 2) → Score-level testing (Box 4) → Appropriate method (Box 5) → High certainty that measure results are reliable (Box 6a) → High

Clinician-level (Palliative)
Precise specifications (Box 1) → Empiric
al reliability testing with measure as specified (Box 2) → Score-level testing not conducted (Box 4) → Data elements tested (Box 8) → Appropriate method (Box 9) → High level of agreement between raters (Box 10a) → Moderate

Questions for the Committee:

- *Are there any concerns about accurately and consistently identifying the denominator and exclusions in the hospital setting?*
- *Is the test sample adequate to generalize for widespread implementation for palliative care? If not, is current testing sufficient until more data are available (assuming the developer is planning to do additional testing)?*

- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

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**

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:

- In the prior review, the measure was tested for the palliative care (hospital) setting using face and construct validity.

Describe any updates to validity testing

- Additional validity testing of the measure score at the facility level of analysis (for the hospice setting) was conducted using FY15 data from the Hospice Quality Reporting System (HQRP).

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

Validity testing method:

- Facility-level (Hospice)
  - Using FY15 data from the Hospice Quality Reporting System (n=3,992 hospice organizations; n=1,218,786 patient stays), the developer conducted non-parametric Spearman rank correlation analysis between this measure and 5 other hospice quality measures. They hypothesized that agencies should perform similarly on assessment processes at hospice admission and expected the resulting correlations to be high. This is an appropriate method of score-level validation.
- Clinician Group/Practice level (Palliative Care)
  - Face validity was assessed by a group of nursing and physician stakeholders who were asked to comment on the validity, accuracy, and actionability of the measure. According to NQF guidance, the face validity of the measure score as a quality indicator may be adequate if it explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. It does appear that this testing conforms to NQF's requirements for face validity.
  - Construct validity was tested by comparing measure results for seriously ill patients seen in specialty interdisciplinary palliative care consultations (n=102) in one hospital to those who did not receive these services (n=460). The developers did not explain what they expected to find with this analysis and how it would demonstrate that the measure results reflect quality of care.

Validity testing results:

- Facility-level (Hospice)
  - Correlation results were positive and statistically significant, confirming the developer's hypothesis. However, the magnitude of the correlation was lower than they had expected. The developer states reasons for this may be due to clustering of scores, skewed distributions, and low score variability.
- Clinician Group/Practice level (Palliative Care)
  - Face validity results from the stakeholder group indicated broad endorsement of the face validity of the measure.
  - The construct validity results likely were inconclusive, as almost all patients were screened for pain, regardless of receipt of specialty palliative care services.

Questions for the Committee:

- *Are the test samples adequate to generalize for widespread implementation?*
- *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 examined the effect of removing the previous length of stay <7 days exclusion criterion for hospice patients. Using FY15 data from the HQRP, the developers found that the removal of the criterion increased the average size of the denominator per hospice organization and had little effect on the distribution of the measure scores. The developers state these findings support the removal of the length of stay criterion as an exclusion criterion from the measure.
- The developer did not include information on the frequency of exclusions for palliative care patients.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the 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)*:

- Meaningful differences among hospices were examined using FY15 data from the Hospice Quality Reporting Program. The developers examined 95% confidence intervals to determine the proportion of hospices with results significantly different from the national hospice-level mean. Results indicated 50.7% of hospices had a score that was significantly different from the national mean. Hospices were more likely to report scores above the national mean than below the national mean (33.8% vs 16.9%, respectively).
- Developers compared measure score results for patients in one acute care hospital who received palliative care consults to those who did not receive palliative care consults. However, this analysis does not speak to whether the measure results reflect meaningful differences between clinician groups at different acute care hospitals.

Question for the Committee:

- *Does this measure identify meaningful differences about quality between hospice facilities? Between clinician groups/practices in the hospital setting?*

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

2b7. Missing Data

- For hospices, the rate of missingness is low (0.02% - 0.11% at the patient-level). More than 90% of hospices had no missing information for three key data elements from the Hospice Item set that make up this measure.
- The developer does not provide any information on missing data specific to palliative care.

Guidance from the Validity Algorithm
Facility-level (Hospice)
Specifications consistent with evidence (Box 1) → potential threats to validity assessed (Box2) →empirical validity

testing (Box 3) → validity testing of the measure score (Box 6) →method appropriate (Box7) →Moderate certainty (Box 8b) → Moderate

Clinician group/practice level (Palliative Care)
Specifications consistent with evidence (Box 1) → potential threats to validity only somewhat assessed (Box2) → empirical validity testing (Box 3) → validity testing of the measure score (Box 6) →method not well described (Box7) → face validity systematically assessed (Box 4) →results indicate substantial agreement as to validity (Box 5) →Moderate (assuming no concerns around exclusions, missing data, or meaningful differences).

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

Committee pre-evaluation comments

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

2a, 2b.
Reliability – Specifications
* The data elements are clear because CMS implemented the Hopsice Item Set, which is a standard instrument. Palliative Care relied on nurse abstractors for which there was good inter-rater reliability. Case abstraction, or medical record entry is subject to the clinician's accurately documenting their activities. The measure is reliable
* Specifications clear. I have no concerns that a screening for pain can be consistently implemented, whether for clinician, facility, hospice or palliative care. (level of analysis is clinician: group practice/facility.

Validity – Specifications
* Pain screening (and pain assessment) was positively correlated with overall quality measures. There is face validity to these measures from clinicians.
* Agree with the moderate rating of validity.

Reliability – Testing
* There was testing for both hospice and Palliative care inter-rater reliability, with high correlation
* Facility: Reliability of the measure score was assessed using FY15 data from the Hospice Quality Reporting Program (n=3,992 hospice organizations and n=1,218,786 patient stays)? Rated high, and agree.
Clinician group:of the data elements was assessed using data obtained from the medical charts for 20 seriously ill patients without specialty palliative care admitted to an acute care hospital. Agree with moderate rating.

Validity Testing
* There is a very large sample size for hospice 1.2 million stays and a high correlation, as mentioned between this measure and overall quality. For palliative care the analysis includes a small sample size, but is valid
* Testing for facility with adequate scope. Clinician group (palliative care):o Face validity results from the stakeholder group indicated broad endorsement of the face validity of the measure.
The construct validity results likely were inconclusive, as almost all patients were screened for pain, regardless of receipt of specialty palliative care services.

Threats to Validity
* The criteria require a patient to be enrolled in hospice for 7 days, which excludes 40% of the potential population. This is unfortunate but necessary at this time. In the future, one might argue to reduce the number of days enrolled to be part of the sample. There is an argument to be made that these measures should be separated into Palliative Care and Hospice measu
* Agree with the exclusion. No concerns related to missing data, and there is no risk adjustment.

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 are routinely collected and are incorporated in hospice providers' electronic clinical documentation system (i.e., electronic health record (EHR)). Data collection for all measures is achieved through use of the Hospice Item Set (HIS). It is unclear if the HIS is automatically populated by the EHR.

- The developer describes one challenge to implementation related to skip patterns built into the EHR. For patients who have an active pain problem, but who are not in pain when they are admitted and screened, the EHR does not allow the clinician to document the completion of a pain assessment. This conflicts with standard clinical practice. The developer states CMS is planning to modify the Hospice Item Set to address this problem, but specific details on these modifications are not provided.
- The developer does not provide any information on potential or actual implementation challenges for palliative care.
- The developer states formal cost estimates of data collection are not available. According to the Guidance Manual for Completion of the Hospice Item Set, completing an entire response for any one admission takes 20 minutes (though seven performance measures, of which this measure is only one, are collected as part of this process).
- In the 2012 endorsement evaluation, the Palliative and End-of-Life Care Steering Committee noted that a substantial data collection effort may be needed to abstract data if electronic data are not available.

Questions for the Committee:

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

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

Committee pre-evaluation comments

Criteria 3: Feasibility

* as the nation's hospitals move to EHRs, inpatient palliative care programs can use "administrative" abstraction for this measure, and it is highly feasible. for Hospices, there are variable charting mechanisms, and chart abstraction by nurses is required. this is very expensive but feasible.

* • The developer describes one challenge to implementation related to skip patterns built into the EHR. For patients who have an active pain problem, but who are not in pain when they are admitted and screened, the EHR does not allow the clinician to document the completion of a pain assessment. This conflicts with standard clinical practice. The developer states CMS is planning to modify the Hospice Item Set to address this problem, but specific details on these modifications are not provided.

• The developer does not provide any information on potential or actual implementation challenges for palliative care.

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

Publicly reported? ☐ **Yes** ☒ **No**

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

OR

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

Accountability program details:

- This measure is included in the Hospice Quality Reporting Program (HQRP). Medicare-certified hospices are required to submit a Hospice Item Set (HIS)-Admission record and an HIS-Discharge record for each patient admission on or after July 1st, 2015. In FY2015, 3,992 hospice organizations provided measure data for 1,218,786 patient stays. The distribution of reporting hospice agencies by geographic area is provided.

Improvement results:

- Longitudinal data for this measure 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 2012, the Measure Applications Partnership (MAP) supported the inclusion of this measure in the Hospice Quality Reporting program. Public comments generally concurred with the recommendation.
- In February 2013, the Measure Application Partnership supported the measure for inclusion in Physician Quality Reporting System (PQRS) to address a core concept (i.e., pain management) not addressed in the program's measure set. Public comments from American Academy of Hospice and Palliative Medicine (AAHMP), Center to Advance Palliative Care (CAPC), and National Coalition for Hospice and Palliative Care (NCHPC) supported the inclusion of this measure in the PQRS. CAPC and NCHPC suggested that the measure should be expanded beyond hospice patients. However, this measure is not included in the PQRS program.
- In 2014, the AAHMP and the Hospice and Palliative Nurses Association published a list of "Top Ten Measures That Matter". Measure #2 of this set (Screening for Physical Symptoms: Percentage of seriously ill patients receiving specialty palliative care in an acute hospital setting >1 day or patients enrolled in hospice >7 days who had a screening for physical symptoms (pain, dyspnea, nausea, and constipation)) "includes" this measure.
- In February 2016, the Measures Application Partnership Post-Acute Care and Long-Term Care workgroup encouraged the continued development of a composite measure of Hospice Item Set measures, which includes NQF #1634.

***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**

Committee pre-evaluation comments

Criteria 4: Usability and Use

* These measures are significant tools in achieving high quality care for these subpopulations
* Not currently publicly reported.
- This measure is included in the Hospice Quality Reporting Program (HQRP). Medicare-certified hospices are required to submit a Hospice Item Set (HIS)-Admission record and an HIS-Discharge record for each patient admission on or after July 1st, 2015. In FY2015, 3,992 hospice organizations provided measure data for 1,218,786 patient stays. The distribution of reporting hospice agencies by geographic area is provided.

Criterion 5: Related and Competing Measures

Competing measures

- 0383: Percentage of visits for patients, regardless of age, with a diagnosis of cancer currently receiving chemotherapy or radiation therapy who report having pain with a documented plan of care to address pain
- 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]*
- 1628: Adult patients with advanced cancer who are screened for pain with a standardized quantitative tool at each outpatient visit *[facility-level and health plan-level process measure in ambulatory setting]*
- 1637: Percentage of hospice or palliative care patients who screened positive for pain and who received a clinical assessment of pain within 24 hours of screening *[clinician-level & facility-level process measure in hospice and hospital setting]*

Harmonization

- Due to differences in care settings, the Committee likely will not be asked to select a best-in-class measure. However, 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: Recommend making the measure specific to palliative or hospice care, not both in the same measure. Palliative care is not restricted to inpatient treatment.

Consider reworking measures 1634 & 1637 to be a singular, stronger measure related to screening for pain.

Consider reviewing measure 209 along with measures 1634 & 1637 to strengthen measures for pain assessment and intervention.

Consider incoporating recommended intervals for screening as the current measure indicates one assessment but one screening is not sufficient in this setting. Perhaps "at each patient encounter" is more appropriate?

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.

NQF #: 1634	**NQF Project:** Palliative Care and End-of-Life Care

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)***:**

Pain is under-recognized by clinicians and undertreated, resulting in excess suffering for patients with serious illness. Pain screening and assessment are necessary processes in order to improve the patient-centered outcome of pain, and its effects on global outcomes of function and quality of life. Pain, like other symptoms, can only be understood through patient self-report and patient observation. Screening and assessment for pain are essential steps in pain management. Without initial screening to identify patients in pain, and clinical assessment to determine the severity, etiology, and effect on function of this symptom, effective treatment cannot be administered. Additional guidelines from American College of Physicians and the American Pain Society recommend systematic pain screening and assessment. Additional evidence includes numerous systematic reviews and individual studies.

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

Clinical Practice Guideline

Selected individual studies (rather than entire body of evidence)

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)***:**

Pain is a highly prevalent, distressing, and functionally limiting symptom common to many serious illness conditions, and its relief is an important priority for patients and families.

a)Strong evidence supports the effectiveness of medical treatment for pain in cancer and in other serious illnesses to improve pain outcomes (randomized controlled trials, systematic reviews)

b)Moderate quality evidence supports the effectiveness of expert pain assessment and structural innovations such as specialty palliative care teams to improve pain outcomes. Published after the most recent systematic reviews, 3 new randomized trials and 1 controlled observational study report interventions enhancing the structure and process of palliative care delivery. Two report improved pain and quality of life outcomes (Casarett, Temel), one reports improved quality of life but no change in symptom scores (Bakitas), and one reports no change in symptom scores (Gade) (randomized and non-randomized trials, systematic reviews, individual studies)

c)Three comprehensive practice guidelines support the importance of screening, assessing, and treating pain for seriously ill and terminally ill patient populations with a wide range of diagnoses.

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)*: 1. Systematic review of varied strategies in palliative care (Lorenz, 2004, 2008): reported results for pain management of 9 systematic reviews, 24 individual studies of interventions

2. Systematic review of RCTs of specialty palliative care services(Zimmerman, 2008): systematic review of 22 randomized trials of specialty palliative care effects on various outcomes, including symptom distress.

3. 2 additional guidelines from American College of Physicians, American Pain Society recommend systematic pain screening and assessment

4. 3 additional RCTs and 1 controlled observational study of palliative care interventions (Gade, Casarett, Bakitas, Temel)

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):* a) Multiple high quality randomized trials of pain treatment, and high quality systematic reviews address interventions to improve pain management and outcomes.

b) Studies of pain treatment include expert screening and assessment in the protocols, but do not define the effect of these processes alone on pain outcomes. Studies of complex structural and process interventions to improve specialty palliative care show varied effects on pain outcomes; however, these interventions are complex and heterogeneous in design, and provide less direct evidence for targeted interventions to improve pain as a primary outcome.

c) Included studies range in size, but many of the highest quality randomized trials have adequate power for hypothesis testing.

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* Evidence from studies explicitly targeting pain outcomes with more discrete interventions is very consistent, and provides strong evidence for improved outcomes. Evidence from studies of complex palliative care interventions shows less consistent effect on pain outcomes, and pain is used as a secondary rather than a primary outcome measure.

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

Net benefit for pain screening and assessment is strong, and clearly outweighs potential harms. Benefit is high due to the high numbers of patients affected and the marked suffering experienced as a result of current under-reporting and undertreatment, a problem more marked among patients of minority race or ethnicity. Potential harms from quality measures focused on pain management include over-treatment with medication toxicities and inattention to other symptoms. These harms are reduced in the context of hospice and specialty palliative care, delivered by professional teams with appropriate expertise and training.

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? Yes

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: (Lorenz, 2008) Strong evidence for treatment of cancer pain. Weak evidence

supports complex interventions such as multidisciplinary teams.

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

1c.12 If other, identify and describe the grading scale with definitions:

1c.13 Grade Assigned to the Body of Evidence: See 1c.10.

1c.14 Summary of Controversy/Contradictory Evidence: N/A

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

OTHER GUIDELINES:

1. Qaseem A, Snow V, Shekelle P et al. Evidence-based interventions to improve the palliative care of pain, dyspnea and depression at the end of life: a clinical practice guideline from the American College of Physicians. Ann Intern Med 2008; 148:141-146.

2. Gordon DB, Dahl JL, Miaskowski C et al. American Pain Society recommendations for improving the quality of acute and cancer pain management. Arch Intern Med 2005; 165:1574-1580.

3. Medical Services Commission. Palliative care for the patient with incurable cancer or advanced disease. Part 2: pain and symptom management. Victoria (BC): British Columbia Medical Services Commission; 2011 Sep 30. 44 p.

SYSTEMATIC REVIEWS:

3. Lorenz KA, Lynn J, Dy SM et al. Evidence for improving palliative care at the end of life: a systematic review. Ann Intern Med 2008; 148:147-159.

4. Lorenz KA, Lynn J et al. End-of-life care and outcomes. AHRQ Publication No. 05-E004-2, December 2004.

5. Zimmerman C, Riechelmann R, Krzyzanowska M et al. Effectiveness of specialized palliative care: a systematic review. JAMA 2008; 299:1698-1709.

ADDITIONAL INDIVIDUAL STUDIES

6. Bakitas M, Lyons KD, Hegel MT et al. Effects of a palliative care intervention on clinical outcomes in patients with advanced cancer: the Project ENABLE II randomized controlled trial. JAMA 2009; 302:741-749.

7. Casarett D, Pickard A, Bailey FA et al. Do palliative consultations improve patient outcomes? J Am Geriatr Soc 2008; 56:593-599.

8. Temel JS, Greer JA, Muzikansky A et al. Early palliative care for patients with metastatic non-small-cell lung cancer. N Engl J Med 2010; 363:733-742.

9. Gade G, Venohr I, Conner D et al. Impact of an inpatient palliative care team: a randomized controlled trial. J Palliat Med 2008; 11:180-190.

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

Health Care Guideline: Palliative Care for Adults

Guideline 4. Physical Aspects of Care: The physical aspects of the patient's serious illness should be an integral component of the palliative care plan. (p.21)

Guideline 2.1 Pain, other symptoms, and side effects are managed based upon the best available evidence, with attention to disease-specific pain and symptoms, which is skillfully and systematically applied.

• Regular, ongoing assessment of pain, non-pain symptoms (including but not limited to shortness of breath, nausea, fatigue and weakness, anorexia, insomnia, anxiety, depression, confusion, and constipation), treatment side effects, and functional capacities are documented through a systematic process. Validated instruments, where available, should be utilized. Symptom assessment in children and cognitively impaired patients should be performed by appropriately trained professionals with appropriate tools.

1c.17 **Clinical Practice Guideline Citation:** Palliative Care for Adults: McCusker M, Ceronsky L, Crone C, Epstein H, Greene B, Halvorson J, Kephart K, Mallen E, Nosan B, Rohr M, Rosenberg E, Ruff R, Schlecht K, Setterlund L. Institute for Clinical Systems Improvement. Palliative Care for Adults. Updated November 2013.p.21

1c.18 **National Guideline Clearinghouse or other URL:**

Palliative Care for Adults: http://www.guideline.gov/content.aspx?id=47629

1c.19 **Grading of Strength of Guideline Recommendation.** Has the recommendation been graded? Yes

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

Palliative Care for Adults: Institute for Clinical Systems Improvement (ICSI)

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

Palliative Care for Adults: GRADE Metholodogy

1c.22 **If other, identify and describe the grading scale with definitions:**

1c.23 **Grade Assigned to the Recommendation:**

Palliative Care for Adults: Low Quality Evidence, Strong Recommendation

1c.24 **Rationale for Using this Guideline Over Others:** We chose to use the Palliative Care for Adults guideline from the Institute for Clinical Systems Improvement (ICSI) because it is evidenced based, uses GRADE methodology, and is included in the National Guideline Clearinghouse.

In addition, the National Consensus Project for Quality Palliative Care was the first United States national guidelines development project for palliative care quality, inclusive of hospice care. This set of guidelines, along with 38 preferred practices, has been rigorously reviewed and endorsed by the National Quality Forum. Although specific investigative

groups and specialty organizations have published other guidelines in pain management or hospice or palliative care practice for specific settings, practices or populations, none have been as comprehensive or comprehensively debated and peer reviewed.

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 #: 1634
De.2. **Measure Title:** Hospice and Palliative Care -- Pain Screening
Co.1.1. **Measure Steward:** University of North Carolina-Chapel Hill
De.3. **Brief Description of Measure:** Percentage of hospice or palliative care patients who were screened for pain during the hospice admission evaluation / palliative care initial encounter.
1b.1. **Developer Rationale:** Pain is prevalent and undertreated for many populations of seriously ill patients, including those patients nearing the end of life. Poor screening, assessment and undertreatment of pain is more common for patients with serious illness who are also of minority race ethnicity. Use of the Pain Screening and Pain Assessment quality measures will increase reporting and efforts to improve awareness of the presence of pain (screening) and assessment of severity, etiology and effect on function (assessment) which are the two essential first steps required for quality pain management and treatment.

S.4. **Numerator Statement:** Patients who are screened for the presence or absence of pain (and if present, rating of its severity) using a standardized quantitative tool during the admission evaluation for hospice / initial encounter for palliative care.
S.7. **Denominator Statement:** Patients enrolled in hospice OR patients receiving specialty palliative care in an acute hospital setting.
S.10. **Denominator Exclusions:** Patients with length of stay < 1 day in palliative care.

De.1. **Measure Type:** Process
S.23. **Data Source:** Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record
S.26. **Level of Analysis:** Clinician : Group/Practice, 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?** Part of the PEACE Measures Set
Paired with Hospice and Palliative Care - Pain Assessment (percentage of hospice or palliative care patients who screen positive for pain and who received a clinical assessment of pain within 24 hours of screening.

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
1634_Evidence_MSF5.0_Data.doc,1634_Evidence_3.17.16.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)*
Pain is prevalent and undertreated for many populations of seriously ill patients, including those patients nearing the end of life. Poor screening, assessment and undertreatment of pain is more common for patients with serious illness who are also of minority race ethnicity. Use of the Pain Screening and Pain Assessment quality measures will increase reporting and efforts to improve awareness of the presence of pain (screening) and assessment of severity, etiology and effect on function (assessment) which are the two essential first steps required for quality pain management and treatment.

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.*
Hospice: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1634 - Hospice and Palliative Care Pain Screening.

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.

The mean score for this QM was 93.5% with a range from 0% to 100%, the median was 96.8%, the interquartile range was 7.6, and the standard deviation was 9.3. For this QM, 20.1% of hospices had perfect scores and 20.7% of hospices scored below 90%.

Scores by decile:
10th percentile 82.6%
25th percentile 91.5%
Median 96.8%
75th percentile 99.4%
90th percentile 100%

Palliative Care:

This submission to the Palliative and End-of-Life Care project updates hospice setting data for NQF #s 1634, 1637, 1638, 1639, 1641, 1647. We are currently in the process of updating palliative care data by collecting and analyzing data in multiple non-hospice settings but final analyses are not available for this submission cycle.(1) Data comes from two sources -- a multi-site study of quality of care in palliative care (R18HS022763 Implementing Best Practice in Palliative Care, PI Johnson) and (CMS Health Care Innovation Award: Increasing patient and system value with community based palliative care, PI Bull / Four Seasons Compassion for Life). We anticipate these data will become available for NQF review next year. This data will allow further updates to the evidence base for non-hospice palliative care beyond what is currently submitted.

(1) Kamal AH, Bull J, Ritchie CS, Kutner JS, Hanson LC, Friedman F, Taylor DH Jr; AAHPM Research Committee Writing Group. Adherence to Measuring What Matters measures using point-of-care data collection across settings. J Pain Symptom Manage 2016; 51:497-503.

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.**
Pain is prevalent, underdiagnosed and undertreated in cancer and other life-limiting or serious illnesses. The prevalence of pain ranges from 40-80% in seriously ill patient populations. As detailed in a systematic review from AHRQ and the American Pain Society Quality of Care guidelines, pain screening and assessment are the essential steps required to ensure that pain is detected by clinicians and appropriate treatment implemented.(1,2) Failure to screen, assess, and treat pain results in functional limitations, physiologic stress, and psychological harms such as social withdrawal and depression.
The current quality of pain screening, assessment, and treatment is poor, as documented in systematic pain prevalence and treatment studies from hospital, outpatient, cancer and nursing home settings. (3,4,5,6) In a systematic review of quality of pain care for diverse patient populations, Gordon reported high average pain severity (6.17-8.37 on 10 point scale) and moderate rates of pain

severity screening or other assessment (47%-96%). These findings did not vary by underlying diagnosis. (7)

1. Wells N, Pasero C, McCaffery M. Improving the Quality of Care through Pain Assessment and Management. In: Hughes RG, editor. Patient Safety and Quality: An Evidence-Based Handbook for Nurses. Rockville (MD): Agency for Healthcare Research and Quality (US); 2008 Apr. Chapter 17.
2. Gordon DB, Dahl JL, Miaskowski C et al. American Pain Society recommendations for improving the quality of acute and cancer pain management. Arch Intern Med 2005; 165:1574-1580.
3. Reynolds K, Henderson M, Schulman A, Hanson LC. Needs of the dying in nursing homes. J Pall Med 2002; 5:895-901.
4. Deandria S, Montanri M, Moja L et al. Prevalence of undertreatment of cancer pain: a review of published literature. Ann Oncol 2008; 19:1985-91.
5. Mularski R, White-Chu F, Overbay D et al. Measuring pain as the 5th vital sign does not improve quality of pain management. J Gen Intern Med 2006; 6:607-612.
6. Erdek MA, Pronovost PA. Improving assessment and treatment of pain in the critically ill. Int J Qual Health Care 2004; 16:59-64.
7. Gordon DB, Pelliano TA, Miaskowski C et al. A 10-year review of quality improvement monitoring in pain management: recommendations for standardized outcome measures. Pain Manage Nurs 2002; 4:116-130.

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.*
Hospice:

Racial disparity analysis. We conducted racial and ethnic disparity analyses at both the patient-stay and hospice levels. At the patient-stay level, we compared the percentage of patients who received pain screening within 2 days after admission among different racial and ethnic groups. Patients were grouped by their racial and ethnic identification as follows: white non-Hispanic, black non-Hispanic, other non-Hispanic, or Hispanic. Other non-Hispanic includes patients who identify as American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, or who identify as more than one race. A Chi-square test was performed to determine if there were any statistically significant differences in pain screening between groups. The lowest rate of pain screening was found for Hispanic patients (91.3%), and the highest rate was among patients identifying as other non-Hispanic (94.3%). Differences in the rate of pain screening by racial identification were found to be statistically significant ($p<0.001$).

Analyses at the hospice level examined the differences in this measure across two groups: hospices with proportions of non-white patients that are greater than or equal to the national median proportion (11.9%), and hospices with fewer non-white patients than the national median. For this analysis, white non-Hispanic patients were included in the white group, and all other racial and ethnic identifications, as described above, were grouped as non-white. We ran a Wilcoxon-Mann-Whitney test for statistical dependence between group and QM score. The results showed that the QM score was not significantly different between the two groups of hospices (93.7% compared to 93.3%, $p=0.45$).

Gender disparity analysis. We performed both the patient stay- and hospice-level analyses on gender disparity. At the patient-stay level, we compared the percentage of patients who received pain screening within 2 days after admission between female and male patients. A Chi-square test was performed to determine if there were any statistically significant differences in pain screening between groups. We found a slightly lower rate of pain screening for female patients (93.1%) than for male patients (93.4%). Differences in the rate of pain screening by gender were statistically significant ($p<0.001$). At the hospice level, we examined the differences in this measure across hospices with proportions of female patients that are greater than or equal to the national median proportion (55.2%). We ran a Wilcoxon-Mann-Whitney test for statistical dependence between group and QM score. The results showed that the QM score was statistically significantly different between the two groups of hospices split by median proportion of female patients (93.3% compared to 93.7%, $p=0.005$). Although statistically significant results were found at both the patient and hospice level, actual differences in screening rates do not seem to be clinically substantial.

Socioeconomic disparity analysis. We performed socioeconomic disparity analyses at both the patient-stay and hospice levels using the same methods described in previous disparity analyses. Medicaid status was used as a proxy measure of low socioeconomic

status. A statistically significant (p<0.001) lower rate of pain screening was found for non-Medicaid patients (92.5%) than for Medicaid patients (93.6%). The highest rate of pain screening was seen for patients with Medicaid status missing (94.9%). At the hospice level, the results showed that the QM score was significantly different between the two groups of hospices split by median proportion of Medicaid patients, 21.5% (92.1% compared to 94.6%, p<0.001). The statistically significant results at the patient-stay level may indicate that quality of hospice care, measured by pain screening within 2 days after admission, for non-Medicaid patients is lower than for Medicaid patients. The significant findings at the hospice level indicate that hospices with a smaller proportion of Medicaid patients are less likely to provide pain screening to patients at admission. A potential caveat of this analysis is that the data used for this analysis may not be a reliable indicator of a patient's Medicaid eligibility status. This data is missing for nearly a quarter of patients. In addition, some of the Medicaid numbers submitted do not appear to be legitimate Medicaid numbers. However, the data used in this analysis is the only national patient-level assessment data that is currently available in hospices. We will update this analysis as more-accurate data sources are available and accessible.

Rural-urban disparity analysis. We compared the QM score between rural and urban hospices using a Wilcoxon-Mann-Whitney test. The results showed that the QM score was not significantly different between rural and urban hospices (93.2% compared to 93.6%, p=0.98). This indicates that rural hospices and urban hospices perform similarly on this quality measure.

Palliative Care: Disparities data not available.

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.**
Extensive evidence documents disparities in cancer pain treatment and control.(1,2) Nursing home residents with advanced cancer receive less effective pain treatment if they are African American.(3,4) The Eastern Cooperative Oncology Group Minority Outpatient Pain Study enrolled 1308 patients with advanced cancer. After clinic visits, physicians underestimated pain severity for 64% of Hispanic and 74% of African American patients.(5) Among patients with pain, 65% of Hispanic and African American patients received inadequate treatment relative to practice guidelines, as did 50% of white patients.(6,7)

1.Pletcher MJ, Kertesz SG, Kohn MA, Gonzales R. Trends in opioids prescribing by race for patients seeking care in US emergency departments. JAMA 2008; 299:70-78.

2.Green CR, Montague L, Hart-Johnson TA. Consistent and breakthrough pain in diverse advanced cancer patients: a longitudinal examination. J Pain Sympt Manage 2009; 37:831-847.

3.Bernabei R, Gambassi G, Lapane K, Landi F, Gatsonis C, Dunlop R, Lipsitz L, Steel K, Mor V. Management of pain in elderly patients with cancer. SAGE Study Group. JAMA 1998; 279:1877-82.

4.Engle VF, Fox-Hill E, Graney MJ. The experience of living-dying in a nursing home: self-reports of black and white older adults. JAGS 1998; 46:1091-96.

5.Anderson KO, Mendoza TR, Valero V, Richman SP, Russell C, Hurley J, DeLeon C, Washington P, Palos G, Payne R, Cleeland CS. Minority cancer patients and their providers: pain management attitudes and practices. Cancer 2000; 88: 1929-38.

6.Cleeland CS, Gonin R, Baez L et al. Pain and treatment of pain in minority patients with cancer. The ECOG Minority Outpatient Pain Study. Ann Intern Med 1997; 127:813-16.

7.Cleeland CS, Gonin R, Hatfield AD et al. Pain and its treatment in outpatients with metastatic cancer. N Engl J Med 1994; 330:592-96.

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**
Affects large numbers, Patient/societal consequences of poor quality, Severity of illness

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.**
The Hospice and Palliative Care - Pain Screening measure addresses pain for patients with high severity of illness and risk of death, including seriously and incurably ill patients enrolled in hospice or hospital-based palliative care. Research on care of patients with serious incurable illness and those nearing the end of life shows they experience high rates of pain (40-70% prevalence) and other physical, emotional, and spiritual causes of distress. (1,2) The National Priorities Partnership has identified palliative and end-of-life care as one of its national priorities. A goal of this priority is to ensure that all patients with life-limiting illness have access to effective treatment for symptoms such as pain and shortness of breath.(3) The affected populations are large; in 2014, 1.66 million people with life-limiting illness received hospice care.(4) In 2013, 67% of US hospitals with 50 or more beds had some form of palliative care service, up from 58.5% in 2008, and national trends show steady expansion of these services.(5) Patients and family caregivers rate pain management as a high priority when living with serious and life-limiting illnesses. (6) The consequences of inadequate screening, assessment and treatment for pain include physical suffering, functional limitation, and development of apathy and depression. (7)

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
1. The Writing Group for the SUPPORT Investigators. A controlled trial to improve care for seriously ill hospitalized patients. The study to understand prognosis and preferences for outcomes and risks of treatments (SUPPORT). JAMA. 1995;274:1591-1598.

2. Gade G, Venohr I, Conner D, et al. Impact of an inpatient palliative care team: a randomized control trial. J Palliat Med. 2008;11(2):180–190.

3. http://www.nationalprioritiespartnership.org/PriorityDetails.aspx?id=608

4. NHPCO Facts and figures: hospice care in America 2015 edition
http://www.nhpco.org/hospice-statistics-research-press-room/facts-hospice-and-palliative-care

5. Dumanovsky T, Augustin R, Rogers M, Lettang K, Morrison RS. The growth of palliative care in U.S. hospitals: a status report. J Pall Med. 2016; 19(1): 8-15

6. Singer PA, Martin DK, Kelner M. Quality end-of-life care: patients´ perspective. JAMA 1999; 281: 163-168.

7. Gordon DB, Dahl JL, Miaskowski C et al. American Pain Society recommendations for improving the quality of acute and cancer pain management. Arch Intern Med 2005; 165:1574-1580.

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):*

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.)*
PEACE Hospice and Palliative Care Quality Measures: http://www.med.unc.edu/pcare/resources/PEACE-Quality-Measures

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.
We would like to propose the removal of the less than 7 day length of stay (LOS) denominator exclusion for hospice patients.

Background: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1634 - Hospice and Palliative Care Pain Screening.

Six of the seven QMs exclude patient stays that are less than 7 days from the measure denominator. When the length of stay (LOS) is too short, hospices may not have enough time to complete all the clinically recommended care processes. Thus, at the time the measures were developed, technical experts recommended that short patient stays be excluded from those measure denominators for assessing quality of care in hospices. However, no national data regarding the implications of the LOS exclusion was available to the Technical Expert Panel (TEP) at that time. Under contract to the Centers for Medicare and Medicaid Services (CMS), the Research Triangle Institute (RTI) performed descriptive analyses on 5 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.

Rationale for inclusion of all hospice patients regardless of LOS: At the patient level, approximately 40% of patients were excluded based on the LOS < 7 days exclusion, thus omitting this essential quality of care standard for many seriously ill patients. Further, excluding short-stay patients omits measures of quality of care for a subset of the sickest patients entering hospice care, and those who arguably may be in greatest need of rapid symptom assessment and treatment. In addition, the original rationale for this exclusion -- allowing time for hospice providers to complete this care process -- does not appear to be necessary. Our analyses show that a large portion of pain screenings were performed on day 1 of admission to hospice, demonstrating a normative standard of care includes prompt attention to symptom distress.

At the hospice level, applying different LOS exclusion criteria had an impact on denominator size. Smaller denominator size at the hospice level may generate less stable and reliable QM scores. Under the LOS < 7 days exclusion, the median number of qualifying stays in the denominator was 94 stays per hospice vs. a median number of 136 stays per hospice with no LOS exclusion. Applying the LOS exclusion resulted in some hospices with no qualifying stays, which excluded 14 hospices from QM score calculations altogether.

Applying or removing the LOS exclusion generally had little impact on the distribution of hospices' QM scores. Under the 7-day LOS exclusion, the mean score was 94.4%, the median score was 97.6%, and the score for hospices in the 10th percentile distribution was 84.8%. With no LOS exclusions, the mean score was 93.5%, the median score was 96.8%, and the score for hospices in the 10th percentile distribution was 82.6%. The impact of the different LOS criteria on the distribution of QM scores was consistent across quarters.

In summary, these new analyses demonstrate that removing the LOS exclusion in QM calculations, thus including all hospice patients in the denominator, is feasible and appropriate for quality of care at the patient level, and enhances completeness and statistical stability of QM reporting at the hospice level. Further, though the mean and median scores on this QM are high, hospice

organizations in the 10th percentile have meaningful opportunity to improve in their timely screening of pain. Since this quality measure is paired, successful screening is linked to subsequent measurement of pain assessment (NQF #1637).

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 who are screened for the presence or absence of pain (and if present, rating of its severity) using a standardized quantitative tool during the admission evaluation for hospice / initial encounter for palliative care.

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.)*
Hospice admission evaluation / initial clinical encounter for palliative care

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.
Patients who are screened for the presence or absence of pain (and if present, rating of its severity) using a standardized tool during the admission evaluation for hospice / initial encounter for hospital-based palliative care. Screening may be completed using verbal, numeric, visual analog, rating scales designed for use the non-verbal patients, or other standardized tools.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Patients enrolled in hospice OR patients receiving specialty palliative care in an acute hospital setting.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, 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)*
The Pain Screening quality measure is intended for patients with serious illness who are enrolled in hospice care OR receive specialty palliative care in an acute hospital setting. Conditions may include, but are not limited to: cancer, heart disease, pulmonary disease, dementia and other progressive neurodegenerative diseases, stroke, HIV/AIDS, and advanced renal or hepatic failure.

[NOTE: This quality measure should be paired with the Pain Assessment quality measure (NQF #1637) to ensure that all patients who report significant pain are clinically assessed.]

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Patients with length of stay < 1 day in palliative care.

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)*
Calculation of length of stay: discharge date is identical to date of initial encounter.

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)*
N/A

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*)

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.)*
Screened for pain:

a. Step 1- Identify all patients with serious, life-limiting illness who are enrolled in hospice OR received specialty palliative care in an acute hospital setting.

b. Step 2- Exclude palliative care patients if length of stay is < 1 day.

c. Step 3- Identify patients who were screened for pain during the admission evaluation (hospice) OR initial encounter (palliative care) using a standardized tool.

Quality Measure =
Numerator: Patients screened for pain in Step 3 / Denominator: Patients in Step 1-Patients excluded in Step 2

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.
Hospice: The hospice analysis was not based on a sample. It was conducted on the entire hospice population that had admission and discharge records in the specified period of analysis.

Palliative Care: consecutive sample of equal numbers of admissions + decedent records beginning with a randomly selected date.

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.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Hospice: For the pain screening measure, there are three items on the HIS that can include missing data – J0900B, J0900C, and J0900D. If missing, these items are coded as dashes and analyzed as though the process did not happen.

In order to assess how these missing data impact the validity of the pain screening measure, we conducted the patient stay- and hospice-level analyses. For the patient stay-level analysis, we calculated the number and percentage of eligible patient stays for which the HIS-Admission records included a dash for each of these three items. The overall rate of missing data ranged from 0.02 percent to 0.11 percent between October 2014 and September 2015. For the hospice-level analysis we calculated each hospice's percent of eligible admissions that included missing data for items J0900B, J0900C, and J0900D. Over 90% of hospices did not have any admissions with missing data for these items.
Overall, we found that only a very small number of admission records for the eligible stays did not include data, i.e. coded as dash, on pain screening items J0900B, J0900C, and J0900D. And a vast majority of hospices did not have any missing data for these items. These results indicate that missing data for these items should not have a negative impact on the validity of the QM.

Palliative care: N/A

S.23. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record

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.
Hospice: Hospice analysis uses the Hospice Item Set (HIS) as the data source to calculate the quality measure.

Palliative Care: Structured medical record abstraction tool with separate collection of numerator and denominator data values.

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 in attached appendix at A.1

S.26. Level of Analysis *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Group/Practice, Facility

S.27. Care Setting *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospice, 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
1634_MeasureTesting_MSF5.0_Data.doc,1634_MeasureTestingAttachment_2.25.16.docx

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 1634	**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)***:**

Hospice: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1634 - Hospice and Palliative Care Pain Screening.

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.

Palliative Care: Two research nurse abstractors independently recorded quality measures data on a random subset of 20 seriously ill patients. Abstractors used the pre-defined operational definitions and a structured chart abstraction tool to record numerator and denominator data separately. Patients were a subsample of 460 seriously ill patients without specialty palliative care admitted to an acute care hospital for at least 1 day to 4 inpatient services. Records eligible for sampling included all seriously ill adult patients admitted to medical and surgical intensive care, medically complex patients aged 65 and older admitted to an Acute Care of the Elderly Unit, and medical oncology patients with Stage IV carcinoma.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)***:**

Hospice:

Split-half reliability. Split-half reliability assesses the internal consistency of a QM in different samples by randomly dividing the patient stays within each hospice into two halves, and calculating correlation between the QM scores on the basis of the two randomly divided halves. In this analysis, we conducted a split-half reliability analysis on all facilities with 20 or more patient stays counted in the measure denominator, and used the Interclass Correlation (ICC) coefficients to measure the internal reliability. In general, the ICC coefficient varies between 0 and 1, where an ICC of 0 indicates no reliability and an ICC of 1 indicates perfect reliability.

Signal-to-noise analysis. If a measure is reliable, then true differences in provider performance should explain a substantial proportion of the variance in QM scores. We conducted an analysis of variance (ANOVA) to determine what proportion of total variance in the measure is attributable to differences among providers. A higher proportion indicates better reliability.

Stability analysis. Stability analysis describes the extent to which providers' performance assessed by a QM changes across time. We analyzed the change in facility scores between four consecutive quarters (Q4 2014 and Q1 2015, Q1 and Q2 2015, and Q2 and Q3 2015). The changes in facility scores are reported in standard deviations.

Palliative Care: Inter-rater reliability between the two abstractors was assessed using kappa statistics. Two research nurse abstractors independently recorded quality measures data on a random subset of 20 seriously ill patients. Abstractors used the pre-defined operational definitions and a structured chart abstraction tool to record numerator and denominator data separately.

2a2.3 **Testing Results** *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**

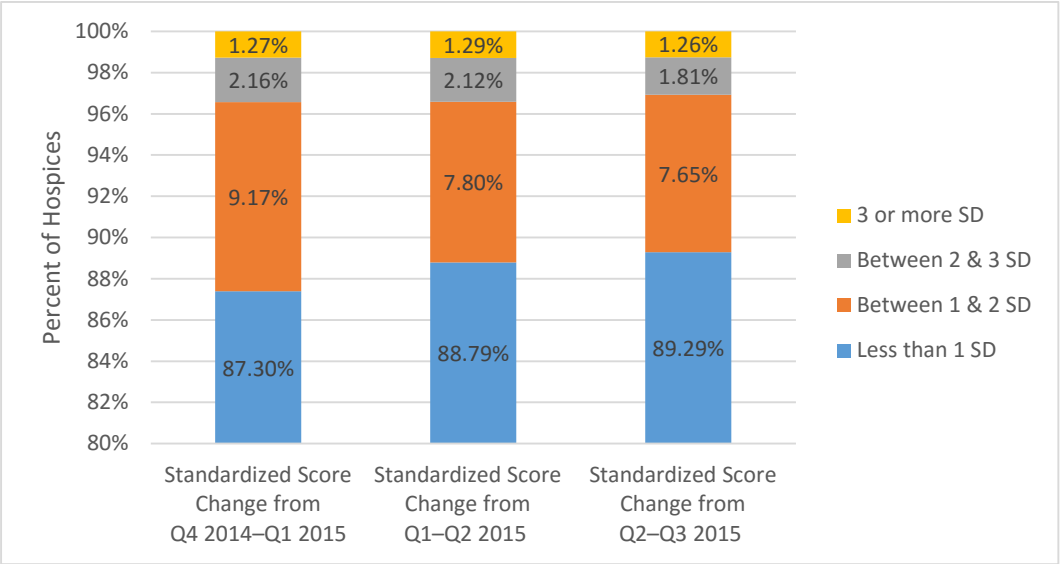
Hospice:

Split-half reliability: In general, the ICC coefficient varies between 0 and 1, where an ICC of 0 indicates no reliability and an ICC of 1 indicates perfect reliability. The ICC coefficient for this measure is 0.86, indicating high internal reliability.

Signal-to-noise analysis: The analysis results found the signal-to-noise ratio to be 0.97, indicating that about 97% of the variance in this measure is because of differences among facilities. This proportion indicates strong reliability for this measure.

Stability analysis: The results of this analysis indicated that facility scores were very stable. Slightly less than 90% of facilities had a change in QM score of less than one standard deviation, indicating high stability of the QM. The number of facilities with a change in QM of less than one standard deviation increased across quarters from 87.3% to 89.3%, suggesting improved reliability across time. Roughly one-tenth of facilities had a change in QM between one and two standard deviations. These results indicate a measure that is generally quite stable. Figure 1 illustrates the change in facility scores between the four consecutive quarters.

Figure 1
Standardized Score Change in QM Score from Q4 2014 to Q3 2015



Palliative Care: Kappa scores range from 0 to 1 with higher scores indicating better agreement. The nurse abstractors achieved excellent inter-rater reliability for this measure with Kappa=1.0 Landis and Koch describe kappa values that range from 0.81 – 0.99 as almost perfect and Fleiss describes kappas over 0.75 as excellent.

Landis, J.R.; Koch, G.G. (1977). "The measurement of observer agreement for categorical data". Biometrics 33 (1): 159–174

Fleiss, J.L. (1981). Statistical methods for rates and proportions (2nd ed.). New York: John Wiley

2b. VALIDITY. Validity, Testing, including all Threats to Validity: 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:**

The measure focus is Pain Screening, designed to pair with Pain Assessment (NQF # 1637) to ensure quality care processes for pain. The target populations are hospice patients, and seriously ill hospitalized patients with diverse underlying diagnoses who are at high risk for palliative care clinical needs, including pain.

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)***:**

Hospice: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1634 - Hospice and Palliative Care Pain Screening.

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.

Palliative Care: The total patient sample size was 562. Chart abstractions were completed for 102 consecutive seriously ill patients with specialty palliative care consultation, and a random sample of 460 seriously ill patients without specialty

palliative care admitted to an acute care hospital for at least 1 day to four inpatient services with high proportions of seriously ill patients. Records eligible for sampling included all patients admitted to medical and surgical intensive care, medically complex patients aged 65 and older admitted to a Geriatric Evaluation Unit, and medical oncology patients with Stage IV carcinoma. Because palliative care domains become even more relevant closer to death, patients dying in hospital were over-sampled to ensure a final ratio of 1 decedent to 1 live discharge. Consistent with oversampling of decedent records, 55% of these patients died in hospital.

The age of the patients ranged from 16 to 99 years, with the mean age 61. Patients were predominantly Caucasian (65%), with smaller subgroups who were African American (24%) and Hispanic / Latino (4%) The most common life-limiting diagnoses were infections (37%), cancer (34%), pulmonary (29%), and neurologic diseases (21%).

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

Hospice:

Correlations. Providers should perform similarly on QMs that reflect the quality of similar care processes. Thus, a common strategy used to evaluate validity of a QM is to examine the relative performance of the provider on QMs that are conceptually related, either due to some core competency manifested across a stay or for processes that occur in close proximity. We conducted nonparametric Spearman rank correlation analysis among all seven HQRP QMs, which address care processes around hospice admission that are clinically recommended or required in the hospice Conditions of Participation. We used nonparametric methods since the data are heavily skewed (most providers perform well, while a small proportion have lower scores). Since all of the seven current QMs measure desirable care assessment processes upon hospice admission, higher positive Spearman correlation coefficients among these QMs will demonstrate higher validity of the QM set. Statistically significant correlations, indicated by p-values of the Spearman correlation coefficients less than 0.05, also support the validity of the QM set.

Missing data. Missing data represent a potential threat to the validity of an HIS item, which in turn may harm the validity of the QM that relies on the item. For the pain screening measure, there are three items on the HIS that can include missing data, which are coded as dashes (J0900B, J0900C, and J0900D). In order to assess how these missing data impact the validity of the pain screening measure, we conducted the patient stay- and hospice-level analyses.

Palliative Care sample:

Face validity of PEACE quality measures for hospital-based specialty palliative care was addressed using stakeholder review and feedback. Investigators prepared data reports in a summary format with detailed operational definitions, and led a 1-hour discussion with nursing and physician leaders from each service group – MICU, SICU, Acute Care for the Elderly (Geriatrics), Oncology, and Palliative Care. The discussion included feedback of quality measure data, response to questions and critiques, and eliciting stakeholder feedback about the validity and actionability of this data for the care of their patients. Stakeholders were specifically asked to comment on the accuracy of the data as a reflection of current care practices, and their highest priority area for future quality improvement.

Construct validity was tested by comparing the PEACE quality measures for patients seen by specialty interdisciplinary palliative care consultants to those not receiving specialty palliative care services.

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)***:**

Hospice:

Correlations: Table 1 presents the Spearman correlation coefficients. The p-values for all the Spearman correlation coefficients are significant ($p < 0.001$). The significant positive correlations between every pair of QMs indicate that high-performing hospices in one area also provided high-quality care in other areas at hospice admission. Overall, the correlations between the QMs are low to moderate. The clustering of QM scores, skewed distributions, and low variability across all seven QMs may affect the level of correlations between QMs.

Table 1
Correlation of Hospice QMs, Percentile Ranking

Quality measure	NQF #1647 (modified) Beliefs/Values Addressed	NQF #1634 Pain Screening	NQF #1637 Pain Assessment	NQF #1639 Dyspnea Screening	NQF #1638 Dyspnea Treatment	NQF #1617 Bowel Regimen
NQF #1641 Treatment Preferences	0.64***	0.37***	0.21***	0.44***	0.28***	0.25***
NQF #1647 (modified) Beliefs/Values Addressed		0.42***	0.17***	0.43***	0.31***	0.25***
NQF #1634 Pain Screening			0.17***	0.41***	0.28***	0.15***
NQF #1637 Pain Assessment				0.17***	0.08***	0.20***
NQF #1639 Dyspnea Screening					0.30***	0.25***
NQF #1638 Dyspnea Treatment						0.34***

NOTE: The correlation is on the basis of each hospice's percentile ranking on the QMs.
*** indicates significant correlation at $p < 0.001$

Overall, NQF #1634, Pain Screening, has significant positive correlations with the other QMs, indicating hospices providing higher-quality care in this area also performed better in other areas at hospice admission. However, the correlation coefficients are low, indicating relatively weaker relationships. The conceptual tie between performance on pain screening and other measures may be low.

Missing data. For the patient stay-level analysis, we calculated the number and percentage of eligible patient stays for which the HIS-Admission records included a dash for each of these three items. The overall rate of missing data ranged from 0.02 percent to 0.11 percent between October 2014 and September 2015. For the hospice-level analysis we calculated each hospice's percent of eligible admissions that included missing data for items J0900B, J0900C, and J0900D. Over 90% of hospices did not have any admissions with missing data for these items.

Overall, we found that only a very small number of admission records for the eligible stays did not include data, i.e. coded as dash, on pain screening items J0900B, J0900C, and J0900D. And a vast majority of hospices did not have any missing data for these items. These results indicate that missing data for these items should not have a negative impact on the validity of the QM.

Palliative Care sample:

Face Validity: Stakeholder discussions provided broad endorsement of face validity, with some considerations for specific patient populations. Intensive care and geriatrics clinicians endorsed the primary importance of pain screening and assessment, but expressed doubts about the validity of numerical pain severity ratings when used for nonverbal or confused

patients. Medical oncologists endorsed the face validity of these quality measures, but favored quality measures endorsed by oncology professional organizations.

Construct Validity: Screening for pain with a numerical pain scale was nearly universal for all seriously ill patients (99.5%), regardless of use of specialty palliative care, and half had moderate or severe pain. Patients with moderate or severe pain were more likely to have a clinical assessment of pain if seen by specialty palliative care (67% vs 42%, p=0.002).

CITATION: Schenck AP, Rokoske FS, Durham DD et al. The PEACE Project: identification of quality measures for hospice and palliative care. J Palliat Med 2010; 13:1451-1459.

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)***:**

Hospice: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1634 - Hospice and Palliative Care Pain Screening.

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.

Palliative Care: N/A

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

Hospice: Under contract to the Centers for Medicare and Medicaid Services (CMS), the Research Triangle Institute (RTI) performed descriptive analyses on 5 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.

Palliative Care: N/A

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

Hospice: At the patient level, approximately 40% of patients were excluded based on the LOS < 7 days exclusion, thus omitting this essential quality of care standard for many seriously ill patients. Further, excluding short-stay patients omits measures of quality of care for a subset of the sickest patients entering hospice care, and those who arguably may be in greatest need of rapid symptom assessment and treatment. In addition, the original rationale for this exclusion -- allowing time for hospice providers to complete this care process -- does not appear to be necessary. Our analyses show that a large portion of pain screenings were performed on day 1 of admission to hospice, demonstrating a normative standard of

care includes prompt attention to symptom distress.

At the hospice level, applying different LOS exclusion criteria had an impact on denominator size. Smaller denominator size at the hospice level may generate less stable and reliable QM scores. Under the LOS < 7 days exclusion, the median number of qualifying stays in the denominator was 94 stays per hospice vs. a median number of 136 stays per hospice with no LOS exclusion. Applying the LOS exclusion resulted in some hospices with no qualifying stays, which excluded 14 hospices from QM score calculations altogether.

Applying or removing the LOS exclusion generally had little impact on the distribution of hospices' QM scores. Under the 7-day LOS exclusion, the mean score was 94.4%, the median score was 97.6%, and the score for hospices in the 10th percentile distribution was 84.8%. With no LOS exclusions, the mean score was 93.5%, the median score was 96.8%, and the score for hospices in the 10th percentile distribution was 82.6%. The impact of the different LOS criteria on the distribution of QM scores was consistent across quarters.

In summary, these new analyses demonstrate that removing the LOS exclusion in QM calculations, thus including all hospice patients in the denominator, is feasible and appropriate for quality of care at the patient level, and enhances completeness and statistical stability of QM reporting at the hospice level. Further, though the mean and median scores on this QM are high, hospice organizations in the 10th percentile have meaningful opportunity to improve in their timely screening of pain. Since this quality measure is paired, successful screening is linked to subsequent measurement of pain assessment (NQF #1637).

Palliative Care: N/A

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)***:**

N/A

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

N/A

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)***:**

N/A

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

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):

Hospice: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1634 - Hospice and Palliative Care Pain Screening.

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.

Palliative Care: The total patient sample size was 562. Chart abstractions were completed for 102 consecutive seriously ill patients with specialty palliative care consultation, and a random sample of 460 seriously ill patients without specialty palliative care admitted to an acute care hospital for at least 1 day to four inpatient services with high proportions of seriously ill patients. Records eligible for sampling included all patients admitted to medical and surgical intensive care, medically complex patients aged 65 and older admitted to a Geriatric Evaluation Unit, and medical oncology patients with Stage IV carcinoma. Because palliative care domains become even more relevant closer to death, patients dying in hospital were over-sampled to ensure a final ratio of 1 decedent to 1 live discharge. Consistent with oversampling of decedent records, 55% of these patients died in hospital.

The age of the patients ranged from 16 to 99 years, with the mean age 61. Patients were predominantly Caucasian (65%), with smaller subgroups who were African American (24%) and Hispanic / Latino (4%) The most common life-limiting diagnoses were infections (37%), cancer (34%), pulmonary (29%), and neurologic diseases (21%).

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

Hospice: Confidence interval analysis. We examined proportions of hospices with the QM scores that are significantly different from the national hospice-level mean. If a high proportion of hospices have a measure score significantly different from the mean, the QM can identify facilities with different levels of performance. For this analysis, statistical significance was determined using 95 percent confidence intervals: a hospice's QM score was significantly different from the national mean if the national mean was not included within the hospice's 95 percent confidence interval. High-performing facilities should have scores that are significantly below average, and low-performing facilities should be significantly above average.

Palliative Care: Construct validity was tested by comparing the PEACE quality measures for patients seen by specialty interdisciplinary palliative care consultants to those not receiving specialty palliative care services. Percentage of patients with and without specialty palliative care for whom the quality measure was met was compared for difference using the chi-square statistic.

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)*:

Hospice:

The mean score for this QM was 93.5% with a range from 0% to 100%, the median was 96.8%, the interquartile range was 7.6, and the standard deviation was 9.3. For this QM, 20.1% of hospices had perfect scores and 20.7% of hospices scored below 90%.

Scores by decile:

10th percentile 82.6%

25th percentile 91.5%

Median 96.8%

75th percentile 99.4%

90th percentile 100%

Across all hospices, 50.7% had a QM score that is significantly different than the national mean. Hospices were more likely to report scores above the national mean than below the national mean (33.8% vs 16.9%, respectively, overall). The QM is able to identify those hospices that are performing well (higher than the national mean) and those that are performing less well (lower than the national mean).

Palliative Care – Pain Screening:

Seriously ill patients with palliative care sample: 99% met quality measure

Seriously ill patients without palliative care: 100% met quality measure (p=NS)

*Note additional data under 3.b1 Use in QI

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)*:

N/A

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

N/A

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)*:

N/A

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)*:

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

Future research with larger sample sizes can be used to test for differential performance by race/ethnicity, gender, cognitive status, age or other characteristics.

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)*

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

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.

Data collection for all measures is achieved through use of the Hospice Item Set (HIS). Most hospice providers and EMR vendors implemented HIS items by integrating them as structured data elements into clinical documentation systems (i.e., comprehensive assessment forms). This presented difficulty in the two pain items when HIS skip patterns were built into the clinical documentation systems. See further details below.

The existing HIS-Admission record includes two pain items: Item J0900, Pain Screening, and Item J0910, Comprehensive Pain Assessment. These items correspond to the National Quality Forum (NQF) #1634 Pain Screening quality measure and the NQF #1637 Pain Assessment quality measure. NQF #1634 calculates the percentage of patients who were screened for pain within 48 hours of admission. Patients who screen positive for pain are included in the denominator for NQF #1637, which measures the percentage of patients who screened positive for pain who received a comprehensive pain assessment within 24 hours.

Under current specifications for NQF #1634 and NQF #1637, if a patient is not in pain at the time of the first screening, that patient is not included in the denominator for NQF #1637—even if pain is an active problem for the patient. As such, if a patient is not in current pain at the time of the first pain screening, HIS skip patterns direct providers to skip Item J0910, the comprehensive pain assessment item. RTI received feedback from the provider community that the measure specifications and associated skip pattern between J0900 and J0910 do not align with clinical practice, as clinicians will often complete a comprehensive pain assessment for

patients when pain is an active problem, but the patient is not in pain at the time of the screening.

Since EMR vendors built HIS items directly into clinical documentation systems, if the associated skip pattern was also built into the clinical record, this prevented clinicians from completing/documenting data that was not required by the HIS but, nonetheless, may have been clinically appropriate. With current HIS items, the main area where this situation arose was the skip pattern between the pain screening and comprehensive assessment items.

This finding has 2 implications:
• Implications for specific pain items: CMS is planning on modifying the HIS to capture whether pain is an active problem and will tie the skip pattern between the pain screening and pain assessment item to the pain active problem component
• General considerations of how items are implemented: if data collection items will continue to be integrated directly into clinical record/EMR systems as structured data elements, this is an important consideration for measure developers to keep in mind as they are creating new items and skip patterns.

Missing data. Missing data represent a potential threat to the validity of an HIS item, which in turn may harm the validity of the QM that relies on the item. For the pain screening measure, there are three items that can include missing data, which are coded as dashes (J0900B, J0900C, and J0900D). In order to assess how these missing data impact the validity of the pain screening measure, we conducted the patient stay- and hospice-level analyses. For the patient stay-level analysis, we calculated the number and percentage of eligible patient stays for which the HIS-Admission records included a dash for each of these three items. The overall rate of missing data ranged from 0.02 percent to 0.11 percent between October 2014 and September 2015. For the hospice-level analysis we calculated each hospice's percent of eligible admissions that included missing data for items J0900B, J0900C, and J0900D. Over 90% of hospices did not have any admissions with missing data for these items.

Overall, we found that only a very small number of admission records for the eligible stays did not include data, i.e. coded as dash, on pain screening items J0900B, J0900C, and J0900D. And a vast majority of hospices did not have any missing data for these items. These results indicate that missing data for these items should not have a negative impact on the validity of the QM. There was no missing data for the elements needed to calculate this subset of pain measures for the Palliative Care samples.

Record abstraction does not require collection of unique patient identifiers and thus protects confidentiality. Timing of data collection can be concurrent with admission / initial encounter care, or can be retrospective based on medical record sampling.

Costs have not been formally estimated; medical record abstraction or electronic capture of the elements of a pain assessment will have more modest costs compared to survey data.

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)*.
N/A

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)
Public Reporting	Quality Improvement with Benchmarking (external benchmarking to multiple organizations)

	HIS/CMS https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Hospice-Quality-Reporting/Current-Measures.html Quality Improvement (Internal to the specific organization) Voluntary use by UNC Palliative Care Program http://www.med.unc.edu/pcare/resources/PEACE-Quality-Measures

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

• Name of program and sponsor: Hospice: HIS/CMS; Palliative Care: Voluntary use by PC organizations

• Purpose: Hospice: Quality reporting for hospice; Palliative Care: Internal quality improvement

• Geographic area and number and percentage of accountable entities and patients included: United States and all accountable entities and 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**

a. Due to insufficient longitudinal data, we are currently unable to discuss progress.

b. Analyses encompassed 3,922 hospice organizations and 1,218,786 patient stays across the United States.
Region:
South: 39.3%
West: 25.1%
Midwest: 23.1%
Northeast: 11.3%
Territories: 0.94%
Unknown: 0.25%
Urban/rural status:
Urban: 75.8%
Rural: 23.9%

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.**

We only have one year of data to report, which is not enough to show trends over time.

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.**

None

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 NPCRC Key Palliative Care Measures Bundle. Refer to the NPCRC cover letter and table of bundle measures for description of the selection and harmonization of the Key Palliative Care Measures Bundle.

This measure has been harmonized with ACOVE / ASSIST Measure 1628: Patients with advanced cancer screened for pain at outpatient visits. The two measures have the same focus, populations are different (although both include patients with advanced cancer), apply in different settings with different timing.

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 **Attachment:** Appendix_A.1_NQF_1634.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** University of North Carolina-Chapel Hill
Co.2 **Point of Contact:** Laura, Hanson, lhanson@med.unc.edu, 919-843-4096-
Co.3 **Measure Developer if different from Measure Steward:** University of North Carolina-Chapel Hill
Co.4 **Point of Contact:** Laura, Hanson, lhanson@med.unc.edu, 919-843-4096-

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.
The Carolinas Center for Medical Excellence PEACE Project Technical Expert Panel

The PEACE project team convened a 14-member Technical Expert Panel (TEP) of nationally recognized experts with extensive experience in the following areas: medical or nursing expertise in hospice and palliative care, methods and instrumentation, and quality improvement. Using criteria provided by the CCME study team, TEP members rated each potential quality measure on four criteria: importance, scientific soundness, feasibility and usability.

Mary Ersek, PhD, RN, Research Associate Professor, Swedish Medical Center- Pain Research Department, Seattle, WA
Betty R. Ferrell, PhD, FAAN, Research Scientist, City of Hope National Medical Center, Duarte, CA
Sean Morrison, MD, Mount Sinai Medical Center, NY, NY
Richard Payne, MD, Director, Duke Institute on Care at the End of Life, Duke Divinity School, Durham, NC
Chris Feudtner, MD, PHD, MPH, Children's Hospital of Philadelphia, Philadelphia, PA
Karen Steinhauser, PhD, Research Health Scientists, Center for Health Services Research in Primary Care, Durham VA Medical Center and Duke University, Durham, NC
Joan M. Teno, MD, Professor of Community Health and Medicine, Center for Gerontology and Health Care Research, Brown University, Providence, RI
Melanie Merriman, PhD, MBA, Touchstone Consulting, North Bay Village, FL
Sydney Dy, MD, MSc, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
David Casarett, MA, MD, Assistant Professor, Division of Geriatrics, Institute on Aging and Center for Bioethics, University of Pennsylvania School of Medicine and NHPCO Board of Directors
Judi Lund-Person, Vice President, Division of Quality, National Hospice and Palliative Care Organization, Washington, DC
Jean Kutner, MD, MSPH, Associate Professor, University of Colorado Health Sciences Center, Denver, CO
Lin Simon, Analyst, National Hospice and Palliative Care Organization, Washington, DC
Karen Pace, NAHC

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

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

Ad.8 **Additional Information/Comments:** This measure is part of the NPCRC Key Palliative Measures Bundle. Refer to the NPCRC cover letter and table of bundle measures for description of the selection and harmonization of the Key Palliative Measures Bundle.

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 #: 1637
Measure Title: Hospice and Palliative Care -- Pain Assessment
Measure Steward: University of North Carolina-Chapel Hill
Brief Description of Measure: This quality measure is defined as:
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.
Developer Rationale: Pain is prevalent and undertreated for many populations of seriously ill patients, including those patients nearing the end of life. Poor screening, assessment, and under treatment of pain is more common for patients with serious illness who are also of minority race ethnicity. Use of the Pain Screening and Pain Assessment quality measures will increase reporting and efforts to improve awareness of the presence of pain (screening) and assessment of severity, etiology and effect on function (assessment) which are the essential first steps required for quality pain management and treatment.

Numerator Statement: Patients who received a comprehensive clinical assessment to determine the severity, etiology and impact of their pain within 24 hours of screening positive for pain.
Denominator Statement: Patients enrolled in hospice OR receiving specialty palliative care in an acute hospital setting who report pain when pain screening is done on the admission evaluation / initial encounter.
Denominator Exclusions: Patients with length of stay < 1 day in palliative care. Patients who screen negative for pain are excluded from the denominator.

Measure Type: Process
Data Source: Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record
Level of Analysis: Clinician : Group/Practice, 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 or **Summary of prior review in 2012**

- The developer provides a rationale for the relationship of this process of care (pain assessment) and patient outcomes (improvements in pain and reductions in the effect of pain on function and quality of life).
- The developer cited selected individual studies, systematic reviews, and clinical practice guidelines to support the effectiveness of medical treatment for pain, the effectiveness of expert pain assessment and specialty care teams to improve pain, and the importance of screening, assessing, and treating pain in seriously and terminally ill patient populations.
    - Of these, the 2005 American Pain Society (APS) guidelines seem to best match the focus on the measure (recommendation states that all patients should be routinely screened for pain, and when it is present, pain intensity should be recorded). The 2005 APS recommendations result from literature reviews, expert experience, and consensus; the evidence was not graded and the evidence for screening was not summarized.
    - The 2008 American College of Physician Clinical Practice Guideline (Qaseem et al., 2008) includes a recommendation regarding assessment ("In patients with serious illness at the end of life, clinicians should regularly assess patients for pain, dyspnea, and depression" Grade: strong recommendation, moderate quality of evidence.) While graded, the evidence summarized was in fact relevant to treatment modalities, not screening or assessment, and was for the most part limited to cancer patients.
- During its 2012 evaluation of the measure, the Palliative/End-of-Life (EOL) Committee highlighted the uncertainty around what degree the components of screening and assessment are associated with better outcomes. Members also noted consistent follow-up assessments may have more therapeutic value than an initial assessment alone, but conceded this may be difficult to capture through measurement.
- The Committee also recommended that this measure and the pain screening measure (#1634) be paired.

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:

- Although the developer attested that there have been no changes in the evidence since it was endorsed in 2012, two new guidelines are cited.
    - 2011 British Columbia Medical Services Commission guideline relates to palliative care for patients with incurable cancer or advanced disease; it is based on a systematic review of evidence but the evidence is not graded. This guideline calls for the assessment of pain using the OPQRSTUV mnemonic (onset, provoking, quality, region, severity, treatment, understanding, values).
    - 2013 ICSI guideline Palliative Care for Adults. The pertinent recommendation from this graded guideline states that "The physical aspects of the patient's serious illness [including pain] should be an integral component of the palliative care plan (Low Quality Evidence, Strong Recommendation)". The guideline goes on to say "Each patient should be frequently evaluated for these issues."

Exception to evidence:

- Because the evidence for this measure is mostly tangential to the measure focus or based on expert opinion only, it is insufficient to meet NQF's criterion for evidence. However, an exception to the evidence criterion is allowed if the Committee agrees that empirical evidence is not needed to hold providers accountable for the measure and that other outcome measures or evidence-based measures are not available or feasible at this time.

Guidance from the Evidence Algorithm

Process measure but mostly tangential evidence (Box 3) → Evidence not graded (Box 7) → An outcome measure for pain exists, but is limited to hospice only (Box 10) → Systematic assessment of expert opinion (Box 11) → If Committee agrees it is OK/beneficial to hold providers accountable for performance in the absence of empirical evidence of benefits to patients → rate as INSUFFICIENT WITH EXCEPTION

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?*
- *Are there, or could there be, performance measures of a related health outcome, OR evidence-based intermediate clinical outcomes, intervention/treatment?*
- *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*
- *Does the SC agree that it is acceptable (or beneficial) to hold providers accountable without empirical evidence?*

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

Rationale: No evidence links assessment for pain to improved patient outcomes.

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.

- Facility-level (Hospice)
  - Data from the Hospice Quality Reporting Program (HQRP) for FY15 were provided. The data were collected from 3,922 hospice organizations and approximately 1.2 million patient stays.
  - The developer reported that 9.0% of hospices had perfect scores and 81.3% of hospices scores below 90%.

Mean	65.7%
Range	0% - 100%
10th percentile	26.9%
25th percentile	50.0%
Median	70.0%
75th percentile	86.4%
90th percentile	97.8%

- Clinician Group/Practice Level of Analysis (Palliative Care):
  - Data specific to palliative care (for clinician group/practice in the hospital setting) are not yet available, although the developer expects these data will become available for NQF review next year.

Disparities

- Facility-Level (Hospice):

Subgroup	**Hospice-level**	
	Score	**Wilcoxon-Mann-Whitney test p-value**
Race		
Non-white ≥ national median of 11.9%	65.8%	p=0.32
Non-white < national median of 11.9%	65.5%	
Sex		
Female ≥ national median of 55.2%	66.3%	p=0.11
Female < national median of 55.2%	65.1%	
Medicaid status (proxy for SES)		

Medicaid patients >= national median of 21.5%	64.6%	p=0.12
Medicaid patients < national median of 21.5%	66.4%	
Geographic location		
Rural	68.2%	p=0.05
Urban	64.8%	

- Clinician Group/Practice Level of Analysis Palliative Care:
  - Disparities data for palliative care (for clinician group/practice in the hospital setting) are not available, although the developer expects these data will become available for NQF review next year.

Questions for the Committee:

- *Is there a gap in hospice care (for pain assessment) that warrants a national performance measure?*
- *Is the Committee aware of evidence demonstrating a gap in hospital-based palliative care (for pain assessment) that warrants a national performance measure?*
- *Palliative-specific disparities information is not provided. Are you aware of evidence that disparities exist in this area of healthcare?*

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.

* The developer attests that evidence is mostly tangential and based on expert opinion. However it is paired with initial pain screening which was strongly supported. Also since this measure originated, there have been no substantive changes to the evidence. Outcomes in terms of improvement in pain are lacking.
Expert opinion strongly support this measure

1b.

* There is variation in performance from 27% to 100% for hospice. No data is available for palliative care, but will be in 2017

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 sources specified in the submission form include EHRs and other electronic clinical data.
Specifications:

- This measure is specified for the clinician group/practice level of analysis in the hospital setting and for the facility level of analysis in the hospice setting. A higher score indicates better quality.
- For hospice settings, the numerator (patients who receive a comprehensive pain assessment) and denominator (patients enrolled in hospice and screen positive for pain on admission) are identified from the Hospice Item Set, a standardized, patient-level dataset used in the Medicare Hospice Quality Reporting Program.
- For palliative care settings, data for the numerator (patients who receive a comprehensive pain assessment) and denominator (patients receiving palliative care & screen positive for moderate or severe pain (i.e., >4 on a 1-point numerical scale or self-report moderate to severe pain) are collected using a structured medical record abstraction tool. For this setting, patients with a length of stay <1 day are excluded from the denominator.
- The comprehensive clinical assessment includes at least 5 of the following 7 characteristics of the pain: location,

severity, character, duration, frequency, what relieves or worsens the pain, and the effect on function or quality of life.

- Data for the denominator are collected at the hospice admission evaluation or initial clinical encounter for palliative care; data for the numerator are collected within 24 hours of screening positive for pain.
- A calculation algorithm is provided.
- The developer indicates that sampling is permissible for the hospital setting. While some basic instructions are given, there is no guidance about the number of patients needed for the sample.

Changes to specifications since previous evaluation:

- After analysis of FY15 hospice data, the developer has changed the specifications so as to no longer exclude hospice stays of less than 7 days.

Questions for the Committee :

- *Are all the data elements clearly defined?*
- *Is the logic or calculation algorithm clear?*
- *Is it reasonable to exclude palliative care patients with < 1 day length of stay?*
- *Do all hospitals use EHRs or are some still using paper records (note: the clinician group/practice measure for the hospital setting specifies use of EHRs only)*
- *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:

- Previous reliability testing included inter-rater reliability testing of hospital data for 20 patients. Reliability testing using hospice data was not previously conducted.

Describe any updates to testing

- Score-level testing included split-half reliability, signal-to-noise, and stability analyses at the facility level.

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:

- Clinician Group/Practice level (Palliative Care)
  - Reliability of the data elements was assessed using data obtained from the medical charts for 20 seriously ill patients without specialty palliative care admitted to an acute care hospital. Inter-rater reliability between two nurse abstractors was calculated. This is an appropriate method for demonstrating data element reliability.

- Facility-Level (Hospice):
  - Reliability of the measure score was assessed using FY15 data from the Hospice Quality Reporting Program (n=3,992 hospice organizations and n=1,218,786 patient stays). Two types of reliability testing were conducted: the first via a signal-to-noise analysis and the second using a split-sample (or "test-retest") methodology. Both are appropriate methods for testing reliability. Stability analysis examined changes in provider's performance over time. 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:

- Clinician Group/Practice level (Palliative Care):
  - Developers report a single kappa value of 0.94.
  - 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 0.94 means that the raters agreed 94% 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.
  - NQF guidance indicates that data element testing should be conducted for all critical data elements, although at minimum, results about the numerator, denominator, and exclusions should be provided. For this measure, only a single kappa value is reported. Although not explicitly stated, we assume that the 20 charts that were abstracted included only seriously ill patients who received specialty palliative care in an acute hospital setting and whose length of stay was at least one day, and thus this kappa value applies to the numerator.

- Facility-Level (Hospice):
  - Split-half analysis ICC=0.91.
    - The ICC reflects the percentage of variance in score results that is due to "true" or real variance between the hospices. A value of 0.7 is often regarded as a minimum acceptable reliability value.
  - Developers report a signal-to-noise ratio of 0.98.
    - A signal-to-noise analysis quantifies the amount of variation in a performance measure that is due to true differences between hospices (i.e., signal) as opposed to random measurement error (i.e., noise). Results will vary based on the amount of variation between the providers and the number of patients treated by each provider. This method results in a reliability statistic that ranges from 0 to 1 for each hospital. A value of 0 indicates that all variation is due to measurement error and a value of 1 indicates that all variation is due to real differences in hospices performance. A value of 0.7 often is regarded as a minimum acceptable reliability value.

Guidance from the Reliability Algorithm:
Facility-level (hospice):
Precise specifications (Box 1) → Empirical reliability testing with measure as specified (Box 2) → Score-level testing (Box 4) → Appropriate method (Box 5) → High certainty that measure results are reliable (Box 6a) → High

Clinician-level (Palliative):
Precise specifications (Box 1) → Empirical reliability testing with measure as specified (Box 2) → Score-level testing not conducted (Box 4) → Data elements tested (Box 8) → Appropriate method (Box 9) → High level of agreement between raters (Box 10a) → Moderate

Questions for the Committee:

- *Are there any concerns about accurately and consistently identifying the denominator and exclusions in the hospital setting?*
- *Is the test sample adequate to generalize for widespread implementation for palliative care? If not, is current testing sufficient until more data are available (assuming the developer is planning to do additional testing)?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

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**

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:

- In the prior review, the measure was tested for the palliative care (hospital) setting using face and construct validity.

Describe any updates to validity testing:

- Additional validity testing of the measure score at the facility level of analysis (for the hospice setting) was conducted using FY15 data from the Hospice Quality Reporting System (HQRP).

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

Validity testing method:

- Facility-level (Hospice):
  - Using FY15 data from the Hospice Quality Reporting System (n=3,992 hospice organizations; n=1,218,786 patient stays), the developer conducted non-parametric Spearman rank correlation analysis between this measure and 5 other hospice quality measures. They hypothesized that agencies should perform similarly on assessment processes at hospice admission and expected the resulting correlations to be high. This is an appropriate method of score-level validation.
- Clinician Group/Practice level (Palliative Care)
  - Face validity was assessed by a group of nursing and physician stakeholders who were asked to comment on the validity, accuracy, and actionability of the measure. According to NQF guidance, the face validity of the measure score as a quality indicator may be adequate if it explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. It does appear that this testing conforms to NQF's requirements for face validity.
  - Construct validity was tested by comparing measure results for seriously ill patients seen in specialty interdisciplinary palliative care consultations (n=102) in one hospital to those who did not receive these services (n=460). The developers did not explain what they expected to find with this analysis and how it would demonstrate that the measure results reflect quality of care.

Validity testing results:

- Facility-level (Hospice)
  - Correlation results were positive and statistically significant, confirming the developer's hypothesis. However, the magnitude of the correlation was lower than they had expected. The developer states reasons for this may be due to clustering of scores, skewed distributions, and low score variability.

- Clinician Group/Practice level (Palliative Care)
  - Face validity results from the stakeholder group indicated broad endorsement of the face validity of the measure.
  - Construct validity: The developers found statistically significant differences in measure results between patients in one hospital who received specialty palliative care services compared to those who did not

(67% vs 42%; p=0.002).

Questions for the Committee:

- *Are the test samples adequate to generalize for widespread implementation?*
- *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 examined the effect of removing the previous length of stay <7 days exclusion criterion for hospice patients. Using FY15 data from the HQRP, the developers found that the removal of the criterion increased the average size of the denominator per hospice organization and had little effect on the distribution of the measure scores. The developers state these findings support the removal of the length of stay criterion as an exclusion criterion from the measure.
- The developer did not include information on the frequency of exclusions for palliative care patients.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the 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)*:

- Meaningful differences among hospices were examined using FY15 data from the Hospice Quality Reporting Program. The developers examined 95% confidence intervals to determine the proportion of hospices with results significantly different from the national hospice-level mean Results indicated 55.5% of hospices had a quality measure score that was significantly difference from the national mean. Hospices were more likely to report scores above the national mean than below the national mean (29.8% vs 25.7%, respectively).
- Developers compared measure score results for patients in one acute care hospital who received palliative care consults to those who did not receive palliative care consults. However, this analysis does not speak to whether the measure results reflect meaningful differences between clinician groups at different acute care hospitals.

Question for the Committee:

- *Does this measure identify meaningful differences about quality between hospice facilities? Between clinician groups/practices in the hospital setting?*

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

2b7. Missing Data

- For hospices, the rate of missingness is low (0.02% - 0.11% at the patient-level). More than 90% of hospices had no missing information for three key data elements from the Hospice Item set that make up this measure.
- The developer does not provide any information on missing data specific to palliative care.

Guidance from the Validity Algorithm
Facility-level (Hospice)
Specifications consistent with evidence (Box 1) → potential threats to validity assessed (Box2) →empirical validity testing (Box 3) → validity testing of the measure score (Box 6) →method appropriate (Box7) → moderate certainty (Box 8b) → Moderate

Clinician group/practice level (Palliative Care)
Specifications consistent with evidence (Box 1) → potential threats to validity only somewhat assessed (Box2) → empirical validity testing (Box 3) → validity testing of the measure score (Box 6) →method not well described (Box7) → face validity systematically assessed (Box 4) →results indicate substantial agreement as to validity (Box 5) →Moderate (assuming no concerns around exclusions, missing data, or meaningful differences).

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

Committee pre-evaluation comments

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

2a, 2b.

Reliability – Specifications

* For Hospice the measure is reliable because it is included in the Hospice Item Set implemented by Medicare, with over 1.2 million patient days. For inpatient palliative care, it is a reliable measure, but as pointed out in the summary, patients in the hospital with pain may have many goals of care, and may undergo a variety of treatments not a of which are focused on symptom management (my editorial)

Validity – Specifications

* It is limited by the criteria of LOS greater than 7 days which excludes many hospice patients. It does not impact its validity or reliability by excluding those patients.

Reliability – Testing

* There is a positive correlation between this measure and overall quality measures for hospice. Interpretation for Palliative care is more problematic, though reliable because of the regular screening of pain in the inpatient setting. Inter-rater reliability of abstraction of records is high

Validity Testing

* Yes, validity was tested with adequate scope in Hospice, as mentioned above with 1.2 million Hospice days.

Threats to Validity

* There is limited data for Palliative care patients, and it should be emphasized that these are only inpatients. Also missing are those hospice patients with a LOS less than 7 days.

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 are routinely collected and are incorporated in hospice providers' electronic clinical documentation system (i.e., electronic health record (EHR)). Data collection for all measures is achieved through use of the Hospice Item Set (HIS). It is unclear if the HIS is automatically populated by the EHR.
- The developer describes one challenge to implementation related to skip patterns built into the EHR. For patients who have an active pain problem, but who are not in pain when they are admitted and screened, the EHR does not allow the clinician to document the completion of a pain assessment. This conflicts with standard clinical practice. The developer states CMS is planning to modify the Hospice Item Set to address this problem, but specific details on these modifications are not provided.
- The developer does not provide any information on potential or actual implementation challenges for palliative care.
- The developer states formal cost estimates of data collection are not available. According to the Guidance Manual for Completion of the Hospice Item Set, completing an entire response for any one admission takes 20 minutes (though seven performance measures, of which this measure is only one, are collected as part of this process).
- In the 2012 endorsement evaluation, the Palliative and End-of-Life Care Steering Committee noted that a substantial data collection effort may be needed to abstract data if electronic data are not available.

Questions for the Committee:

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

- *Is the data collection strategy ready to be put into operational use in the hospital setting?*

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

Committee pre-evaluation comments

Criteria 3: Feasibility

* Because of the HIS data set, this is feasible, but as mentioned in 1734, requiring nurse abstraction is very expensive. Using an EHR from which valid data on pain (in discrete fields) is feasible in the inpatient space (based on reviewer's experience.) similar EHR extraction is possible for Hospice when agencies adopt a robust EHR. Nurse abstraction is reliable, with good inter-rater reliability but necessarily sample size is small.

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

Publicly reported? ☐ **Yes** ☒ **No**

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

OR

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

Accountability program details

- This measure is included in the Hospice Quality Reporting Program (HQRP). Medicare-certified hospices are required to submit an Hospice Item Set (HIS)-Admission record and an HIS-Discharge record for each patient admission on or after July 1st, 2015. In FY2015, 3,992 hospice organizations provided measure data for 1,218,786 patient stays. The distribution of reporting hospice agencies by geographic area is provided.

Improvement results

- Longitudinal data for this measure 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 2016, the Measures Application Partnership Post-Acute Care and Long-Term Care workgroup encouraged the continued development of a composite measure of Hospice Item Set measures, which includes NQF #1637.
- In February 2013, the Measure Application Partnership supported the measure for inclusion in Physician Quality Reporting System (PQRS) to address a core concept (i.e., pain assessment) not addressed in the program's measure set. Public comments from American Academy of Hospice and Palliative Medicine (AAHMP), Center to

Advance Palliative Care (CAPC), and National Coalition for Hospice and Palliative Care (NCHPC) supported the inclusion of this measure in the PQRS. CAPC and NCHPC suggested that the measure should be expanded beyond hospice patients.

***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**

Committee pre-evaluation comments

Criteria 4: Usability and Use

* As yet, data about improvement is not available, but will be in the future, time not specified. It may be premature to stop this measure before that question is answered, whether the measurement itself improves performance in pain management for these patients. Since this measure is based on documentation and improvement "gets" a better score, it may incent providers to show improvement. I think that is a small concern, and the larger picture of improving pain management outweighs these concerns

Criterion 5: Related and Competing Measures

Related or competing measures

- 0383: Percentage of visits for patients, regardless of age, with a diagnosis of cancer currently receiving chemotherapy or radiation therapy who report having pain with a documented plan of care to address pain *[clinician-level process measure in ambulatory 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]*
- 1628: Adult patients with advanced cancer who are screened for pain with a standardized quantitative tool at each outpatient visit *[facility-level and health plan-level process measure in ambulatory setting]*
- 1634: Percentage of hospice or palliative care patients who were screened for pain during the hospice admission evaluation / palliative care initial encounter *[clinician-level & facility-level process measure in hospice and hospital setting]*

Harmonization

- Due to differences in care settings, the Committee likely will not be asked to select a best-in-class measure. However, 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: Recommend making the measure specific to palliative or hospice care, not both in the same measure. Palliative care is not restricted to inpatient treatment.

Consider reworking measures 1634 & 1637 to be a singular, stronger measure related to screening for pain.

Consider reviewing measure 209 along with measures 1634 & 1637 to strengthen measures for pain assessment and intervention.

Consider incorporating recommended intervals for screening as the current measure indicates one assessment but one screening is not sufficient in this setting. Perhaps "at each patient encounter" is more appropriate?

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.

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 1637	**NQF Project:** Palliative Care and End-of-Life Care

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)***:**

Pain is under-recognized by clinicians and undertreated, resulting in excess suffering for patients with serious illness. Pain screening and assessment are necessary in order to improve the patient-centered outcome of pain, and its effects on global outcomes of function and quality of life. Pain, like other symptoms, can only be understood through patient self-report and patient observation. Screening and assessment for pain are essential steps in pain management. Without initial screening to identify patients in pain, and clinical assessment to determine the severity, etiology, and effect on function of this symptom, effective treatment cannot be administered. Additional guidelines from the American College of Physicians and the American Pain Society recommend systematic pain screening and assessment. Additional evidence includes numerous systematic reviews and individual studies.

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

Clinical Practice Guideline

Selected individual studies (rather than entire body of evidence)

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)***:**

Pain is a highly prevalent, distressing and functionally limiting symptom common to many serious illness conditions, and its relief is an important priority for patients and families.

a)Strong evidence supports the effectiveness of medical treatment for pain in cancer and in other serious illnesses to improve pain outcomes (randomized controlled trials, systematic reviews)

b)Moderate quality evidence supports the effectiveness of expert pain assessment and structural innovations such as specialty palliative care teams to improve pain outcomes. Published after the most recent systematic reviews, 3 new randomized trials and 1 controlled observational study report interventions enhancing the structure and process of palliative care delivery. Two report improved pain and quality of life outcomes (Casarett, Temel), one reports improved quality of life but no change in symptom scores (Bakitas), and one no change in symptom scores (Gade) (randomized and non-

randomized trials, systematic reviews, individual studies)

c)Three comprehensive practice guidelines support the importance of screening, assessing, and treating pain for seriously ill and terminally ill patient populations with a wide range of diagnoses.

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)*: 1. Systematic reviews of varied strategies in palliative care (Lorenz, 2004, 2008): reported results for pain management of 9 systematic reviews, 24 individual studies of interventions

2. Systematic review of RCTs of specialty palliative care services(Zimmerman, 2008): systematic review of 22 randomized trials of specialty palliative care effects on various outcomes, including symptom distress.

3. 2 additional guidelines from American College of Physicians, American Pain Society recommend systematic pain screening and assessment

4. 3 additional RCTs and 1 controlled observational study of palliative care interventions (Gade, Casarett, Bakitas, Temel)

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):* a) Multiple high quality randomized trials of pain treatment, and high quality systematic reviews address interventions to improve pain management and outcomes.

b) Studies of pain treatment include expert screening and assessment in the protocols, but do not define the effect of these processes alone on pain outcomes. Studies of complex structural and process interventions to improve specialty palliative care show varied effects on pain outcomes; however, these interventions are complex and heterogeneous in design, and provide less direct evidence for targeted interventions to improve pain as a primary outcome.

c) Included studies range in size, but many of the highest quality randomized trials have adequate power for hypothesis testing.

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* Evidence from studies explicitly targeting pain outcomes with more discrete interventions is very consistent, and provides strong evidence for improved outcomes. Evidence from studies of complex palliative care interventions shows less consistent effect on pain outcomes, and pain is used as a secondary rather than a primary outcome measure.

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

Net benefit for pain screening and assessment is strong, and clearly outweighs potential harms. Benefit is high due to the high numbers of patients affected and the marked suffering experienced as a result of current under-reporting and undertreatment, a problem more marked among patients of minority race or ethnicity. Potential harms of quality measures in pain management include over-treatment with medication toxicities and inattention to other symptoms. These harms are reduced in the context of hospice and specialty palliative care, delivered by professional teams with appropriate expertise and training.

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? Yes

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: (Lorenz, 2008) Strong evidence for treatment of cancer pain. Weak evidence supports complex interventions such as multidisciplinary teams.

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

1c.12 If other, identify and describe the grading scale with definitions:

1c.13 Grade Assigned to the Body of Evidence: See 1c.10.

1c.14 Summary of Controversy/Contradictory Evidence: N/A

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

OTHER GUIDELINES:

1. Qaseem A, Snow V, Shekelle P et al. Evidence-based interventions to improve the palliative care of pain, dyspnea and depression at the end of life: a clinical practice guideline from the American College of Physicians. Ann Intern Med 2008; 148:141-146.

2. Gordon DB, Dahl JL, Miaskowski C et al. American Pain Society recommendations for improving the quality of acute and cancer pain management. Arch Intern Med 2005; 165:1574-1580.

3. Medical Services Commission. Palliative care for the patient with incurable cancer or advanced disease. Part 2: pain and symptom management. Victoria (BC): British Columbia Medical Services Commission; 2011 Sep 30. 44 p.

SYSTEMATIC REVIEWS:

3. Lorenz KA, Lynn J, Dy SM et al. Evidence for improving palliative care at the end of life: a systematic review. Ann Intern Med 2008; 148:147-159.

4. Lorenz KA, Lynn J et al. End-of-life care and outcomes. AHRQ Publication No. 05-E004-2, December 2004.

5. Zimmerman C, Riechelmann R, Krzyzanowska M et al. Effectiveness of specialized palliative care: a systematic review. JAMA 2008; 299:1698-1709.

ADDITIONAL INDIVIDUAL STUDIES

6. Bakitas M, Lyons KD, Hegel MT et al. Effects of a palliative care intervention on clinical outcomes in patients with advanced cancer: the Project ENABLE II randomized controlled trial. JAMA 2009; 302:741-749.

7. Casarett D, Pickard A, Bailey FA et al. Do palliative consultations improve patient outcomes? J Am Geriatr Soc 2008; 56:593-599.

8. Temel JS, Greer JA, Muzikansky A et al. Early palliative care for patients with metastatic non-small-cell lung cancer. N Engl J Med 2010; 363:733-742.

9. Gade G, Venohr I, Conner D et al. Impact of an inpatient palliative care team: a randomized controlled trial. J Palliat Med 2008; 11:180-190.

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

Health Care Guideline: Palliative Care for Adults

Guideline 4. Physical Aspects of Care: The physical aspects of the patient's serious illness should be an integral component of the palliative care plan. (p.21)

Guideline 2.1 Pain, other symptoms, and side effects are managed based upon the best available evidence, with attention to disease-specific pain and symptoms, which is skillfully and systematically applied.

•Regular, ongoing assessment of pain, nonpain symptoms (including but not limited to shortness of breath, nausea, fatigue and weakness, anorexia, insomnia, anxiety, depression, confusion, and constipation), treatment side effects, and functional capacities are documented through a systematic process. Validated instruments, where available, should be utilized. Symptom assessment in children and cognitively impaired patients should be performed by appropriately trained professionals with appropriate tools.

1c.17 Clinical Practice Guideline Citation:

Palliative Care for Adults: McCusker M, Ceronsky L, Crone C, Epstein H, Greene B, Halvorson J, Kephart K, Mallen E, Nosan B, Rohr M, Rosenberg E, Ruff R, Schlecht K, Setterlund L. Institute for Clinical Systems Improvement. Palliative Care for Adults. Updated November 2013.p.21

1c.18 National Guideline Clearinghouse or other URL: Palliative Care for Adults: http://www.guideline.gov/content.aspx?id=47629

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? Yes

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

Palliative Care for Adults: Institute for Clinical Systems Improvement (ICSI)

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

Palliative Care for Adults: GRADE Metholodogy

1c.22 If other, identify and describe the grading scale with definitions:

1c.23 Grade Assigned to the Recommendation:

Palliative Care for Adults: Low Quality Evidence, Strong Recommendation

1c.24 **Rationale for Using this Guideline Over Others:** We chose to use the Palliative Care for Adults guideline from the Institute for Clinical Systems Improvement (ICSI) because it is evidenced based, uses GRADE methodology, and is included in the National Guideline Clearinghouse.

In addition, The National Consensus Project for Quality Palliative Care was the first United States national guidelines development project for palliative care quality, inclusive of hospice care. This set of guidelines, along with 38 preferred practices, has been rigorously reviewed and endorsed by the National Quality Forum. Although specific investigative groups and specialty organizations have published other guidelines in pain management or hospice or palliative care practice for specific settings, practices or populations, none have been as comprehensive or comprehensively debated, peer reviewed, or NQF endorsed.

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

1c.25 Quantity: High 1c.26 Quality: Moderate1c.27 Consistency: High

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
1637_Evidence_MSF5.0_Data.doc,1637_Evidence_3.17.16.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)*
Pain is prevalent and undertreated for many populations of seriously ill patients, including those patients nearing the end of life. Poor screening, assessment, and undertreatment of pain is more common for patients with serious illness who are also of minority race ethnicity. Use of the Pain Screening and Pain Assessment quality measures will increase reporting and efforts to improve awareness of the presence of pain (screening) and assessment of severity, etiology and effect on function (assessment) which are the essential first steps required for quality pain management and treatment.

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.*
Hospice: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1637 - Hospice and Palliative Care Pain Assessment.

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 serving approximately 1,218,786 patients.

The mean score for this QM was 65.7% with a range from 0% to 100%, the median was 70.0%, the interquartile range was 36.4, and the standard deviation was 26.2. For this QM, 9.0% of hospices had perfect scores and 81.3% of hospices scored below 90%.

Scores by decile:
10th percentile 26.9%
25th percentile 50.0%
Median 70.0%
75th percentile 86.4%
90th percentile 97.8%

Palliative Care:

This submission to the Palliative and End-of-Life Care project updates hospice setting data for NQF #s 1634, 1637, 1638, 1639, 1641, 1647. We are currently in the process of updating palliative care data by collecting and analyzing data in multiple non-hospice settings but final analyses are not available for this submission cycle.(1) Data comes from two sources -- a multi-site study of quality of care in palliative care (R18HS022763 Implementing Best Practice in Palliative Care, PI Johnson) and (CMS Health Care Innovation Award: Increasing patient and system value with community based palliative care, PI Bull / Four Seasons Compassion for Life). We anticipate these data will become available for NQF review next year. This data will allow further updates to the evidence base for non-hospice palliative care beyond what is currently submitted.

(1) Kamal AH, Bull J, Ritchie CS, Kutner JS, Hanson LC, Friedman F, Taylor DH Jr; AAHPM Research Committee Writing Group. Adherence to Measuring What Matters measures using point-of-care data collection across settings. J Pain Symptom Manage 2016; 51:497-503.

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.

Pain is prevalent, underdiagnosed and undertreated in cancer and other life-limiting or serious illnesses. The prevalence of pain ranges from 40-80% in seriously ill patient populations. As detailed in a systematic review from AHRQ and the American Pain Society Quality of Care guidelines, pain screening and assessment are the essential steps required to ensure that pain is detected by clinicians and appropriate treatment implemented.(1,2) Failure to screen, assess, and treat pain results in functional limitations, physiologic stress, and psychological harms such as social withdrawal and depression.

The current quality of pain screening, assessment and treatment is poor, as documented in systematic pain prevalence and treatment studies from hospital, outpatient, cancer and nursing home settings. (3,4,5,6) In a systematic review of quality of pain care for diverse patient populations, Gordon reported high average pain severity (6.17-8.37 on 10 point scale) and moderate rates of pain severity screening or other assessment (47%-96%). These findings did not vary by underlying diagnosis. (7)

1. Wells N, Pasero C, McCaffery M. Improving the Quality of Care through Pain Assessment and Management. In: Hughes RG, editor. Patient Safety and Quality: An Evidence-Based Handbook for Nurses. Rockville (MD): Agency for Healthcare Research and Quality (US); 2008 Apr. Chapter 17.
2. Gordon DB, Dahl JL, Miaskowski C et al. American Pain Society recommendations for improving the quality of acute and cancer pain management. Arch Intern Med 2005; 165:1574-1580.
3. Reynolds K, Henderson M, Schulman A, Hanson LC. Needs of the dying in nursing homes. J Pall Med 2002; 5:895-901.
4. Deandria S, Montanri M, Moja L et al. Prevalence of undertreatment of cancer pain: a review of published literature. Ann Oncol 2008; 19:1985-91.
5. Mularski R, White-Chu F, Overbay D et al. Measuring pain as the 5th vital sign does not improve quality of pain management. J Gen Intern Med 2006; 6:607-612.
6. Erdek MA, Pronovost PA. Improving assessment and treatment of pain in the critically ill. Int J Qual Health Care 2004; 16:59-64.
7. Gordon DB, Pelliano TA, Miaskowski C et al. A 10-year review of quality improvement monitoring in pain management: recommendations for standardized outcome measures. Pain Manage Nurs 2002; 4:116-130.

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.*

Hospice:

Racial disparity analysis. We conducted racial and ethnic disparity analyses at both the patient-stay and hospice levels. At the patient-stay level, we compared the percentage of patients who received pain assessment among different racial and ethnic groups. Patients were grouped by their racial and ethnic identification as follows: white non-Hispanic, black non-Hispanic, other non-Hispanic, or Hispanic. Other non-Hispanic includes patients who identify as American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, or who identify as more than one race. A Chi-square test was performed to determine if there were any statistically significant differences in pain assessment between groups. The lowest rate of pain assessment was found for patients with racial and ethnic group missing (50.8%), and the highest rate was among patients identifying as other non-Hispanic (63.9%). Differences in the rate of pain screening by racial identification were found to be statistically significant ($p<0.001$).
Analyses at the hospice level examined the differences in this measure across two groups: hospices with proportions of non-white patients that are greater than or equal to the national median proportion (11.9%), and hospices with fewer non-white patients than the national median. For this analysis, white non-Hispanic patients were included in the white group, and all other racial and ethnic identifications, as described above, were grouped as non-white. We ran a Wilcoxon-Mann-Whitney test for statistical dependence between group and QM score. The results showed that the QM score was not significantly different between the two groups of hospices (65.8% compared to 65.5%, $p=0.32$).

Gender disparity analysis. We performed both the patient stay- and hospice-level analyses on gender disparity. At the patient-stay level, we compared the percentage of patients who received pain assessment between female and male patients. A Chi-square test was performed to determine if there were any statistically significant differences in pain assessment between groups. We found a slightly lower rate of pain assessment for female patients (59.9%) than for male patients (62.1%). Differences in the rate of pain assessment by gender were statistically significant ($p<0.001$). At the hospice level, we examined the differences in this measure across hospices with proportions of female patients that are greater than or equal to the national median proportion (55.2%). We ran a Wilcoxon-Mann-Whitney test for statistical dependence between group and QM score. The results showed that the QM score was not statistically significantly different between the two groups of hospices split by median proportion of female patients (66.3% compared to 65.1%, $p=0.11$).

Socioeconomic disparity analysis. We performed socioeconomic disparity analyses at both the patient-stay and hospice levels using the same methods described in previous disparity analyses. Medicaid status was used as a proxy measure of low socioeconomic status. A statistically significant ($p<0.001$) lower rate of pain assessment was found for non-Medicaid patients (59.9%) than for Medicaid patients (60.3%). The highest rate of pain assessment was seen for patients with Medicaid status missing (63.8%). At the hospice level, the results showed that the QM score was not significantly different between the two groups of hospices split by median proportion of Medicaid patients, 21.5% (64.6% compared to 66.4%, $p=0.12$). The statistically significant results at the patient-stay level may indicate that quality of hospice care, measured by pain assessment, for non-Medicaid patients is lower than for Medicaid patients. A potential caveat of this analysis is that the data used for this analysis may not be a reliable indicator of a patient's Medicaid eligibility status. This data is missing for nearly a quarter of patients. In addition, some of the Medicaid numbers submitted do not appear to be legitimate Medicaid numbers. However, the data used in this analysis is the only national patient-level assessment data that is currently available in hospices. We will update this analysis as more-accurate data sources are available and accessible.

Rural-urban disparity analysis. We compared the QM score between rural and urban hospices using a Wilcoxon-Mann-Whitney test. We found a lower rate of pain assessment in urban hospices (64.8%) than for rural hospices (68.2%). Differences in the rate of pain assessment by location were statistically significant ($p=0.05$). Although statistically significant results were found, actual differences in assessment rates do not seem to be clinically substantial.

Palliative Care: Disparities data not available.

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.
Extensive evidence documents disparities in cancer pain treatment and control.(1,2) Nursing home residents with advanced cancer receive less effective pain treatment if they are African American.(3,4) The Eastern Cooperative Oncology Group Minority Outpatient Pain Study enrolled 1308 patients with advanced cancer. After clinic visits, physicians underestimated pain severity for 64% of Hispanic and 74% of African American patients.(5) Among patients with pain, 65% of Hispanic and African American patients received inadequate treatment relative to practice guidelines, as did 50% of white patients.(6,7)

1.Pletcher MJ, Kertesz SG, Kohn MA, Gonzales R. Trends in opioids prescribing by race for patients seeking care in US emergency departments. JAMA 2008; 299:70-78.

2.Green CR, Montague L, Hart-Johnson TA. Consistent and breakthrough pain in diverse advanced cancer patients: a longitudinal examination. J Pain Sympt Manage 2009; 37:831-847.
3.Bernabei R, Gambassi G, Lapane K, Landi F, Gatsonis C, Dunlop R, Lipsitz L, Steel K, Mor V. Management of pain in elderly patients with cancer. SAGE Study Group. JAMA 1998; 279:1877-82.
4.Engle VF, Fox-Hill E, Graney MJ. The experience of living-dying in a nursing home: self-reports of black and white older adults. JAGS 1998; 46:1091-96.
5.Anderson KO, Mendoza TR, Valero V, Richman SP, Russell C, Hurley J, DeLeon C, Washington P, Palos G, Payne R, Cleeland CS. Minority cancer patients and their providers: pain management attitudes and practices. Cancer 2000; 88: 1929-38.
6.Cleeland CS, Gonin R, Baez L et al. Pain and treatment of pain in minority patients with cancer. The ECOG Minority Outpatient Pain Study. Ann Intern Med 1997; 127:813-16.
7.Cleeland CS, Gonin R, Hatfield AD et al. Pain and its treatment in outpatients with metastatic cancer. N Engl J Med 1994; 330:592-96.

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**

Affects large numbers, Patient/societal consequences of poor quality, Severity of illness

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.**

The Hospice and Palliative Care - Pain Assessment measure addresses pain for patients with high severity of illness and risk of death, including seriously and incurably ill patients enrolled in hospice or hospital-based palliative care. Research on care of patients with serious incurable illness and those nearing the end of life shows they experience high rates of pain (40-70% prevalence) and other physical, emotional, and spiritual causes of distress. (1,2) The National Priorities Partnership has identified palliative and end-of-life care as one of its national priorities. A goal of this priority is to ensure that all patients with life-limiting illness have access to effective treatment for symptoms such as pain and shortness of breath.(3) The affected populations are large; in 2014, 1.66 million people with life-limiting illness received hospice care.(4) In 2013, 67% of US hospitals with 50 or more beds had some form of palliative care service, up from 58.5% in 2008, and national trends show steady expansion of these services.(5) Patients and family caregivers rate pain management as a high priority when living with serious and life-limiting illnesses. (6) The consequences of inadequate screening, assessment and treatment for pain include physical suffering, functional limitation, and development of apathy and depression. (7)

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

1. The Writing Group for the SUPPORT Investigators. A controlled trial to improve care for seriously ill hospitalized patients. The study to understand prognosis and preferences for outcomes and risks of treatments (SUPPORT). JAMA. 1995;274:1591-1598.

2. Gade G, Venohr I, Conner D, et al. Impact of an inpatient palliative care team: a randomized control trial. J Palliat Med. 2008;11(2):180–190.

3. http://www.nationalprioritiespartnership.org/PriorityDetails.aspx?id=608

4. NHPCO Facts and figures: hospice care in America 2015 edition
http://www.nhpco.org/hospice-statistics-research-press-room/facts-hospice-and-palliative-care

5. Dumanovsky T, Augustin R, Rogers M, Lettang K, Morrison RS. The growth of palliative care in U.S. hospitals: a status report. J Pall Med. 2016; 19(1): 8-15

6. Singer PA, Martin DK, Kelner M. Quality end-of-life care: patients´ perspective. JAMA 1999; 281: 163-168.

7. Gordon DB, Dahl JL, Miaskowski C et al. American Pain Society recommendations for improving the quality of acute and cancer

pain management. Arch Intern Med 2005; 165:1574-1580.

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):*

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.)*
PEACE Hospice and Palliative Care Quality Measures: http://www.med.unc.edu/pcare/resources/PEACE-Quality-Measures

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.
We would like to propose the removal of the less than 7 day length of stay (LOS) denominator exclusion for hospice patients.

Background: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1637 - Hospice and Palliative Care Pain Assessment.
Six of the seven QMs exclude patient stays that are less than 7 days from the measure denominator. When the length of stay (LOS) is too short, hospices may not have enough time to complete all the clinically recommended care processes. Thus, at the time the measures were developed, technical experts recommended that short patient stays be excluded from those measure denominators for assessing quality of care in hospices. However, no national data regarding the implications of the LOS exclusion was available to the Technical Expert Panel (TEP) at that time. Under contract to the Centers for Medicare and Medicaid Services (CMS), the Research Triangle Institute (RTI) performed descriptive analyses on 5 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.

Rationale for inclusion of all hospice patients regardless of LOS: At the patient level, approximately 40% of patients were excluded

based on the LOS < 7 days exclusion, thus omitting this essential quality of care standard for many seriously ill patients. Further, excluding short-stay patients omits measures of quality of care for a subset of the sickest patients entering hospice care, and those who arguably may be in greatest need of rapid symptom assessment and treatment. In addition, the original rationale for this exclusion -- allowing time for hospice providers to complete this care process -- does not appear to be necessary. Our analyses show that a large portion of pain assessments were performed on day 1 of admission to hospice, demonstrating a normative standard of care includes prompt attention to symptom distress and treatment.

At the hospice level, applying different LOS exclusion criteria had an impact on denominator size. Smaller denominator size at the hospice level may generate less stable and reliable QM scores. Under the LOS < 7 days exclusion, the median number of qualifying stays in the denominator was 35 stays per hospice vs. a median number of 53 stays per hospice with no LOS exclusion. Applying the LOS exclusion resulted in some hospices with no qualifying stays, which excluded 26 hospices from QM score calculations altogether. Applying or removing the LOS exclusion generally had little impact on the distribution of hospices' QM scores. Under the 7-day LOS exclusion, the mean score was 70.7%, the median score was 77.0%, and the score for hospices in the 10th percentile distribution was 32.5%. With no LOS exclusions, the mean score was 65.7%, the median score was 70.0%, and the score for hospices in the 10th percentile distribution was 26.9%. The impact of the different LOS criteria on the distribution of QM scores was consistent across quarters. In summary, these new analyses demonstrate that removing the LOS exclusion in QM calculations, thus including all hospice patients in the denominator, is feasible and appropriate for quality of care at the patient level, and enhances completeness and statistical stability of QM reporting at the hospice level. Hospice organizations have meaningful opportunity to improve in their timely assessment of pain. Since this quality measure is paired, successful screening (NQF #1634) is linked to subsequent measurement of pain assessment.

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 who received a comprehensive clinical assessment to determine the severity, etiology and impact of their pain within 24 hours of screening positive for pain.

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.)*
24 hours

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.
Patients with a comprehensive clinical assessment including at least 5 of the following 7 characteristics of the pain: location, severity, character, duration, frequency, what relieves or worsens the pain, and the effect on function or quality of life.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Patients enrolled in hospice OR receiving specialty palliative care in an acute hospital setting who report pain when pain screening is done on the admission evaluation / initial encounter.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, 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)*
The Pain Assessment quality measure is intended for patients with serious illness who are enrolled in hospice care OR receive specialty palliative care in an acute hospital setting. Conditions may include, but are not limited to: cancer, heart disease, pulmonary disease, dementia and other progressive neurodegenerative diseases, stroke, HIV/AIDS, and advanced renal or hepatic failure.

For patients enrolled in hospice, a positive screen is indicated by any pain noted in screening (any response other than none on verbal scale, any number >0 on numerical scale or any observation or self-report of pain), due to the primacy of pain control and comfort care goals in hospice care.

For patients receiving specialty palliative care, a positive screen is indicated by moderate or severe pain noted in screening (response of moderate or severe on verbal scale, >4 on a 10-point numerical scale, or any observation or self-report of moderate to severe pain). Only management of moderate or severe pain is targeted for palliative care patients, who have more diverse care goals. Individual clinicians and patients may still decide to assess mild pain, but this subset of patients is not included in the quality measure denominator.

[NOTE: This quality measure should be paired with the Pain Screening quality measure (NQF #1634) to ensure that all patients are screened and therefore given the opportunity to report pain and enter the denominator population for Pain Assessment.]

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Patients with length of stay < 1 day in palliative care. Patients who screen negative for pain are excluded from the denominator.

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)*
Calculation of length of stay; discharge date is identical to date of initial encounter.

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)*
N/A

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*)

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.)*
Clinical assessment of Pain:
a.Step 1- Identify all patients with serious, life-limiting illness who are enrolled in hospice OR received specialty palliative care in an acute hospital setting
b.Step 2- Exclude palliative care patients if length of stay is < 1 day.
c.Step 3- Identify patients who were screened for pain during the admission evaluation (hospice) OR initial encounter (palliative care)

d.Step 4- Identify patients who screened positive for pain [any pain if hospice; moderate or severe pain if palliative care].
e.Step 5- Exclude patients who screened negative for pain
f.Step 6- Identify patients who received a clinical assessment for pain within 24 hours of screening positive for pain

Quality Measure= Numerator: Patients who received a clinical assessment for pain in Step 6 / Denominator: Patients in Step 4

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.
Hospice: The hospice analysis was not based on a sample. It was conducted on the entire hospice population that had admission and discharge records in the specified period of analysis.

Palliative Care: consecutive sample of equal numbers of admissions + decedents beginning with a randomly selected date.

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.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Hospice: For the pain assessment measure, there are two items on the HIS that can include missing data – J0910B and J0910C. If missing, these items are coded as dashes and analyzed as though the process did not happen.

In order to assess how these missing data impact the validity of the pain assessment measure, we conducted the patient stay- and hospice-level analyses. For the patient stay-level analysis, we calculated the number and percentage of eligible patient stays for which the HIS-Admission records included a dash for each of these items. The overall rate of missing data ranged from 0.04 percent to 0.09 percent between October 2014 and September 2015. For the hospice-level analysis we calculated each hospice's percent of eligible admissions that included missing data for items J0910B and J0910C. Almost 95% of hospices did not have any admissions with missing data for these items.

Overall, we found that only a very small number of admission records for the eligible stays did not include data, i.e. coded as dash, on pain assessment items J0910B and J0910C and a vast majority of hospices did not have any missing data for these items. These results indicate that missing data for these items should not have a negative impact on the validity of the QM.

Palliative care: N/A

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record

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.
Hospice: Hospice analysis uses the Hospice Item Set (HIS) as the data source to calculate the quality measure.

Palliative Care: Structured medical record abstraction tool with separate collection of numerator and denominator values.

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 in attached appendix at A.1

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

Clinician : Group/Practice, Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospice, 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
1637_MeasureTesting_MSF5.0_Data.doc,1637_MeasureTestingAttachment_2.26.16.docx

NQF #: 1637 **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)***:**

Hospice: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1637 - Hospice and Palliative Care Pain Assessment.

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.

Palliative Care: Two research nurse abstractors independently recorded quality measures data on a random subset of 20 seriously ill patients. Abstractors used the pre-defined operational definitions and a structured chart abstraction tool to record numerator and denominator data separately. Patients were a subsample of 460 seriously ill patients without specialty palliative care admitted to an acute care hospital for at least 1 day to four inpatient services. Records eligible for sampling included all seriously ill adult patients admitted to medical and surgical intensive care, medically complex patients aged 65 and older admitted to an Acute Care of the Elderly Unit, and medical oncology patients with Stage IV carcinoma.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)***:**

Hospice:

Split-half reliability. Split-half reliability assesses the internal consistency of a QM in different samples by randomly dividing the patient stays within each hospice into two halves, and calculating correlation between the QM scores on the basis of the two randomly divided halves. In this analysis, we conducted a split-half reliability analysis on all facilities with 20 or more patient stays counted in the measure denominator, and used the Interclass Correlation (ICC) coefficients to measure the internal reliability. In general, the ICC coefficient varies between 0 and 1, where an ICC of 0 indicates no reliability and an ICC of 1 indicates perfect reliability.

Signal-to-noise analysis. If a measure is reliable, then true differences in provider performance should explain a substantial proportion of the variance in QM scores. We conducted an analysis of variance (ANOVA) to determine what proportion of total variance in the measure is attributable to differences among providers. A higher proportion indicates better reliability

Stability analysis. Stability analysis describes the extent to which providers' performance assessed by a QM changes across time. We analyzed the change in facility scores between four consecutive quarters (Q4 2014 and Q1 2015, Q1 and Q2 2015, and Q2 and Q3 2015). The changes in facility scores are reported in standard deviations.

Palliative Care:Inter-rater reliability between the two abstractors was assessed using kappa statistics. Two research nurse abstractors independently recorded quality measures data on a random subset of 20 seriously ill patients. Abstractors used the pre-defined operational definitions and a structured chart abstraction tool to record numerator and denominator data separately

2a2.3 **Testing Results** *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**

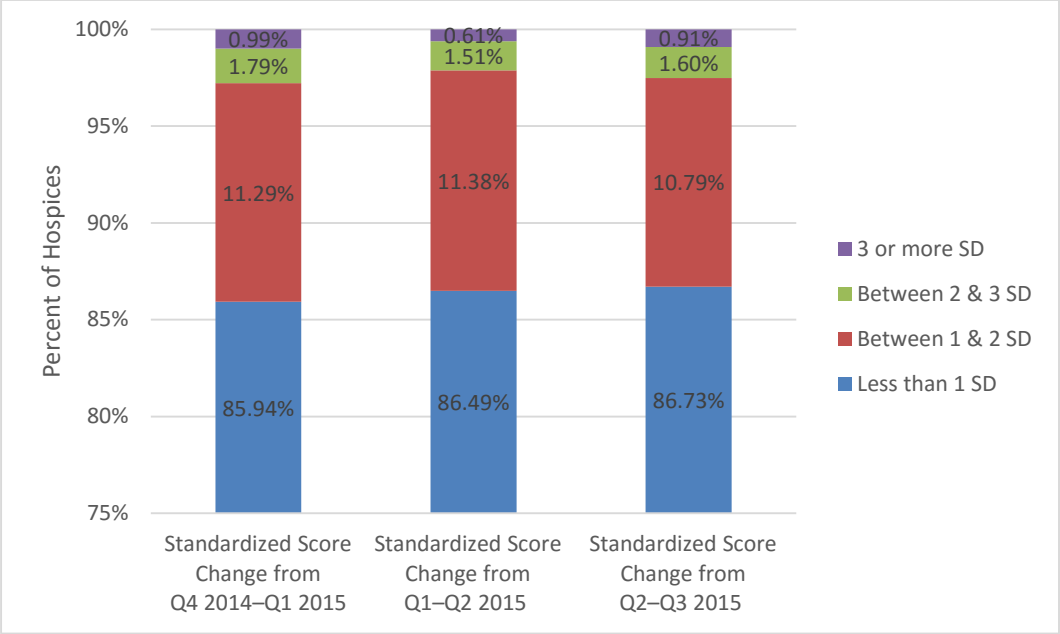
Hospice:

Split-half reliability: In general, the ICC coefficient varies between 0 and 1, where an ICC of 0 indicates no reliability and an ICC of 1 indicates perfect reliability. The ICC coefficient for this measure is 0.91, indicating high internal reliability.

Signal-to-noise analysis: The analysis results found the signal-to-noise ratio to be 0.98, indicating that about 98% of the variance in this measure is because of differences among facilities. This proportion indicates strong reliability for this measure.

Stability analysis: The results of this analysis indicated that facility scores were very stable. Less than 90% of facilities had a change in QM score of less than one standard deviation, indicating high stability of the QM. The number of facilities with a change in QM of less than one standard deviation increased across quarters from 85.9% to 86.7%, suggesting improved reliability across time. Roughly one-tenth of facilities had a change in QM between one and two standard deviations. These results indicate a measure that is generally quite stable. Figure 1 illustrates the change in facility scores between the four consecutive quarters.

Figure 1
Standardized Score Change in QM Score from Q4 2014 to Q3 2015



Palliative Care: Kappa scores range from 0 to 1 with higher scores indicating better agreement. The nurse abstractors achieved excellent inter-rate reliability for this measure with Kappa=0.94. Landis and Koch describe kappa values that range from 0.81 – 0.99 as almost perfect and Fleiss describes kappas over 0.75 as excellent.

Landis, J.R.; Koch, G.G. (1977). "The measurement of observer agreement for categorical data". Biometrics 33 (1): 159–174

Fleiss, J.L. (1981). Statistical methods for rates and proportions (2nd ed.). New York: John Wiley

2b. VALIDITY. Validity, Testing, including all Threats to Validity: 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:**

The measure focus is Pain Assessment, designed to pair with Pain Screening to ensure quality care processes for pain. The target populations are hospice patients, and seriously ill hospitalized patients with diverse underlying diagnoses who are at high risk for palliative care clinical needs, including pain.

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)***:**

Hospice: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1637 - Hospice and Palliative Care Pain Assessment.

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.

Palliative Care: The total patient sample size was 562. Chart abstractions were completed for 102 consecutive seriously ill patients with specialty palliative care consultation, and a random sample of 460 seriously ill patients without specialty palliative care admitted to an acute care hospital for at least 1 day to four inpatient services with high proportions of seriously ill patients. Records eligible for sampling included all patients admitted to medical and surgical intensive care, medically complex patients aged 65 and older admitted to a Geriatric Evaluation Unit, and medical oncology patients with Stage IV carcinoma. Because palliative care domains become even more relevant closer to death, patients dying in hospital were over-sampled to ensure a final ratio of 1 decedent to 1 live discharge. Consistent with oversampling of decedent records, 55% of these patients died in hospital.

The age of the patients ranged from 16 to 99 years, with the mean age 61. Patients were predominantly Caucasian (65%), with smaller subgroups who were African American (24%) and Hispanic / Latino (4%) The most common life-limiting diagnoses were infections (37%), cancer (34%), pulmonary (29%), and neurologic diseases (21%).

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

Hospice:

Correlations. Providers should perform similarly on QMs that reflect the quality of similar care processes. Thus, a common strategy used to evaluate validity of a QM is to examine the relative performance of the provider on QMs that are conceptually related, either due to some core competency manifested across a stay or for processes that occur in close proximity We conducted nonparametric Spearman rank correlation analysis among all seven HQRP QMs, which address care processes around hospice admission that are clinically recommended or required in the hospice Conditions of Participation. We used nonparametric methods since the data are heavily skewed (most providers perform well, while a small proportion have lower scores). Since all of the seven current QMs measure desirable care assessment processes upon hospice admission, higher positive Spearman correlation coefficients among these QMs will demonstrate higher validity of the QM set. Statistically significant correlations, indicated by p-values of the Spearman correlation coefficients less than 0.05, also support the validity of the QM set.

Missing data. Missing data represent a potential threat to the validity of an HIS item, which in turn may harm the validity of the QM that relies on the item. For the pain assessment measure, there are two items on the HIS that can include missing data, which are coded as dashes – J0910B and J0910C. In order to assess how these missing data impact the validity of the pain assessment measure, we conducted the patient stay- and hospice-level analyses.

Palliative Care sample: Face validity of PEACE quality measures for hospital-based specialty palliative care was addressed using stakeholder review and feedback. Investigators prepared data reports in a summary format with detailed operational definitions, and led a 1-hour discussion with nursing and physician leaders from each service group – MICU, SICU, Acute Care for the Elderly (Geriatrics), Oncology, and Palliative Care. The discussion included feedback of quality measure data, response to questions and critiques, and eliciting stakeholder feedback about the validity and actionability of this data for the care of their patients. Stakeholders were specifically asked to comment on the accuracy of the data as a reflection of current care practices, and their highest priority area for future quality improvement.

Construct validity was tested by comparing the PEACE quality measures for patients seen by specialty interdisciplinary palliative care consultants to those not receiving specialty palliative care services.

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)***:**

Hospice:

Correlations: Table 1 presents the Spearman correlation coefficients. The p-values for all the Spearman correlation coefficients are significant (p <0.001). The significant positive correlations between every pair of QMs indicate that high-performing hospices in one area also provided high-quality care in other areas at hospice admission. Overall, the correlations between the QMs are low to moderate. The clustering of QM scores, skewed distributions, and low variability across all seven QMs may affect the level of correlations between QMs.

Table 1
Correlation of Hospice QMs, Percentile Ranking

Quality measure	NQF #1647 (modified) Beliefs/Values Addressed	NQF #1634 Pain Screening	NQF #1637 Pain Assessment	NQF #1639 Dyspnea Screening	NQF #1638 Dyspnea Treatment	NQF #1617 Bowel Regimen
NQF #1641 Treatment Preferences	0.64***	0.37***	0.21***	0.44***	0.28***	0.25***
NQF #1647 (modified) Beliefs/Values Addressed		0.42***	0.17***	0.43***	0.31***	0.25***
NQF #1634 Pain Screening			0.17***	0.41***	0.28***	0.15***
NQF #1637 Pain Assessment				0.17***	0.08***	0.20***
NQF #1639 Dyspnea Screening					0.30***	0.25***
NQF #1638 Dyspnea Treatment						0.34***

NOTE: The correlation is on the basis of each hospice's percentile ranking on the QMs.
*** indicates significant correlation at p < 0.001

Overall, NQF #1637, Pain Assessment, has significant positive correlations with the other QMs, indicating hospices providing higher-quality care in this area also performed better in other areas at hospice admission. However, the correlation coefficients are low, indicating relatively weaker relationships. The conceptual tie between performance on pain assessment and other measures may be low.

Missing data. For the patient stay-level analysis, we calculated the number and percentage of eligible patient stays for which the HIS-Admission records included a dash for each of these items. The overall rate of missing data ranged from 0.04 percent to 0.09 percent between October 2014 and September 2015. For the hospice-level analysis we calculated each hospice's percent of eligible admissions that included missing data for items J0910B and J0910C. Almost 95% of hospices did not have any admissions with missing data for these items.

Overall, we found that only a very small number of admission records for the eligible stays did not include data, i.e. coded as dash, on pain assessment items J0910B and J0910C and a vast majority of hospices did not have any missing data for these items. These results indicate that missing data for these items should not have a negative impact on the validity of the QM.

Palliative Care sample:

Face Validity: Stakeholder discussions provided broad endorsement of face validity, with some considerations for specific patient populations. Intensive care and geriatrics clinicians endorsed the primary importance of pain screening and

assessment, but expressed doubts about the validity of numerical pain severity ratings when used for nonverbal or confused patients. Medical oncologists endorsed the face validity of these quality measures, but favored quality measures endorsed by oncology professional organizations.

Construct Validity: Screening for pain with a numerical pain scale was nearly universal for all seriously ill patients, regardless of use of specialty palliative care, and half had moderate or severe pain. Patients with moderate or severe pain were more likely to have a clinical assessment of pain if seen by specialty palliative care (67% vs 42%, p=0.002).

CITATION: Schenck AP, Rokoske FS, Durham DD et al. The PEACE Project: identification of quality measures for hospice and palliative care. J Palliat Med 2010; 13:1451-1459.

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)***:**

Hospice: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1637 - Hospice and Palliative Care Pain Assessment.

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.

Palliative Care: N/A

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

Hospice: Under contract to the Centers for Medicare and Medicaid Services (CMS), the Research Triangle Institute (RTI) performed descriptive analyses on 5 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.

Palliative Care: N/A

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

Hospice: At the patient level, approximately 40% of patients were excluded based on the LOS < 7 days exclusion, thus omitting this essential quality of care standard for many seriously ill patients. Further, excluding short-stay patients omits measures of quality of care for a subset of the sickest patients entering hospice care, and those who arguably may be in greatest need of rapid symptom assessment and treatment. In addition, the original rationale for this exclusion -- allowing time for hospice providers to complete this care process -- does not appear to be necessary. Our analyses show that a large portion of pain assessments were performed on day 1 of admission to hospice, demonstrating a normative standard of

care includes prompt attention to symptom distress and treatment.

At the hospice level, applying different LOS exclusion criteria had an impact on denominator size. Smaller denominator size at the hospice level may generate less stable and reliable QM scores. Under the LOS < 7 days exclusion, the median number of qualifying stays in the denominator was 35 stays per hospice vs. a median number of 53 stays per hospice with no LOS exclusion. Applying the LOS exclusion resulted in some hospices with no qualifying stays, which excluded 26 hospices from QM score calculations altogether.

Applying or removing the LOS exclusion generally had little impact on the distribution of hospices' QM scores. Under the 7-day LOS exclusion, the mean score was 70.7%, the median score was 77.0%, and the score for hospices in the 10th percentile distribution was 32.5%. With no LOS exclusions, the mean score was 65.7%, the median score was 70.0%, and the score for hospices in the 10th percentile distribution was 26.9%. The impact of the different LOS criteria on the distribution of QM scores was consistent across quarters.

In summary, these new analyses demonstrate that removing the LOS exclusion in QM calculations, thus including all hospice patients in the denominator, is feasible and appropriate for quality of care at the patient level, and enhances completeness and statistical stability of QM reporting at the hospice level. Hospice organizations have meaningful opportunity to improve in their timely screening of pain. Since this quality measure is paired, successful screening (NQF #1634) is linked to subsequent measurement of pain assessment.

Palliative Care: N/A

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)***:**

N/A

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

N/A

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)***:**

N/A

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

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)***:**

Hospice: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1637 - Hospice and Palliative Care Pain Assessment.

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.

Palliative Care: The total patient sample size was 562. Chart abstractions were completed for 102 consecutive seriously ill patients with specialty palliative care consultation, and a random sample of 460 seriously ill patients without specialty palliative care admitted to an acute care hospital for at least 1 day to four inpatient services with high proportions of seriously ill patients. Records eligible for sampling included all patients admitted to medical and surgical intensive care, medically complex patients aged 65 and older admitted to a Geriatric Evaluation Unit, and medical oncology patients with Stage IV carcinoma. Because palliative care domains become even more relevant closer to death, patients dying in hospital were over-sampled to ensure a final ratio of 1 decedent to 1 live discharge. Consistent with oversampling of decedent records, 55% of these patients died in hospital.

The age of the patients ranged from 16 to 99 years, with the mean age 61. Patients were predominantly Caucasian (65%), with smaller subgroups who were African American (24%) and Hispanic / Latino (4%) The most common life-limiting diagnoses were infections (37%), cancer (34%), pulmonary (29%), and neurologic diseases (21%).

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

Hospice: Confidence interval analysis. We examined proportions of hospices with the QM scores that are significantly different from the national hospice-level mean. If a high proportion of hospices have a measure score significantly different from the mean, the QM can identify facilities with different levels of performance. For this analysis, statistical significance was determined using 95 percent confidence intervals: a hospice's QM score was significantly different from the national mean if the national mean was not included within the hospice's 95 percent confidence interval. High-performing facilities should have scores that are significantly below average, and low-performing facilities should be significantly above average.

Palliative Care: Construct validity was tested by comparing the PEACE quality measures for patients seen by specialty interdisciplinary palliative care consultants to those not receiving specialty palliative care services. Percentage of patients with and without specialty palliative care for whom the quality measure was met was compared for difference using the chi-square statistic.

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)*:

Hospice:

The mean score for this QM was 65.7% with a range from 0% to 100%, the median was 70.0%, the interquartile range was 36.4, and the standard deviation was 26.2. For this QM, 9.0% of hospices had perfect scores and 81.3% of hospices scored below 90%.

Scores by decile:

10th percentile 26.9%

25th percentile 50.0%

Median 70.0%

75th percentile 86.4%

90th percentile 97.8%

Across all hospices, 55.5% had a QM score that is significantly different than the national mean. Hospices were more likely to report scores above the national mean than below the national mean (29.8% vs 25.7%, respectively, overall). The QM is able to identify those hospices that are performing well (higher than the national mean) and those that are performing less well (lower than the national mean).

Palliative care sample: Patients with moderate or severe pain were more likely to have a documented clinical assessment of pain if seen by specialty palliative care (67% vs 42%, p=0.002).

* For additional data, see 3b.1. Use in QI

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)*:

N/A

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

N/A

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)*:

N/A

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)*:

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

Future research with larger sample sizes can be used to test for differential performance by race / ethnicity and by gender.

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)*

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

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.

Data collection for all measures is achieved through use of the Hospice Item Set (HIS). Most hospice providers and EMR vendors implemented HIS items by integrating them as structured data elements into clinical documentation systems (i.e., comprehensive assessment forms). This presented difficulty in the two pain items when HIS skip patterns were built into the clinical documentation systems. See further details below.

The existing HIS-Admission record includes two pain items: Item J0900, Pain Screening, and Item J0910, Comprehensive Pain Assessment. These items correspond to the National Quality Forum (NQF) #1634 Pain Screening quality measure and the NQF #1637 Pain Assessment quality measure. NQF #1634 calculates the percentage of patients who were screened for pain within 48 hours of admission. Patients who screen positive for pain are included in the denominator for NQF #1637, which measures the percentage of patients who screened positive for pain who received a comprehensive pain assessment within 24 hours.

Under current specifications for NQF #1634 and NQF #1637, if a patient is not in pain at the time of the first screening, that patient is not included in the denominator for NQF #1637—even if pain is an active problem for the patient. As such, if a patient is not in current pain at the time of the first pain screening, HIS skip patterns direct providers to skip Item J0910, the comprehensive pain assessment item. RTI received feedback from the provider community that the measure specifications and associated skip pattern between J0900 and J0910 do not align with clinical practice, as clinicians will often complete a comprehensive pain assessment for

patients when pain is an active problem, but the patient is not in pain at the time of the screening.

Since EMR vendors built HIS items directly into clinical documentation systems, if the associated skip pattern was also built into the clinical record, this prevented clinicians from completing/documenting data that was not required by the HIS but, nonetheless, may have been clinically appropriate. With current HIS items, the main area where this situation arose was the skip pattern between the pain screening and comprehensive assessment items.

This finding has 2 implications:
• Implications for specific pain items: CMS is planning on modifying the HIS to capture whether pain is an active problem and will tie the skip pattern between the pain screening and pain assessment item to the pain active problem component
• General considerations of how items are implemented: if data collection items will continue to be integrated directly into clinical record/EMR systems as structured data elements, this is an important consideration for measure developers to keep in mind as they are creating new items and skip patterns.

Missing data. Missing data represent a potential threat to the validity of an HIS item, which in turn may harm the validity of the QM that relies on the item. For the pain assessment measure, there are two items on the HIS that can include missing data – J0910B and J0910C. If missing, these items are coded as dashes.

In order to assess how these missing data impact the validity of the pain assessment measure, we conducted the patient stay- and hospice-level analyses. For the patient stay-level analysis, we calculated the number and percentage of eligible patient stays for which the HIS-Admission records included a dash for each of these three items. The overall rate of missing data ranged from 0.04 percent to 0.09 percent between October 2014 and September 2015. For the hospice-level analysis we calculated each hospice's percent of eligible admissions that included missing data for items J0910B and J0910C. Almost 95% of hospices did not have any admissions with missing data for these items.

Overall, we found that only a very small number of admission records for the eligible stays did not include data, i.e. coded as dash, on pain assessment items J0910B and J0910C and a vast majority of hospices did not have any missing data for these items. These results indicate that missing data for these items should not have a negative impact on the validity of the QM. There was no missing data for the elements needed to calculate this subset of pain measures for the Palliative Care sample.

Record abstraction does not require collection of unique patient identifiers and thus protects confidentiality. Timing of data collection can be concurrent with admission / initial encounter care, or can be retrospective based on medical record sampling.

Costs have not been formally estimated; medical record abstraction or electronic capture of the elements of a pain assessment will have more modest costs compared to survey data.

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)*.
N/A

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)

Public Reporting	Quality Improvement with Benchmarking (external benchmarking to multiple organizations) HIS/CMS https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Hospice-Quality-Reporting/Current-Measures.html Quality Improvement (Internal to the specific organization) Voluntary use by UNC Palliative Care Program http://www.med.unc.edu/pcare/resources/PEACE-Quality-Measures

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

• Name of program and sponsor: Hospice: HIS/CMS; Palliative Care: Voluntary use by PC organizations
• Purpose: Hospice: Quality reporting for hospice; Palliative Care: Internal quality improvement
• Geographic area and number and percentage of accountable entities and patients included: United States and all accountable entities and 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?*)
N/A

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.*)
N/A

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**

a. Due to insufficient longitudinal data, we are currently unable to discuss progress.

b. Analyses encompassed 3,922 hospice organizations and 1,218,786 patient stays across the United States.
Region:
South: 39.3%
West: 25.1%
Midwest: 23.1%
Northeast: 11.3%
Territories: 0.94%
Unknown: 0.25%
Urban/rural status:
Urban: 75.8%
Rural: 23.9%
Unknown: 0.31%

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.
We only have one year of data to report, which is not enough to show trends over time.

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.**
None

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.)
This measure was part of the NPCRC Key Palliative Care Measures Bundle. Refer to the NPCRC cover letter and table of bundle measures for description of the selection and harmonization of the Key Palliative Care Measures Bundle.

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 **Attachment:** Appendix_A.1_NQF_1637.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** University of North Carolina-Chapel Hill
Co.2 **Point of Contact:** Laura, Hanson, lhanson@med.unc.edu, 919-843-4096-
Co.3 **Measure Developer if different from Measure Steward:** University of North Carolina- Chapel Hill
Co.4 **Point of Contact:** Laura, Hanson, lhanson@med.unc.edu, 919-843-4096-

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.
The Carolinas Center for Medical Excellence PEACE Project Technical Expert Panel

The PEACE project team convened a 14-member Technical Expert Panel (TEP) of nationally recognized experts with extensive experience in the following areas: medical or nursing expertise in hospice and palliative care, methods and instrumentation, and quality improvement. Using criteria provided by the CCME study team, TEP members rated each potential quality measure on four criteria: importance, scientific soundness, feasibility and usability.

Mary Ersek, PhD, RN, Research Associate Professor, Swedish Medical Center- Pain Research Department, Seattle, WA
Betty R. Ferrell, PhD, FAAN, Research Scientist, City of Hope National Medical Center, Duarte, CA
Sean Morrison, MD, Mount Sinai Medical Center, NY, NY
Richard Payne, MD, Director, Duke Institute on Care at the End of Life, Duke Divinity School, Durham, NC
Chris Feudtner, MD, PHD, MPH, Children's Hospital of Philadelphia, Philadelphia, PA
Karen Steinhauser, PhD, Research Health Scientists, Center for Health Services Research in Primary Care, Durham VA Medical Center and Duke University, Durham, NC
Joan M. Teno, MD, Professor of Community Health and Medicine, Center for Gerontology and Health Care Research, Brown University, Providence, RI
Melanie Merriman, PhD, MBA, Touchstone Consulting, North Bay Village, FL
Sydney Dy, MD, MSc, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
David Casarett, MA, MD, Assistant Professor, Division of Geriatrics, Institute on Aging and Center for Bioethics, University of Pennsylvania School of Medicine and NHPCO Board of Directors
Judi Lund-Person, Vice President, Division of Quality, National Hospice and Palliative Care Organization, Washington, DC
Jean Kutner, MD, MSPH, Associate Professor, University of Colorado Health Sciences Center, Denver, CO
Lin Simon, Analyst, National Hospice and Palliative Care Organization, Washington, DC
Karen Pace, NAHC

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

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

Ad.8 **Additional Information/Comments:** This measure is part of the NPCRC Key Palliative Measures Bundle. Refer to the NPCRC cover letter and table of bundle measures for description of the selection and harmonization of the Key Palliative Measures Bundle.

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 #: 1638
Measure Title: Hospice and Palliative Care -- Dyspnea Treatment
Measure Steward: University of North Carolina-Chapel Hill
Brief Description of Measure: Percentage of patients who screened positive for dyspnea who received treatment within 24 hours of screening.
Developer Rationale: Dyspnea is prevalent and undertreated for many populations of seriously ill patients, including those patients nearing the end of life. Screening for dyspnea is necessary to determine its presence and severity, and forms the basis for treatment decision-making. Unlike pain, structured clinical assessment of the symptom is less well-defined, yet similar to pain, effective treatment is available to alleviate symptom distress.

Numerator Statement: Patients who screened positive for dyspnea who received treatment within 24 hours of screening.
Denominator Statement: Patients enrolled in hospice OR patients receiving hospital-based palliative care for 1 or more days.
Denominator Exclusions: Patients with length of stay < 1 day in palliative care, patients who were not screened for dyspnea, and/or patients with a negative screening.

Measure Type: Process
Data Source: Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record
Level of Analysis: Clinician : Group/Practice, 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**

Summary of prior review in 2012

- The developer provides a rationale for the importance of dyspnea screening, assessment and treatment. An explicit link to patient outcomes is not provided.
- The developer cited systematic reviews of randomized trials and small trials supporting 1) the use of opioids for breathlessness; 2) shorter follow-up times; 3)the use of beta agonists for dyspnea in COPD; 4) the use of oxygen for hypoxic patients; and 5)the benefit of coping or relaxation interventions.
- In 2012, the Steering Committee agreed dyspnea is a problem for a large number of patients and this measure would likely benefit these patients.
- This measure is paired with # 1639: Hospice and Palliative Care- Dyspnea Screening

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:

- Although the developer attested that there have been no changes in the evidence since it was endorsed in 2012, they added two new guidelines to the submission.
    - The 2011 British Columbia Medical Services Commission guideline relates to palliative care for patients with incurable cancer or advanced disease; it is based on a systematic review of evidence but the evidence is not graded. This guideline recommends pharmacological (e.g., opioids, benzodiazepines, neuroleptics, corticosteroids, supplemental O_2) and non-pharmacological treatment options for dyspnea.
    - The 2013 ICSI guideline on Palliative Care for Adults. The pertinent recommendation from this graded guideline states that "The physical aspects of the patient's serious illness should be an integral component of the palliative care plan (Low Quality Evidence, Strong Recommendation)".

Exception to evidence: N/A

Guidance from the Evidence Algorithm

Process measure based on systematic review (Box 3) → QQC presented (Box 4) → Quantity: high; Quality: moderate; Consistency: high (Box 5) → Moderate (Box 5b) → Moderate

Questions for the Committee:

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

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

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.

- Facility-Level (Hospice) :
    - Data from the Hospice Quality Reporting Program (HQRP) for FY15 are provided. The data are from 3,922 hospice organizations and approximately 1.2 million patient stays.
    - The developer reported that 31.3% of hospices had perfect scores, and 21.7% of hospices scored below

90%.

Mean	93.3%
Range	0% - 100%
10th percentile	81.6%
25th percentile	91.7%
Median	97.5%
75th percentile	100%
90th percentile	100%

- Clinician Group/Practice Level of Analysis (Palliative Care):
  - Data specific to palliative care (for clinician group/practice in the hospital setting) are not yet available, although the developer expects these data will become available for NQF review next year.

Disparities

- Facility-level (Hospice):

Subgroup	Hospice-level	
	Score	Wilcoxon-Mann-Whitney test p-value
Race		
Non-white ≥ national median of 11.9%	92.5%	p<0.001
Non-white < national median of 11.9%	93.9%	
Sex		
Female ≥ national median of 55.2%	93.6%	p=0.13
Female < national median of 55.2%	93.0%	
Medicaid status (proxy for SES)		
Medicaid patients >= national median of 21.5%	91.4%	p<0.001
Medicaid patients < national median of 21.5%	95.0%	
Geographic location		
Rural	92.2%	p=0.85
Urban	93.6%	

- Clinician Group/Practice Level (Palliative Care):
  - Disparities data for palliative care are not available, although the developer expects these data will become available for NQF review next year.

Questions for the Committee:

- *Is there a gap in hospice care (for dyspnea treatment) that warrants a national performance measure?*
- *Is the Committee aware of evidence demonstrating a gap in hospital-based palliative care (for dyspnea treatment) that warrants a national performance measure?*
- *Palliative-specific disparities information is not provided. Are you aware of evidence that disparities exist in this area of healthcare?*

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.

* Maintenance Measure -Process. The evidence relates to the importance of screening and treating dyspnea, using pharmacologic and non pharmacologic treatments.

* HOSPICE: This process measure applies directly to patients who screened positive for dyspnea on admission (paired with NQF #1639 Dyspnea Screening) that received treatment within 24 hours. Since dyspnea is a real and frequent symptom at end of life, the outcome of dyspnea treatment initiated within 24 hours of admission will correlate to better end of life care.

PALLIATIVE CARE: n/a

1b.
* Performance data was provided. "Effective treatment for dyspnea is available, but not consistently administered. Evidence-based treatments include pharmacologic interventions such as opioids and inhaled bronchodilators, and non-pharmacologic interventions including oxygen for hypoxic patients, pulmonary rehabilitation and exercise in COPD, and drainage of pleural effusion". Palliative care: disparities data not available, study cited. Disparities in dyspnea screening and treatment have not been well characterized in the hospice and palliative care population. Future research with larger sample sizes can be used to test for differential performance by race / ethnicity and by gender.
* HOSPICE: Performance data provided indicates that 31.3% of hospices had perfect scores of initiating treatment within 24 hours of patients screening positive for dyspnea. 21.7% of hospices scored below 90%.

PALLIATIVE CARE: not available at this time

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 sources specified in the submission form include EHRs and other electronic clinical data.

Specifications:

- This measure is specified for the clinician group/practice level of analysis in the hospital setting and for the facility level of analysis in the hospice setting. A higher score indicates better quality.
- For hospice settings, the numerator (patient treated for dyspnea within 24 hours of positive screening) and denominator (patients enrolled in hospice) are identified from the Hospice Item Set, a standardized, patient-level dataset used in the Medicare Hospice Quality Reporting Program.
- For palliative care (hospital) settings, data for the numerator and denominator (patients admitted to palliative care) are collected using a structured medical record abstraction tool.
- Treatment is defined as medical treatment plan, orders, or pharmacy records for inhaled medications, steroids, diuretics, or non-medication strategies such as oxygen and energy conservation; benzodiazepines or opioids may also be considered as treatment if clearly prescribed for dyspnea.
- Data are collected at the hospice admission evaluation or initial clinical encounter for palliative care.
- The following exclusion criteria apply to the denominator:
  - patients with a length of stay <1 day in palliative care
  - patients who were not screened for dyspnea
  - patients with a negative screening
- A calculation algorithm is provided.
- The developer indicates that sampling is permissible for the hospital setting. While some basic instructions are given, there is no guidance about the number of patients needed for the sample.

Changes to specifications since previous evaluation:

- After analysis of FY15 hospice data, the developer has changed the specifications so as to no longer exclude hospice stays of less than 7 days.

Questions for the Committee :

- *Are all the data elements clearly defined?*
- *Is the logic or calculation algorithm clear?*
- *Is it reasonable to exclude palliative care patients with < 1 day length of stay?*
- *Do all hospitals use EHRs or are some still using paper records (note: the clinician group/practice measure for the hospital setting specifies use of EHRs only)?*
- *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:

- Previous reliability testing included inter-rater reliability testing of hospital data for 20 patients. Reliability testing using hospice data was not previously conducted.

Describe any updates to testing

- Score-level testing for the hospice setting included split-half reliability, signal-to-noise, and stability analyses at the facility level.

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:

- Clinician Group/Practice level (Palliative Care):
  - Reliability of the data elements was assessed using data obtained from the medical charts for 20 seriously ill patients without specialty palliative care admitted to an acute care hospital. Inter-rater reliability between two nurse abstractors was calculated. This is an appropriate method for demonstrating data element reliability.
- Facility-level (Hospice):
  - Reliability of the measure score was assessed using FY15 data from the Hospice Quality Reporting Program (n=3,992 hospice organizations and n=1,218,786 patient stays). Two types of reliability testing were conducted: the first via a signal-to-noise analysis and the second using a split-sample (or "test-retest") methodology. Both are appropriate methods for testing reliability. Stability analysis examined changes in provider's performance over time. 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:

- Clinician Group/Practice level (Palliative Care)
  - Developers report a single kappa value of 0.89.
  - 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 0.89 means that the raters agreed 89.0% 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.
  - NQF guidance indicates that data element testing should be conducted for all critical data elements, although at minimum, results about the numerator, denominator, and exclusions should be provided. For this measure, only a single kappa value is reported. Although not explicitly stated, we assume that the 20 charts that were abstracted included only seriously ill patients who received specialty palliative care in an acute hospital setting and whose length of stay was at least one day, and thus this kappa

value applies to the numerator.

- Facility-level (Hospice):
  - Split-half analysis ICC=0.86.
    - The ICC reflects the percentage of variance in score results that is due to "true" or real variance between the hospices. A value of 0.7 is often regarded as a minimum acceptable reliability value.
  - Developers report a signal-to-noise ratio of 0.96.
    - A signal-to-noise analysis quantifies the amount of variation in a performance measure that is due to true differences between hospices (i.e., signal) as opposed to random measurement error (i.e., noise). Results will vary based on the amount of variation between the providers and the number of patients treated by each provider. This method results in a reliability statistic that ranges from 0 to 1 for each hospital. A value of 0 indicates that all variation is due to measurement error and a value of 1 indicates that all variation is due to real differences in hospices performance. A value of 0.7 often is regarded as a minimum acceptable reliability value.

Guidance from the Reliability Algorithm
Facility-level (hospice):
Precise specifications (Box 1) → Empirical reliability testing with measure as specified (Box 2) → Score-level testing (Box 4) → Appropriate method (Box 5) → High certainty that measure results are reliable (Box 6a) → High

Clinician-level (Palliative):
Precise specifications (Box 1) → Empirical reliability testing with measure as specified (Box 2) → Score-level testing not conducted (Box 4) → Data elements tested (Box 8) → Appropriate method (Box 9) → High level of agreement between raters → Moderate

Questions for the Committee:

- *Are there any concerns about accurately and consistently identifying the denominator and exclusions in the hospital setting?*
- *Is the test sample adequate to generalize for widespread implementation for palliative care? If not, is current testing sufficient until more data are available (assuming the developer is planning to do additional testing?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

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**

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:

- In the prior review, the measure was tested for the palliative care (hospital) setting using face and construct

validity.

Describe any updates to validity testing:

- Additional validity testing of the measure score at the facility level of analysis (for the hospice setting) was conducted using FY15 data from the Hospice Quality Reporting System (HQRP).

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

Validity testing method:

- Facility-level (Hospice):
  - Using FY15 data from the Hospice Quality Reporting System (n=3,992 hospice organizations; n=1,218,786 patient stays), the developer conducted non-parametric Spearman rank correlation analysis between this measure and 5 other hospice quality measures. They hypothesized that agencies should perform similarly on assessment processes at hospice admission and expected the resulting correlations to be high. This is an appropriate method of score-level validation.
- Clinician Group/Practice level (Palliative Care)
  - Face validity was assessed by a group of nursing and physician stakeholders who were asked to comment on the validity, accuracy, and actionability of the measure. According to NQF guidance, the face validity of the measure score as a quality indicator may be adequate if it explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. It does appear that this testing conforms to NQF's requirements for face validity.
  - Construct validity was tested by comparing measure results for seriously ill patients seen in specialty interdisciplinary palliative care consultations (n=102) in one hospital to those who did not receive these services (n=460). The developers did not explain what they expected to find with this analysis and how it would demonstrate that the measure results reflect quality of care.

Validity testing results :

- Facility-level (Hospice):
  - Correlation results were positive and statistically significant, confirming the developer's hypothesis. However, the magnitude of the correlation was lower than they had expected. The developer states reasons for this may be due to clustering of scores, skewed distributions, and low score variability.

- Clinician Group/Practice (Palliative Care):
  - Face validity results from the stakeholder group indicated broad endorsement of the face validity of the measure.
  - Construct validity results found patients with dyspnea were likely to receive some form of treatment within 24 hours, with or without the addition of specialty palliative care (96% vs 93%, p=NS).

Questions for the Committee:

- *Are the test samples adequate to generalize for widespread implementation?*
- *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 examined the effect of removing the previous length of stay <7 days exclusion criterion for hospice patients. Using FY15 data from the HQRP, the developers found that the removal of the criterion

increased the average size of the denominator per hospice organization and had little effect on the distribution of the measure scores. The developers stated these finds supported the removal of the length of stay criterion as an exclusion criterion from the measure.
- The developer did not include information on exclusion analyses for palliative care patients.

Questions for the Committee:
- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the 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)*:

- Meaningful differences among hospices were examined using FY15 data from the Hospice Quality Reporting Program. The developers examined 95% confidence intervals to determine the proportion of hospices with results significantly different from the national hospice-level mean. Results indicated 39.8% of hospices had a score that was significantly different from the national mean. Hospices were more likely to report scores above the national mean than below the national mean (27.0% vs 12.8%, respectively).
- Developers compared measure score results for patients in one acute care hospital who received palliative care consults to those who did not receive palliative care consults. However, this analysis does not speak to whether the measure results reflect meaningful differences between clinician groups at different acute care hospitals.

Question for the Committee:
- *Does this measure identify meaningful differences about quality between hospice facilities? Between clinician groups/practices in the hospital setting?*

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

2b7. Missing Data
- For hospices, the rate of missingness is low (0.03% - 0.12% at the patient-level). More than 90% of hospices had no missing information for two key data elements from the Hospice Item set that make up this measure.
- The developer does not provide any information on missing data specific to palliative care.

Guidance from the Validity Algorithm

Facility-level (Hospice)
Specifications consistent with evidence (Box 1) → potential threats to validity assessed (Box2) →empirical validity testing (Box 3) → validity testing of the measure score (Box 6) →method appropriate (Box7) →moderate certainty (Box 8b)→ Moderate

Clinician group/practice level (Palliative Care)
Specifications consistent with evidence (Box 1) → potential threats to validity only somewhat assessed (Box2) → empirical validity testing (Box 3) → validity testing of the measure score (Box 6) →method not well described (Box7) → face validity systematically assessed (Box 4) →results indicate substantial agreement as to validity (Box 5) →Moderate (assuming no concerns around exclusions, missing data, or meaningful differences).

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

Committee pre-evaluation comments

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

2a, 2b.
Reliability – Specifications
* Data elements clearly defined.
Am in agreement with developer on no longer excluding hospice stays of less than 7 days.
* HOSPICE: from a user standpoint, at point of service at admission, one of the questions asked is “WAS TREATMENT FOR SHORTNESS OF BREATH INITIATED?”. This is not an accurate response as it should say “continued or initiated” as

does the pain question. Staff have to answer "YES" that they initiated treatment when in reality they are actually continuing treatment.

Validity – Specifications
* Specifications consistent with the evidence.
* HOSPICE: results were positive and significant, thus supporting the hypothesis for the measure.

PALLIATIVE CARE: high validity results

Reliability – Testing
* Several methods of testing, with resultant moderate rating. for reliability.
* see concerns above

Validity Testing
* Test size appears adequate to generalize for widespread implementation
* 2b3: Exclusions:

HOSPICE: excluding the <7 day criteria increased the denominator but did not effect the distribution of the measure scores. The developers would like to exclude the <7 day LOS and use all data collected.

PALLIATIVE CARE: not included

Threats to Validity
* I do not feel the exclusion plan presents a threat to the validity of the measure.
* none

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 are routinely collected and are incorporated in hospice providers' electronic clinical documentation system (i.e., electronic health record (EHR)). Data collection for all measures is achieved through use of the Hospice Item Set (HIS). It is unclear if the HIS is automatically populated by the EHR.
- The developer does not provide any information on potential or actual implementation challenges for palliative care.
- The developer states formal cost estimates of data collection are not available. According to the Guidance Manual for Completion of the Hospice Item Set, completing an entire response for any one admission takes 20 minutes (though seven performance measures, of which this measure is only one, are collected as part of this process).
- In the 2012 endorsement evaluation, the Palliative and End-of-Life Care Steering Committee noted that a substantial data collection effort may be needed to abstract data if electronic data are not available.

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

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

Committee pre-evaluation comments
Criteria 3: Feasibility

* no concerns

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

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

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

Accountability program details

- This measure is included in the Hospice Quality Reporting Program (HQRP). Medicare-certified hospices are required to submit a Hospice Item Set (HIS)-Admission record and an HIS-Discharge record for each patient admission on or after July 1st, 2015. In FY2015, 3,992 hospice organizations provided measure data for 1,218,786 patient stays. The distribution of reporting hospice agencies by geographic area is provided.

Improvement results

- Longitudinal data for this measure 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.
- In the previous evaluation in 2012, the Steering Committee questioned whether this measure would result in the potential for overtreatment of dyspnea.

Feedback :

- In February 2016, the Measures Application Partnership Post-Acute Care and Long-Term Care workgroup supported the continued development of a composite measure of Hospice Item Set measures, which includes NQF #1638.
- In February 2013, the Measure Application Partnership supported the measure for inclusion in the Physician Quality Reporting System (PQRS) and the Hospice Quality Reporting Program (HQRP) to address a core concept (i.e., dyspnea treatment).Public comments from American Academy of Hospice and Palliative Medicine (AAHMP), Center to Advance Palliative Care (CAPC), and National Coalition for Hospice and Palliative Care (NCHPC) supported the inclusion of this measure. CAPC and NCHPC suggested that the measure should be expanded beyond hospice patients. The measure is not currently included in the PQRS program, but it is included in the Hospice Quality Reporting Program.
- In February 2012, the Measure Applications Partnership (MAP) supported the inclusion of this measure in the Hospice Quality Reporting program. Public comments generally concurred with the recommendation.

***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**

Committee pre-evaluation comments

Criteria 4: Usability and Use

* Currently not being publicly reported.

- This measure is included in the Hospice Quality Reporting Program (HQRP). Medicare-certified hospices are required to submit a Hospice Item Set (HIS)-Admission record and an HIS-Discharge record for each patient admission on or after July 1st, 2015. In FY2015, 3,992 hospice organizations provided measure data for 1,218,786 patient stays. The distribution of reporting hospice agencies by geographic area is provided.

Few harms of screening, or treatment of dyspnea reported.

No unintended consequences foreseen and the identification and prompt treatment of dyspnea will add to the quality of life, relief of the distress and anxiety of dyspnea which is more prevalent than pain in certain populations of seriously ill.

* no public reporting at this time. The measure has led to greater focus on results of assessment leading to more timely treatment.

Criterion 5: Related and Competing Measures

Related measures

- 0179: Improvement in dyspnea: Percentage of home health episodes of care during which the patient became less short of breath or dyspneic *[facility-level outcome measure in home health setting]*
- 1639: Percentage of hospice or palliative care patients who were screened for dyspnea during the hospice admission evaluation / palliative care initial encounter *[clinician-level & facility-level process measure in hospice and hospital setting*]

Harmonization

- Committee 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: Recommend making the measure specific to palliative or hospice care, not both in the same measure. Palliative care is not restricted to inpatient treatment.

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.

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 1638	**NQF Project:** Palliative Care and End-of-Life Care

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)***:**

Dyspnea screening and assessment are necessary to detect the presence of dyspnea (for which physical signs such as hypoxia and tachypnea do not clearly correlate), and to understand its severity and underlying etiology. Evidence-based treatment of dyspnea will vary with its severity and etiology, with treatment options differing for causes such as malignant pleural effusion, bulky tumor mass, congestive heart failure, anemia, COPD, among others. Additional guidelines from the American College of Physicians recommend dyspnea screening and assessment. Additional evidence includes numerous systematic reviews.

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)***:**

Dyspnea is a prevalent, distressing and functionally limiting symptom common to many serious illness conditions, and associated with risk of death.

a) Combining the results of several systematic reviews, moderate evidence supports pharmacologic treatments, including opioids for treatment of dyspnea in cancer and non-cancer diagnoses, and inhaled beta agonists for COPD.

b) Combining the results of several systematic reviews, moderate evidence supports non-pharmacologic treatments, including oxygen for hypoxic patients with cancer and non-cancer diagnoses, exercise interventions for COPD and CHF, thoracentesis for malignant pleural effusions, and nurse-led coping or relaxation interventions in cancer dyspnea.

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)*: 1. Systematic review of dyspnea treatment in palliative care (Lorenz, 2008); reported on 7 systematic reviews and 12 additional individual studies.

2. Systematic review of dyspnea management in cancer care, with evidence included for other diagnoses (Dy, 2008); reported on 25 studies of oxygen, 33 RCTs of beta agonists in COPD and 1 in "terminal illness," 18 studies of opioids, 9 of which were RCTs, and 1 meta-analysis of 36 RCTs of pleurodesis.

3. Systematic review for cancer dyspnea management (Ben-Aharon, 2008); reported on 18 randomized clinical trials.

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):* a) Multiple randomized trials of varied quality support the use of opioids for breathlessness; follow-up times are generally short, and evidence is stronger for COPD than for cancer patients.

b) Multiple randomized trials support the use of beta agonists for dyspnea in COPD

c) Several randomized trials support the use of oxygen, with mixed results and stratification showing beneficial effects for hypoxic but not for non-hypoxic patients.

d) Several small trials provide early evidence for the benefit of coping or relaxation interventions.

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

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

Benefits outweigh harms. Few harms of dyspnea screening or assessment are reported, few harms of treatment are reported.

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? Yes

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: Overall moderate quality (see Lorenz, 2008)

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

1c.12 If other, identify and describe the grading scale with definitions:

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

1c.14 Summary of Controversy/Contradictory Evidence: N/A

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

1. Luce JM, Luce JA. Management of dyspnea in patients with far-advanced lung disease. JAMA 2001; 285:1331-1337.

2. Ben-Aharon I, Gafter-Gvili A, Paul M et al. Interventions for alleviating cancer-related dyspnea: a systematic review. J Clin Oncol 2008; 26:2396-2404.

3. Roberts DK, Thorne SE, Pearson C. Cancer Nurs 1993; 16:310-320.

4. Currow DC, Ward AM, Abernethy AP. Advances in the pharmacologic management of breathlessness. Current Opin Supp Pall Care 2009; 3:103-106.

5. Bausewin C, Booth S, Gysels M et al. Non-pharmacologic interventions for breathlessness in advanced stages of malignant and nonmalignant diseases. Cochrane Database Syst Rev 2009. Apr 16; 2:CD005623.

6. Lorenz KA, Lynn J, Dy SM et al. Evidence for improving palliative care at the end of life: a systematic review. Ann Intern Med 2008; 148:147-159.

7. Dy SM, Lorenz KA, Naeim A et al. Evidence-based recommendations for cancer fatigue, anorexia, depression and dyspnea. J Clin Onc 2008; 26:3886-3895.

8. Lorenz KA, Lynn J et al. End-of-life care and outcomes. AHRQ Publication No. 05-E004-2, December 2004.

OTHER GUIDELINES:

9. Medical Services Commission. Palliative care for the patient with incurable cancer or advanced disease. Part 2: pain and symptom management. Victoria (BC): British Columbia Medical Services Commission; 2011 Sep 30. 44 p.

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

Health Care Guideline: Palliative Care for Adults

Guideline 4. Physical Aspects of Care: The physical aspects of the patient's serious illness should be an integral component of the palliative care plan. (p.21)

Guideline 2.1 Pain, other symptoms, and side effects are managed based upon the best available evidence, with attention to disease-specific pain and symptoms, which is skillfully and systematically applied.

•Regular, ongoing assessment of pain, non-pain symptoms (including but not limited to shortness of breath, nausea, fatigue and weakness, anorexia, insomnia, anxiety, depression, confusion, and constipation), treatment side effects, and functional capacities are documented through a systematic process. Validated instruments, where available, should be utilized. Symptom assessment in children and cognitively impaired patients should be performed by appropriately trained professionals with appropriate tools.

1c.17 Clinical Practice Guideline Citation: Palliative Care for Adults: McCusker M, Ceronsky L, Crone C, Epstein H, Greene B, Halvorson J, Kephart K, Mallen E, Nosan B, Rohr M, Rosenberg E, Ruff R, Schlecht K, Setterlund L. Institute for Clinical Systems Improvement. Palliative Care for Adults. Updated November 2013.p.21

1c.18 National Guideline Clearinghouse or other URL: Palliative Care for Adults: http://www.guideline.gov/content.aspx?id=47629

1c.19 **Grading of Strength of Guideline Recommendation.** Has the recommendation been graded? Yes

1c.20 **If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias:** Palliative Care for Adults: Institute for Clinical Systems Improvement (ICSI)

1c.21 **System Used for Grading the Strength of Guideline Recommendation:** Palliative Care for Adults: GRADE Methodology

1c.22 **If other, identify and describe the grading scale with definitions:**

1c.23 **Grade Assigned to the Recommendation:** Palliative Care for Adults: Low Quality Evidence, Strong Recommendation

1c.24 **Rationale for Using this Guideline Over Others:** We chose to use the Palliative Care for Adults guideline from the Institute for Clinical Systems Improvement (ICSI) because it is evidenced based, uses GRADE methodology, and is included in the National Guideline Clearinghouse.

In addition, the National Consensus Project for Quality Palliative Care was the first United States national guidelines development project for palliative care quality, inclusive of hospice care. This set of guidelines, along with 38 preferred practices, has been rigorously reviewed and endorsed by the National Quality Forum. Although specific investigative groups and specialty organizations have published other guidelines in pain management or hospice or palliative care practice for specific settings or populations, none have been as comprehensive or comprehensively debated, peer reviewed, or NQF endorsed.

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: Moderate1c.27 Consistency**:** High

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
1638_Evidence_MSF5.0_Data.doc,1638_Evidence_3.17.16.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)*
Dyspnea is prevalent and undertreated for many populations of seriously ill patients, including those patients nearing the end of life. Screening for dyspnea is necessary to determine its presence and severity, and forms the basis for treatment decision-making. Unlike pain, structured clinical assessment of the symptom is less well-defined, yet similar to pain, effective treatment is available to alleviate symptom distress.

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.*
Hospice: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1638 - Hospice and Palliative Care Dyspnea Treatment.
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 1,218,786 patient stays.

The mean score for this QM was 93.3% with a range from 0% to 100%, the median was 97.5%, the interquartile range was 8.3, and the standard deviation was 11.2. For this measure, 31.3% of hospices had perfect scores and 21.7% of hospices scored below 90%.

Scores by decile:
10th percentile 81.6%
25th percentile 91.7%
Median 97.5%
75th percentile 100%
90th percentile 100%

Palliative Care:

This submission to the Palliative and End-of-Life Care project updates hospice setting data for NQF #s 1634, 1637, 1638, 1639, 1641, 1647. We are currently in the process of updating palliative care data by collecting and analyzing data in multiple non-hospice settings but final analyses are not available for this submission cycle.(1) Data comes from two sources -- a multi-site study of quality of care in palliative care (R18HS022763 Implementing Best Practice in Palliative Care, PI Johnson) and (CMS Health Care Innovation Award: Increasing patient and system value with community based palliative care, PI Bull / Four Seasons Compassion for Life). We anticipate these data will become available for NQF review next year. This data will allow further updates to the evidence base for non-hospice palliative care beyond what is currently submitted.

(1) Kamal AH, Bull J, Ritchie CS, Kutner JS, Hanson LC, Friedman F, Taylor DH Jr; AAHPM Research Committee Writing Group.

Adherence to Measuring What Matters measures using point-of-care data collection across settings. J Pain Symptom Manage 2016; 51:497-503.

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.

Prevalence of dyspnea in advanced cancer ranges from 50-70%. Among COPD patients with advanced illness enrolled in the SUPPORT Study, dyspnea which was moderate to severe at least half of the time was present for at least 65% of patients throughout the 6 months preceding death.

Effective treatment for dyspnea is available, but not consistently administered. Evidence-based treatments include pharmacologic interventions such as opioids and inhaled bronchodilators, and non-pharmacologic interventions including oxygen for hypoxic patients, pulmonary rehabilitation and exercise in COPD, and drainage of pleural effusion.

1. Luce JM, Luce JA. Management of dyspnea in patients with far-advanced lung disease. JAMA 2001; 285:1331-1337.
2. Ben-Aharon I, Gafter-Gvili A, Paul M et al. Interventions for alleviating cancer-related dyspnea: a systematic review. J Clin Oncol 2008; 26:2396-2404.
3. Roberts DK, Thorne SE, Pearson C. Cancer Nurs 1993; 16:310-320.
4. Currow DC, Ward AM, Abernethy AP. Advances in the pharmacologic management of breathlessness. Current Opin Supp Pall Care 2009; 3:103-106.
5. Bausewin C, Booth S, Gysels M et al. Non-pharmacologic interventions for breathlessness in advanced stages of malignant and nonmalignant diseases. Cochrane Database Syst Rev 2009. Apr 16; 2:CD005623.
6. Lorenz KA, Lynn J, Dy SM et al. Evidence for improving palliative care at the end of life: a systematic review. Ann Intern Med 2008; 148:147-159.
7. Dy SM, Lorenz KA, Naeim A et al. Evidence-based recommendations for cancer fatigue, anorexia, depression and dyspnea. J Clin Onc 2008; 26:3886-3895.

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.*

Hospice:

Racial disparity analysis. We conducted racial and ethnic disparity analyses at both the patient-stay and hospice levels. At the patient-stay level, we compared the percentage of patients who received dyspnea treatment among different racial and ethnic groups. Patients were grouped by their racial and ethnic identification as follows: white non-Hispanic, black non-Hispanic, other non-Hispanic, or Hispanic. Other non-Hispanic includes patients who identify as American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, or who identify as more than one race. A Chi-square test was performed to determine if there were any statistically significant differences in dyspnea treatment between groups. The lowest rate of dyspnea treatment was found for Black non-Hispanic patients (95.1%), and the highest rate was among patients with racial and ethnic group missing (96.3%). Differences in the rate of dyspnea treatment by racial identification were found to be statistically significant ($p<0.001$).

Analyses at the hospice level examined the differences in this measure across two groups: hospices with proportions of non-white patients that are greater than or equal to the national median proportion (11.9%), and hospices with fewer non-white patients than the national median. For this analysis, white non-Hispanic patients were included in the white group, and all other racial and ethnic identifications, as described above, were grouped as non-white. We ran a Wilcoxon-Mann-Whitney test for statistical dependence between group and QM score. The results showed that the QM score was significantly different between the two groups of hospices (92.5% compared to 93.9%, $p<0.001$). Although statistically significant results were found at both the patient and hospice level, actual differences in screening rates do not seem to be clinically substantial.

Gender disparity analysis. We performed both the patient stay- and hospice-level analyses on gender disparity. At the patient-stay level, we compared the percentage of patients who received dyspnea treatment between female and male patients. A Chi-square test was performed to determine if there were any statistically significant differences in dyspnea treatment between groups. We found a slightly lower rate of dyspnea treatment for female patients (95.1%) than for male patients (95.4%). Differences in the rate of dyspnea treatment by gender were statistically significant ($p<0.001$). At the hospice level, we examined the differences in this measure across hospices with proportions of female patients that are greater than or equal to the national median proportion (55.2%). We ran a Wilcoxon-Mann-Whitney test for statistical dependence between group and QM score. The results showed that the QM score was not statistically significantly different between the two groups of hospices split by median proportion of female

patients (93.6% compared to 93.0%, p=0.13).

Socioeconomic disparity analysis. We performed socioeconomic disparity analyses at both the patient-stay and hospice levels using the same methods described in previous disparity analyses. Medicaid status was used as a proxy measure of low socioeconomic status. A statistically significant (p<0.001) lower rate of dyspnea treatment was found for non-Medicaid patients (94.5%) than for Medicaid patients (95.7%). The highest rate of dyspnea treatment was seen for patients with Medicaid status missing (97.0%). At the hospice level, the results showed that the QM score was significantly different between the two groups of hospices split by median proportion of Medicaid patients, 21.5% (91.4% compared to 95.0%, p<0.001). The statistically significant results at the patient-stay level may indicate that quality of hospice care, measured by dyspnea treatment, for non-Medicaid patients is lower than for Medicaid patients. The significant findings at the hospice level indicate that hospices with a smaller proportion of Medicaid patients are less likely to provide dyspnea treatment to patients at admission. A potential caveat of this analysis is that the data used for this analysis may not be a reliable indicator of a patient's Medicaid eligibility status. This data is missing for nearly a quarter of patients. In addition, some of the Medicaid numbers submitted do not appear to be legitimate Medicaid numbers. However, the data used in this analysis is the only national patient-level assessment data that is currently available in hospices. We will update this analysis as more-accurate data sources are available and accessible.

Rural-urban disparity analysis. We compared the QM score between rural and urban hospices using a Wilcoxon-Mann-Whitney test. The results showed that the QM score was not significantly different between rural and urban hospices (92.2% compared to 93.6%, p=0.85). This indicates that rural hospices and urban hospices perform similarly on this quality measure.

Palliative Care: Disparities data not available

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.**
Limited research has explored the nature of health disparities in the experience of dyspnea or in dyspnea management. One observational study of dyspnea in cancer patients provides evidence that dyspnea and other symptoms, in addition to minority race / ethnicity, independently predict worsened survival.

Tammemagi CM, Neslund-Dudas C, Simoff M, Kvale P. Lung carcinoma symptoms – an independent predictor of survival and an important mediator of African-American health disparity in survival. Cancer. 2004 Oct 1;101(7):1655-63.

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**
Affects large numbers, Patient/societal consequences of poor quality, Severity of illness
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.**
The Hospice and Palliative Care Dyspnea Treatment measure addresses dyspnea for patients with high severity of illness and risk of death, including seriously and incurably ill patients enrolled in hospice or hospital-based palliative care. Research on care of patients nearing the end of life shows they experience high rates of physical, emotional, and spiritual causes of distress.(1,2) The National Priorities Partnership has identified palliative and end-of-life care as one of its national priorities. A goal of this priority is to ensure that all patients with life-limiting illness have access to effective treatment for symptoms such as pain and shortness of breath.(3) In 2014, 1.66 million people with life-limiting illness received hospice care.(4) In 2013, 67% of US hospitals with 50 or more beds had some form of palliative care service, up from 58.5% in 2008, and national trends show a steady expansion of these services. (5)

Dyspnea is a common symptom in serious illness, more common than pain for patients with chronic obstructive lung disease, lung cancer, cystic fibrosis, and restrictive lung diseases such as pulmonary fibrosis.(6) Unlike pain, dyspnea severity is associated with the risk of death.(7) Between 50-70% of patients with advanced lung cancer experience dyspnea near the end of life. As detailed in a recent systematic review, opioids, oxygen and non-pharmacologic nursing interventions demonstrate efficacy in randomized

controlled trials of treatment for dyspnea in cancer and in other serious illness.(8,9) Unfortunately, dyspnea is often persistent and under-treated in advanced cancer and other end-stage diseases.(10)

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

1. The Writing Group for the SUPPORT Investigators. A controlled trial to improve care for seriously ill hospitalized patients. The study to understand prognosis and preferences for outcomes and risks of treatments (SUPPORT). JAMA. 1995;274:1591-1598.
2. Gade G, Venohr I, Conner D, et al. Impact of an inpatient palliative care team: a randomized control trial. J Palliat Med. 2008;11(2):180–190.
3. http://www.nationalprioritiespartnership.org/PriorityDetails.aspx?id=608
4. NHPCO Facts and figures: hospice care in America 2015 edition
http://www.nhpco.org/hospice-statistics-research-press-room/facts-hospice-and-palliative-care
5. Dumanovsky T, Augustin R, Rogers M, Lettang K, Morrison RS. The growth of palliative care in U.S. hospitals: a status report. J Pall Med. 2016; 19(1): 8-15
6. Luce JM, Luce JA. Management of dyspnea in patients with far-advanced lung disease. JAMA 2001; 285:1331-1337.
7. Olajidae O, Hanson LC, Usher BM et al. Validation of the Palliative Performance Score in the acute tertiary hospital setting. J Palliat Med 2007; 10:111-117
8. Ben-Aharon I, Gafter-Gvili A, Paul M et al. Interventions for alleviating cancer-related dyspnea: a systematic review. J Clin Oncol 2008; 26:2396-2404.
9. Lorenz KA, Lynn J, Dy SM et al. Evidence for improving palliative care at the end of life: a systematic review. Ann Intern Med 2008; 148:147-159.
10. Roberts DK, Thorne SE, Pearson C. Cancer Nurs 1993; 16:310-320.

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):*

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.)*
PEACE Hospice and Palliative Care Quality Measures: http://www.med.unc.edu/pcare/resources/PEACE-Quality-Measures

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.
We would like to propose the removal of the less than 7 day length of stay (LOS) denominator exclusion for hospice patients.

Background: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1638- Hospice and Palliative Care Dyspnea Treatment.

Six of the seven QMs exclude patient stays that are less than 7 days from the measure denominator. When the length of stay (LOS) is too short, hospices may not have enough time to complete all the clinically recommended care processes. Thus, at the time the measures were developed, technical experts recommended that short patient stays be excluded from those measure denominators for assessing quality of care in hospices. However, no national data regarding the implications of the LOS exclusion was available to the Technical Expert Panel (TEP) at that time. Under contract to the Centers for Medicare and Medicaid Services (CMS), the Research Triangle Institute (RTI) performed descriptive analyses on 5 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 1,218,786 patient stays.

Rationale for inclusion of all hospice patients regardless of LOS: At the patient level, approximately 40% of patients were excluded based on the LOS < 7 days exclusion, thus omitting this essential quality of care standard for many seriously ill patients. Further, excluding short-stay patients omits measures of quality of care for a subset of the sickest patients entering hospice care, and those who arguably may be in greatest need of rapid symptom assessment and treatment. In addition, the original rationale for this exclusion -- allowing time for hospice providers to complete this care process -- does not appear to be necessary. Our analyses show that a large portion of dyspnea treatments were performed on day 1 of admission to hospice, demonstrating a normative standard of care includes prompt attention to symptom distress.

At the hospice level, applying different LOS exclusion criteria had an impact on denominator size. Smaller denominator size at the hospice level may generate less stable and reliable QM scores. Under the LOS < 7 days exclusion, the median number of qualifying stays in the denominator was 41 stays per hospice vs. a median number of 63 stays per hospice with no LOS exclusion. Applying the LOS exclusion resulted in some hospices with no qualifying stays, which excluded 33 hospices from QM score calculations altogether.

Applying or removing the LOS exclusion generally had little impact on the distribution of hospices' QM scores. Under the 7-day LOS exclusion, the mean score was 91.9%, the median score was 96.9%, and the score for hospices in the 10th percentile distribution was 77.4%. With no LOS exclusions, the mean score was 93.3%, the median score was 97.5%, and the score for hospices in the 10th percentile distribution was 81.6%. The impact of the different LOS criteria on the distribution of QM scores was consistent across quarters.

In summary, these new analyses demonstrate that removing the LOS exclusion in QM calculations, thus including all hospice patients in the denominator, is feasible and appropriate for quality of care at the patient level, and enhance completeness and statistical stability of QM reporting at the hospice level. Further, though the mean and median scores on this QM are high, hospice organizations in the 10th percentile have meaningful opportunity to improve in their timely treatment of dyspnea. Since this quality measure is paired, successful screening (NQF #1639) is linked to subsequent measurement of dyspnea treatment.

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 who screened positive for dyspnea who received treatment within 24 hours of screening.

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.)*
24 hours

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.
Treatment is administered if within 24 hours of the positive screen for dyspnea, medical treatment plan, orders or pharmacy records show inhaled medications, steroids, diuretics, or non-medication strategies such as oxygen and energy conservation. Treatment may also include benzodiazepine or opioid if clearly prescribed for dyspnea.

S.7. Denominator Statement *(Brief, narrative description of the target population being measured)*
Patients enrolled in hospice OR patients receiving hospital-based palliative care for 1 or more days.

S.8. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, 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)*
The Dyspnea Treatment quality measure is intended for patients with serious illness who are enrolled in hospice care OR receive specialty palliative care in an acute hospital setting. Conditions may include, but are not limited to: cancer, heart disease, pulmonary disease, dementia and other progressive neurodegenerative diseases, stroke, HIV/AIDS, and advanced renal or hepatic failure.

For patients enrolled in hospice or palliative care, a positive screen is indicated by any dyspnea noted as other than none on a verbal screen, any number > 0 on a numeric scale or any observational or self-report of dyspnea.
[NOTE: This quality measure should be paired with the Dyspnea Screening quality measure (NQF #1639) to ensure that all patients are screened and therefore given the opportunity to report dyspnea and enter the denominator population for Dyspnea Treatment.]

S.10. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*
Patients with length of stay < 1 day in palliative care, patients who were not screened for dyspnea, and/or patients with a negative screening.

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)*
Calculation of length of stay; discharge date is identical to date of initial encounter.

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)*
N/A

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*)

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.)*
Dyspnea treatment:

a. Step 1- Identify all patients with serious, life-limiting illness who received either specialty palliative care in an acute hospital setting or hospice care

b. Step 2- Identify admission evaluation / initial encounter dates; exclude palliative care patients if length of stay is less than one day. Exclude hospice patients if length of stay is less than 7 days

c. Step 3- Identify patients who were screened for dyspnea during the admission evaluation (hospice) / initial encounter (palliative care)

d. Step 4- Identify patients who screened positive for dyspnea

e. Step 5- Identify patients who received treatment within 24 hours of screening positive for dyspnea

Quality Measure= Numerator: Patients who received treatment for dyspnea in Step 5 / Denominator: Patients in Step 4

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.
Hospice: The hospice analysis was not based on a sample. It was conducted on the entire hospice population that had admission and discharge records in the specified period of analysis.

Palliative care: consecutive sample of equal numbers of admissions + decedents beginning with a randomly selected date. Data collection using a structured chart abstraction tool and operational definition

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.
N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Hospice: For the dyspnea treatment measure, there are two items on the HIS that can include missing data – J2040B and J2040C. If missing, these items are coded as dashes and analyzed as though the process did not happen.

In order to assess how these missing data impact the validity of the dyspnea treatment measure, we conducted the patient stay- and hospice-level analyses. For the patient stay-level analysis, we calculated the number and percentage of eligible patient stays for which the HIS-Admission records included a dash for each of these items. The overall rate of missing data ranged from 0.03 percent to 0.12 percent between October 2014 and September 2015. For the hospice-level analysis we calculated each hospice's percent of eligible admissions that included missing data for items J2040B and J2040C. Over 90% of hospices did not have any admissions with missing data for these items.

Overall, we found that only a very small number of admission records for the eligible stays did not include data, i.e. coded as dash, on dyspnea treatment items J2040B and J2040C. And a vast majority of hospices did not have any missing data for these items. These results indicate that missing data for these items should not have a negative impact on the validity of the QM.

Palliative care: N/A

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record

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.
Hospice: Hospice analysis uses the Hospice Item Set (HIS) as the data source to calculate the quality measure.

Palliative Care: Structured medical record abstraction tool, with separate collection of denominator and numerator data

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 in attached appendix at A.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Group/Practice, Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospice, 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
1638_MeasureTesting_MSF5.0_Data.doc,1638_MeasureTestingAttachment_2.26.16.docx

NQF #: 1638 **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)***:**

Hospice: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1638 - Hospice and Palliative Care Dyspnea Treatment.

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.

Palliative Care: Two research nurse abstractors independently recorded quality measures data on a random subset of 20 seriously ill patients. Abstractors used the pre-defined operational definitions and a structured chart abstraction tool to record numerator and denominator data separately. Patients were a sub-sample of 460 seriously ill patients without specialty palliative care admitted to an acute hospital for at least 1 day to 4 inpatient services from February 2008 to November 2009. Records eligible for sampling included all seriously ill adult patients admitted to medical and surgical intensive care, medically complex patients aged 65 and older admitted to an Acute Care of the Elderly Unit, and medical oncology patients with Stage IV carcinoma.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)***:**

Hospice:

Split-half reliability. Split-half reliability assesses the internal consistency of a QM in different samples by randomly dividing the patient stays within each hospice into two halves, and calculating correlation between the QM scores on the basis of the two randomly divided halves. In this analysis, we conducted a split-half reliability analysis on all facilities with 20 or more patient stays counted in the measure denominator, and used the Interclass Correlation (ICC) coefficients to measure the internal reliability. In general, the ICC coefficient varies between 0 and 1, where an ICC of 0 indicates no reliability and an ICC of 1 indicates perfect reliability.

Signal-to-noise analysis. If a measure is reliable, then true differences in provider performance should explain a substantial proportion of the variance in QM scores. We conducted an analysis of variance (ANOVA) to determine what proportion of total variance in the measure is attributable to differences among providers. A higher proportion indicates better reliability.

Stability analysis. Stability analysis describes the extent to which providers' performance assessed by a QM changes across time. We analyzed the change in facility scores between four consecutive quarters (Q4 2014 and Q1 2015, Q1 and Q2 2015, and Q2 and Q3 2015). The changes in facility scores are reported in standard deviations.

Palliative Care: Inter-rater reliability between the two abstractors was assessed using kappa statistics. Two research nurse abstractors independently recorded quality measures data on a random subset of 20 seriously ill patients. Abstractors used the pre-defined operational definitions and a structured chart abstraction tool to record numerator and denominator data separately.

a2.3 **Testing Results** *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:

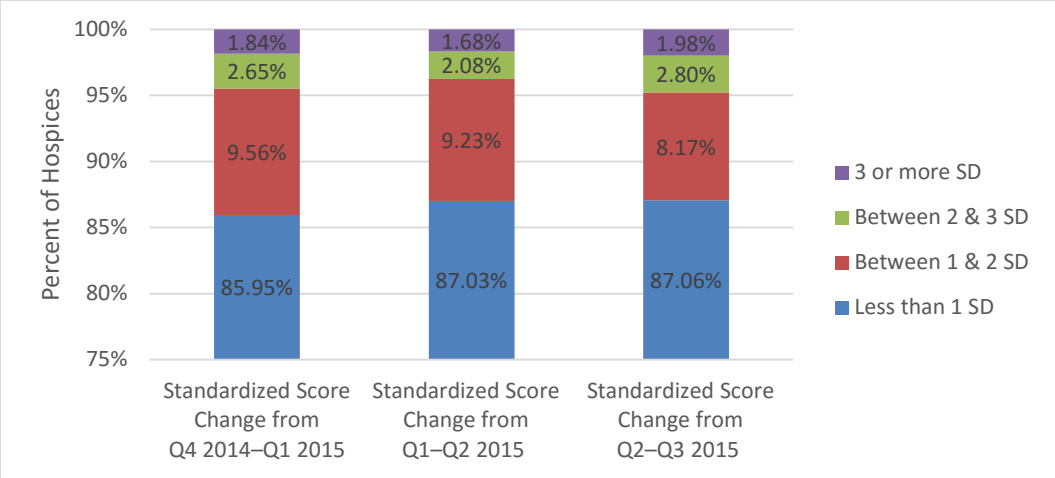
Hospice:

Split-half reliability: In general, the ICC coefficient varies between 0 and 1, where an ICC of 0 indicates no reliability and an ICC of 1 indicates perfect reliability. The ICC coefficient for this measure is 0.86, indicating high internal reliability.

Signal-to-noise analysis: The analysis results found the signal-to-noise ratio to be 0.96, indicating that about 96% of the variance in this measure is because of differences among facilities. This proportion indicates strong reliability for this measure.

Stability analysis: The results of this analysis indicated that facility scores were very stable. Slightly less than 90% of facilities had a change in QM score of less than one standard deviation, indicating high stability of the QM. The number of facilities with a change in QM of less than one standard deviation increased across quarters from 86.0% to 87.1%, suggesting improved reliability across time. Roughly one-tenth of facilities had a change in QM between one and two standard deviations. These results indicate a measure that is generally quite stable. Figure 1 illustrates the change in facility scores between the four consecutive quarters.

Figure 1
Standardized Score Change in QM Score from Q4 2014 to Q3 2015



Palliative Care: Kappa scores range from 0 to 1 with higher scores indicating better agreement. The nurse abstractors achieved excellent inter-rater reliability for this measure: Kappa=0.89. Landis and Koch describe kappa values that range from 0.81 – 0.99 as almost perfect and Fleiss describes kappas over 0.75 as excellent.

Landis, J.R.; Koch, G.G. (1977). "The measurement of observer agreement for categorical data". Biometrics 33 (1): 159–174

Fleiss, J.L. (1981). Statistical methods for rates and proportions (2nd ed.). New York: John Wiley

2b. VALIDITY. Validity, Testing, including all Threats to Validity: 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:**

The measure focus is Dyspnea Treatment, designed to pair with Dyspnea Screening to ensure quality care processes for dyspnea. The target populations are hospice patients, and seriously ill hospitalized patients with diverse underlying diagnoses who are at high risk for palliative care clinical needs, including dyspnea.

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)***:**

Hospice: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1638 - Hospice and Palliative Care Dyspnea Treatment.

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.

Palliative Care: The total patient sample size was 562. Chart abstractions were completed for 102 consecutive seriously ill patients with specialty palliative care consultation, and a random sample of 460 seriously ill patients without specialty palliative care admitted to an acute care hospital for at least 1 day to four inpatient services with high proportions of seriously ill patients. Records eligible for sampling included all patients admitted to medical and surgical intensive care,

medically complex patients aged 65 and older admitted to a Geriatric Evaluation Unit, and medical oncology patients with Stage IV carcinoma. Because palliative care domains become even more relevant closer to death, patients dying in hospital were over-sampled to ensure a final ratio of 1 decedent to 1 live discharge. Consistent with oversampling of decedent records, 55% of these patients died in hospital.

The age of the patients ranged from 16 to 99 years, with the mean age 61. Patients were predominantly Caucasian (65%), with smaller subgroups who were African American (24%) and Hispanic / Latino (4%) The most common life-limiting diagnoses were infections (37%), cancer (34%), pulmonary (29%), and neurologic diseases (21%).

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

Hospice:

Correlations. Providers should perform similarly on QMs that reflect the quality of similar care processes. Thus, a common strategy used to evaluate validity of a QM is to examine the relative performance of the provider on QMs that are conceptually related, either due to some core competency manifested across a stay or for processes that occur in close proximity. We conducted nonparametric Spearman rank correlation analysis among all seven HQRP QMs, which address care processes around hospice admission that are clinically recommended or required in the hospice Conditions of Participation. We used nonparametric methods since the data are heavily skewed (most providers perform well, while a small proportion have lower scores). Since all of the seven current QMs measure desirable care assessment processes upon hospice admission, higher positive Spearman correlation coefficients among these QMs will demonstrate higher validity of the QM set. Statistically significant correlations, indicated by p-values of the Spearman correlation coefficients less than 0.05, also support the validity of the QM set.

Missing data. Missing data represent a potential threat to the validity of an HIS item, which in turn may harm the validity of the QM that relies on the item. For the dyspnea treatment measure, there are two items on the HIS that can include missing data, which are coded as dashes – J2040B and J2040C. In order to assess how these missing data impact the validity of the dyspnea treatment measure, we conducted the patient stay- and hospice-level analyses.

Palliative Care sample:

Face validity of PEACE quality measures for hospital-based specialty palliative care was addressed using stakeholder review and feedback. Investigators prepared data reports in a summary format with detailed operational definitions, and led a 1-hour discussion with nursing and physician leaders from each service group – MICU, SICU, Acute Care for the Elderly (Geriatrics), Oncology, and Palliative Care. The discussion included feedback of quality measure data, response to questions and critiques, and eliciting stakeholder feedback about the validity and actionability of this data for the care of their patients. Stakeholders were specifically asked to comment on the accuracy of the data as a reflection of current care practices, and their highest priority area for future quality improvement.

Construct validity was tested by comparing the PEACE quality measures for patients seen by specialty interdisciplinary palliative care consultants to those not receiving specialty palliative care services.

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)***:**

Hospice:

Correlations : Table 1 presents the Spearman correlation coefficients. The p-values for all the Spearman correlation coefficients are significant ($p < 0.001$). The significant positive correlations between every pair of QMs indicate that high-

performing hospices in one area also provided high-quality care in other areas at hospice admission. Overall, the correlations between the QMs are low to moderate. The clustering of QM scores, skewed distributions, and low variability across all seven QMs may affect the level of correlations between QMs.

Table 1
Correlation of Hospice QMs, Percentile Ranking

Quality measure	NQF #1647 (modified) Beliefs/Values Addressed	NQF #1634 Pain Screening	NQF #1637 Pain Assessment	NQF #1639 Dyspnea Screening	NQF #1638 Dyspnea Treatment	NQF #1617 Bowel Regimen
NQF #1641 Treatment Preferences	0.64***	0.37***	0.21***	0.44***	0.28***	0.25***
NQF #1647 (modified) Beliefs/Values Addressed		0.42***	0.17***	0.43***	0.31***	0.25***
NQF #1634 Pain Screening			0.17***	0.41***	0.28***	0.15***
NQF #1637 Pain Assessment				0.17***	0.08***	0.20***
NQF #1639 Dyspnea Screening					0.30***	0.25***
NQF #1638 Dyspnea Treatment						0.34***

NOTE: The correlation is on the basis of each hospice's percentile ranking on the QMs.
*** indicates significant correlation at $p < 0.001$

Overall, NQF #1638, Dyspnea Treatment, has significant positive correlations with the other QMs, indicating hospices providing higher-quality care in this area also performed better in other areas at hospice admission.

Missing data: For the patient stay-level analysis, we calculated the number and percentage of eligible patient stays for which the HIS-Admission records included a dash for each of these items. The overall rate of missing data ranged from 0.03 percent to 0.12 percent between October 2014 and September 2015. For the hospice-level analysis we calculated each hospice's percent of eligible admissions that included missing data for items J2040B and J2040C. Over 90% of hospices did not have any admissions with missing data for these items.

Overall, we found that only a very small number of admission records for the eligible stays did not include data, i.e. coded as dash, on dyspnea treatment items J2040B and J2040C. And a vast majority of hospices did not have any missing data for these items. These results indicate that missing data for these items should not have a negative impact on the validity of the QM.

Palliative Care sample:

Face Validity: Stakeholder discussions provided broad endorsement of face validity, with some considerations for specific patient populations. Medical oncologists endorsed the face validity of these quality measures, but favored quality measures endorsed by oncology professional organizations.

Construct Validity: Screening for dyspnea was nearly universal for all seriously ill patients, but was more consistently done by specialty palliative care providers (100% vs 95%, p=0.016). Patients with dyspnea were likely to receive some

form of treatment within 24 hours, with or without the addition of specialty palliative care (96% vs 93%, p=NS).

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)***:**

Hospice: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1638 - Hospice and Palliative Care Dyspnea Treatment.

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.

Palliative Care: N/A

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

Hospice: Under contract to the Centers for Medicare and Medicaid Services (CMS), the Research Triangle Institute (RTI) performed descriptive analyses on 5 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.

Palliative Care: N/A

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

Hospice: At the patient level, approximately 40% of patients were excluded based on the LOS < 7 days exclusion, thus omitting this essential quality of care standard for many seriously ill patients. Further, excluding short-stay patients omits measures of quality of care for a subset of the sickest patients entering hospice care, and those who arguably may be in greatest need of rapid symptom assessment and treatment. In addition, the original rationale for this exclusion -- allowing time for hospice providers to complete this care process -- does not appear to be necessary. Our analyses show that a large portion of dyspnea treatments were performed on day 1 of admission to hospice, demonstrating a normative standard of care includes prompt attention to symptom distress.

At the hospice level, applying different LOS exclusion criteria had an impact on denominator size. Smaller denominator size at the hospice level may generate less stable and reliable QM scores. Under the LOS < 7 days exclusion, the median number of qualifying stays in the denominator was 41 stays per hospice vs. a median number of 63 stays per hospice with no LOS exclusion. Applying the LOS exclusion resulted in some hospices with no qualifying stays, which excluded 33 hospices from QM score calculations altogether.

Applying or removing the LOS exclusion generally had little impact on the distribution of hospices' QM scores. Under the 7-day LOS exclusion, the mean score was 91.9%, the median score was 96.9%, and the score for hospices in the 10th percentile distribution was 77.4%. With no LOS exclusions, the mean score was 93.3%, the median score was 97.5%, and the score for hospices in the 10th percentile distribution was 81.6%. The impact of the different LOS criteria on the

distribution of QM scores was consistent across quarters.

In summary, these new analyses demonstrate that removing the LOS exclusion in QM calculations, thus including all hospice patients in the denominator, is feasible and appropriate for quality of care at the patient level, and enhance completeness and statistical stability of QM reporting at the hospice level. Further, though the mean and median scores on this QM are high, hospice organizations in the 10th percentile have meaningful opportunity to improve in their timely treatment of dyspnea. Since this quality measure is paired, successful screening (NQF #1639) is linked to subsequent measurement of dyspnea treatment.

Palliative Care: N/A

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)***:**

N/A

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

N/A

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)***:**

N/A

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

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)***:**

Hospice: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1638 - Hospice and Palliative Care Dyspnea Treatment.

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.

Palliative Care: The total patient sample size was 562. Chart abstractions were completed for 102 consecutive seriously ill patients with specialty palliative care consultation, and a random sample of 460 seriously ill patients without specialty palliative care admitted to an acute care hospital for at least 1 day to four inpatient services with high proportions of seriously ill patients from February 2008 to November 2009. Records eligible for sampling included all patients admitted to medical and surgical intensive care, medically complex patients aged 65 and older admitted to a Geriatric Evaluation Unit, and medical oncology patients with Stage IV carcinoma. Because palliative care domains become even more relevant closer to death, patients dying in hospital were over-sampled to ensure a final ratio of 1 decedent to 1 live discharge. Consistent with oversampling of decedent records, 55% of these patients died in hospital.

The age of the patients ranged from 16 to 99 years, with the mean age 61. Patients were predominantly Caucasian (65%), with smaller subgroups who were African American (24%) and Hispanic / Latino (4%) The most common life-limiting diagnoses were infections (37%), cancer (34%), pulmonary (29%), and neurologic diseases (21%).

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

Hospice: Confidence interval analysis. We examined proportions of hospices with the QM scores that are significantly different from the national hospice-level mean. If a high proportion of hospices have a measure score significantly different from the mean, the QM can identify facilities with different levels of performance. For this analysis, statistical significance was determined using 95 percent confidence intervals: a hospice's QM score was significantly different from the national mean if the national mean was not included within the hospice's 95 percent confidence interval. High-performing facilities should have scores that are significantly below average, and low-performing facilities should be significantly above average.

Palliative Care: Construct validity was tested by comparing the PEACE quality measures for patients seen by specialty interdisciplinary palliative care consultants to those not receiving specialty palliative care services. Percentage of patients with and without specialty palliative care for whom the quality measure was met was compared for difference using chi-square statistics.

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)*:

Hospice:

The mean score for this QM was 93.3% with a range from 0% to 100%, the median was 97.5%, the interquartile range was 8.3, and the standard deviation was 11.2. For this measure, 31.3% of hospices had perfect scores and 21.7% of hospices scored below 90%.

Scores by decile:

10th percentile 81.6%

25th percentile 91.7%

Median 97.5%

75th percentile 100%

90th percentile 100%

Across all hospices, 39.8% had a QM score that is significantly different than the national mean. Hospices were more likely to report scores above the national mean than below the national mean (27.0% vs 12.8%, respectively, overall). The QM is

able to identify those hospices that are performing well (higher than the national mean) and those that are performing less well (lower than the national mean).

Palliative Care:

Seriously ill patients with palliative care sample: 96% met quality measure

Seriously ill patients without palliative care: 93% (p=NS)

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)***:**

N/A

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

N/A

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)***:**

N/A

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)***:**

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

Disparities in dyspnea screening and treatment have not been well characterized in the hospice and palliative care population. Future research with larger sample sizes can be used to test for differential performance by race / ethnicity and by gender.

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)*

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

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.

Missing data represent a potential threat to the validity of an HIS item, which in turn may harm the validity of the QM that relies on the item. For the dyspnea treatment measure, there are two items on the HIS that can include missing data – J2040B and J2040C. If missing, these items are coded as dashes.

In order to assess how these missing data impact the validity of the dyspnea treatment measure, we conducted the patient stay- and hospice-level analyses. For the patient stay-level analysis, we calculated the number and percentage of eligible patient stays for which the HIS-Admission records included a dash for each of these items. The overall rate of missing data ranged from 0.03 percent to 0.12 percent between October 2014 and September 2015. For the hospice-level analysis we calculated each hospice's percent of eligible admissions that included missing data for items J2040B and J2040C. Over 90% of hospices did not have any admissions with missing data for these items.

Overall, we found that only a very small number of admission records for the eligible stays did not include data, i.e. coded as dash, on dyspnea treatment items J2040B and J2040C. And a vast majority of hospices did not have any missing data for these items. These results indicate that missing data for these items should not have a negative impact on the validity of the QM. There was no missing data for the elements needed to calculate this subset of dyspnea measures for the Palliative Care samples.

Record abstraction does not require collection of unique patient identifiers and thus protects confidentiality. Timing of data collection can be concurrent with admission / initial encounter care, or can be retrospective based on medical record sampling.

Costs have not been formally estimated; medical record abstraction or electronic capture of the elements of a dyspnea screen and treatment will have more modest cost compared to survey data.

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)*.
N/A

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)
Public Reporting	Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Hospice Item Set-Hospice Quality Reporting Program https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Hospice-Quality-Reporting/Hospice-Item-Set-HIS.html Quality Improvement (Internal to the specific organization) Voluntary use by UNC Palliative Care Program http://www.med.unc.edu/pcare/resources/PEACE-Quality-Measures

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

Name of program and sponsor: Hospice: HIS/CMS; Palliative Care: Voluntary use by PC organizations

Purpose- Hospice: Quality reporting for hospice; Palliative Care: Internal quality improvement

Geographic area and number and percentage of accountable entities and patients included – United States and all accountable entities and 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**

a. Due to insufficient longitudinal data, we are currently unable to discuss progress.
b. Analyses encompassed 3,922 hospice organizations and 1,218,786 patient stays across the United States.
Region:
South: 39.3%
West: 25.1%
Midwest: 23.1%
Northeast: 11.3%
Territories: 0.94%
Unknown: 0.25%

Urban/rural status:
Urban: 75.8%
Rural: 23.9%
Unknown: 0.31%

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.**
We only have one year of data to report, which is not enough to show trends over time.

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.**
None

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.)
This measure is part of the NPCRC Key Palliative Measures Bundle. Refer to the NPCRC cover letter and table of bundle measures for description of the selection and harmonization of the Key Palliative Measures Bundle.

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 **Attachment:** Appendix_A.1_NQF_1638.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** University of North Carolina-Chapel Hill
Co.2 **Point of Contact:** Laura, Hanson, lhanson@med.unc.edu, 919-843-4096-
Co.3 **Measure Developer if different from Measure Steward:** University of North Carolina- Chapel Hill
Co.4 **Point of Contact:** Laura, Hanson, lhanson@med.unc.edu, 919-843-4096-

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.
The Carolinas Center for Medical Excellence PEACE Project Technical Expert Panel

The PEACE project team convened a 14-member Technical Expert Panel (TEP) of nationally recognized experts with extensive experience in the following areas: medical or nursing expertise in hospice and palliative care, methods and instrumentation, and quality improvement. Using criteria provided by the CCME study team, TEP members rated each potential quality measure on four criteria: importance, scientific soundness, feasibility and usability.

Mary Ersek, PhD, RN, Research Associate Professor, Swedish Medical Center- Pain Research Department, Seattle, WA
Betty R. Ferrell, PhD, FAAN, Research Scientist, City of Hope National Medical Center, Duarte, CA

Sean Morrison, MD, Mount Sinai Medical Center, NY, NY
Richard Payne, MD, Director, Duke Institute on Care at the End of Life, Duke Divinity School, Durham, NC
Chris Feudtner, MD, PHD, MPH, Children's Hospital of Philadelphia, Philadelphia, PA
Karen Steinhauser, PhD, Research Health Scientists, Center for Health Services Research in Primary Care, Durham VA Medical Center and Duke University, Durham, NC
Joan M. Teno, MD, Professor of Community Health and Medicine, Center for Gerontology and Health Care Research, Brown University, Providence, RI
Melanie Merriman, PhD, MBA, Touchstone Consulting, North Bay Village, FL
Sydney Dy, MD, MSc, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
David Casarett, MA, MD, Assistant Professor, Division of Geriatrics, Institute on Aging and Center for Bioethics, University of Pennsylvania School of Medicine and NHPCO Board of Directors
Judi Lund-Person, Vice President, Division of Quality, National Hospice and Palliative Care Organization, Washington, DC
Jean Kutner, MD, MSPH, Associate Professor, University of Colorado Health Sciences Center, Denver, CO
Lin Simon, Analyst, National Hospice and Palliative Care Organization, Washington, DC
Karen Pace, NAHC

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

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

Ad.8 **Additional Information/Comments:** This measure is part of the NPCRC Key Palliative Measures Bundle. Refer to the NPCRC cover letter and table of bundle measures for description of the selection and harmonization of the Key Palliative Measures Bundle.

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 #: 1639
Measure Title: Hospice and Palliative Care -- Dyspnea Screening
Measure Steward: University of North Carolina-Chapel Hill
Brief Description of Measure: Percentage of hospice or palliative care patients who were screened for dyspnea during the hospice admission evaluation / palliative care initial encounter.
Developer Rationale: Dyspnea is prevalent and undertreated for many populations of seriously ill patients, including those patients nearing the end of life. Screening for dyspnea is necessary to determine its presence and severity, and forms the basis for treatment decision-making. Unlike pain, structured clinical assessment of the symptom is less well-defined, yet similar to pain, effective treatment is available to alleviate symptom distress.

Numerator Statement: Patients who are screened for the presence or absence of dyspnea and its severity during the hospice admission evaluation / initial encounter for palliative care.
Denominator Statement: Patients enrolled in hospice OR patients receiving hospital-based palliative care for 1 or more days.
Denominator Exclusions: Patients with length of stay < 1 day in palliative care.

Measure Type: Process
Data Source: Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record
Level of Analysis: Clinician : Group/Practice, 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**

Summary of prior review in 2012

- The developer provides a rationale for the importance of dyspnea screening, assessment and treatment. An explicit link to patient outcomes is not provided.
- The developer cited clinical guidelines and systematic reviews of randomized trials and small trials supporting 1) the use of opioids for breathlessness; 2) shorter follow-up times; 3)the use of beta agonists for dyspnea in COPD; 4) the use of oxygen for hypoxic patients; and 5)the benefit of coping or relaxation interventions. However, none of these sources appear to support a link between screening for dyspnea and improved patient outcomes.
- During its 2012 evaluation of the measure, the Palliative/End-of-Life (EOL) Committee agreed there was no evidence demonstrating that solely screening for dyspnea leads to better outcomes; however, members noted that screening is a necessary step leading to treatment.
- This measure is paired with # 1638: Hospice and Palliative Care- Dyspnea Treatment

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:

- Although the developer attested that there have been no changes in the evidence since it was endorsed in 2012, they added two new guidelines to the submission.
    - The 2011 British Columbia Medical Services Commission guideline relates to palliative care for patients with incurable cancer or advanced disease; it is based on a systematic review of evidence but the evidence is not graded. This guideline calls for the assessment of dyspnea severity using a 1-10 scale.
    - The 2013 ICSI guideline on Palliative Care for Adults. The pertinent recommendation from this graded guideline states that "The physical aspects of the patient's serious illness should be an integral component of the palliative care plan (Low Quality Evidence, Strong Recommendation)". The guideline goes on to say "Each patient should be frequently evaluated for these issues."

Exception to evidence:

- Because the evidence for this measure is mostly tangential to the relationship of dyspnea screening to patient outcomes or based on expert opinion only, it is insufficient to meet NQF's criterion for evidence. However, an exception to the evidence criterion is allowed if the Committee agrees that empirical evidence is not needed to hold providers accountable for the measure and that other outcome measures or evidence-based measures are not available or feasible at this time.

Guidance from the Evidence Algorithm

Process measure but graded systematic reviews mostly tangential evidence (Box 3) → Other evidence not submitted (Box 7) → A process measure for dyspnea treatment is available (#1638) (Box 10) → Systematic assessment of expert opinion (Box 11) → If Committee agrees it is OK/beneficial to hold providers accountable for performance in the absence of empirical evidence of benefits to patients → rate as INSUFFICIENT WITH EXCEPTION

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?*
- *Are there, or could there be, performance measures of a related health outcome, OR evidence-based intermediate clinical outcomes, intervention/treatment?*
- *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*
- *Does the SC agree that it is acceptable (or beneficial) to hold providers accountable without empirical evidence?*

Preliminary rating for evidence: ☐ **High** ☐ **Moderate** ☐ **Low** ☒ **Insufficient**
Rationale: No evidence links screening for dyspnea to improved patient outcomes.

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.

- Facility-level (Hospice):
  - Data from the Hospice Quality Reporting Program (HQRP) for FY15 are provided. The data are from 3,922 hospice organizations and approximately 1.2 million patient stays.
  - The developer reported that 37.1% of hospices had perfect scores, and 6.7% of hospices scored below 90%.

Mean	97.3%
Range	0% - 100%
10th percentile	92.6%
25th percentile	97.1%
Median	99.2%
75th percentile	100%
90th percentile	100%

- Clinician Group/Practice Level (Palliative Care):
  - Data specific to palliative care (for clinician group/practice in the hospital setting) are not yet available, although the developer expects these data will become available for NQF review next year

Disparities

- Facility-level (Hospice):

Subgroup	**Hospice-level**	
	Score	**Wilcoxon-Mann-Whitney test p-value**
Race		
Non-white ≥ national median of 11.9%	97.6%	P=0.008
Non-white < national median of 11.9%	96.9%	
Sex		
Female ≥ national median of 55.2%	97.3%	p=0.96
Female < national median of 55.2%	97.2%	
Medicaid status (proxy for SES)		
Medicaid patients >= national median of 21.5%	96.9%	p<0.001
Medicaid patients < national median of 21.5%	97.7%	
Geographic location		
Rural	97.6%	P<0.001
Urban	97.1%	

- Clinician Group/Practice Level (Palliative Care):
  - Disparities data for palliative care are not available, although the developer expects these data will become available for NQF review next year

Questions for the Committee:

- *Although there are statistically significant differences in performance between hospices for particular subgroups,*

the developer notes these may not be clinically meaningful. Is there a gap in hospice care (for dyspnea screening) that warrants a national performance measure? Is the measure "topped out" for hospices?

- *Is the Committee aware of evidence demonstrating a gap in hospital-based palliative care (for dyspnea screening) that warrants a national performance measure?*
- *Palliative-specific disparities information is not provided. Are you aware of evidence that disparities exist in this area of healthcare?*

Preliminary rating for opportunity for improvement: ☐ High ☐ Moderate ☒ Low ☐ Insufficient
Rationale: Most hospices are reporting on this measure, and performance seems to be topped out.

Committee pre-evaluation comments

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

1a.
* Process Measure for Maintenance Review.
•During its 2012 evaluation of the measure, the Palliative/End-of-Life (EOL) Committee agreed there was no evidence demonstrating that solely screening for dyspnea leads to better outcomes; however, members noted that screening is a necessary step leading to treatment.
* PROCESS:
HOSPICE: This process measure applies directly to ALL patients admitted to hospice (paired with NQF #1638 Dyspnea Treatment)

Dyspnea screening and assessment are necessary to detect the presence of dyspnea (for which physical signs such as hypoxia and tachypnea do not clearly correlate), and to understand its severity and underlying etiology. Evidence-based treatment of dyspnea will vary with its severity and etiology, with treatment options differing for causes such as malignant pleural effusion, bulky tumor mass, congestive heart failure, anemia, COPD, among others.

1b.
* Exception to Evidence: • Because the evidence for this measure is mostly tangential to the relationship of dyspnea screening to patient outcomes or based on expert opinion only, it is insufficient to meet NQF's criterion for evidence. However, an exception to the evidence criterion is allowed if the Committee agrees that empirical evidence is not needed to hold providers accountable for the measure and that other outcome measures or evidence-based measures are not available or feasible at this time.
Preliminary rating for opportunity for improvement: Low. No data available for palliative care at this time.
* HOSPICE: 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 resulted in significant diparities in screening.

Palliative Care: Disparities data not available

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 sources specified in the submission form include EHRs and other electronic clinical data.

Specifications:

- This measure is specified for the clinician group/practice level of analysis in the hospital setting and for the facility level of analysis in the hospice setting. A higher score indicates better quality.
- For hospice settings, the numerator (patients who are screened for dyspnea and asked to rate its severity) and denominator (patients enrolled in hospice) are identified from the Hospice Item Set, a standardized, patient-level dataset used in the Medicare Hospice Quality Reporting Program.
- For palliative care settings, data for the numerator (patients screened for dyspnea and asked to rate its severity) and denominator (patients admitted to palliative care) are collected using a structured medical record

abstraction tool. For this setting, patients with a length of stay <1 day are excluded from the denominator.

- Screening may be completed using verbal, numeric, visual analog, or rating scales designed for use with non-verbal patients.
- A calculation algorithm is provided.
- The developer indicates that sampling is permissible for the hospital setting. While some basic instructions are given, there is no guidance about the number of patients needed for the sample.

Prior Evaluation

- In the 2012 evaluation of this measure, the Committee expressed concern about consistent documentation of dyspnea screening.

Changes to specifications since previous evaluation:

- After analysis of FY15 hospice data, the developer has changed the specifications so as to no longer exclude hospice stays of less than 7 days.

Questions for the Committee :

- *Are all the data elements clearly defined?*
- *Is the logic or calculation algorithm clear?*
- *Is it reasonable to exclude palliative care patients with < 1 day length of stay?*
- *Do all hospitals use EHRs or are some still using paper records (note: the clinician group/practice measure for the hospital setting specifies use of EHRs only)?*
- *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:

- Previous reliability included inter-rater reliability testing of hospital data for 20 patients. Reliability testing using hospice data was not previously conducted.

Describe any updates to testing

- Score-level testing for the hospice setting included split-half reliability, signal-to-noise, and stability analyses at the facility level.

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

- Clinician Group/Practice Level (Palliative Care):
  - Reliability of the data elements was assessed using data obtained from the medical charts for 20 seriously ill patients without specialty palliative care admitted to an acute care hospital. Inter-rater reliability between two nurse abstractors was calculated. This is an appropriate method for demonstrating data element reliability.
- Facility-level (Hospice):
  - Reliability of the measure score was assessed using FY15 data from the Hospice Quality Reporting Program (n=3,992 hospice organizations and n=1,218,786 patient stays). Two types of reliability testing were conducted: the first via a signal-to-noise analysis and the second using a split-sample (or "test-retest") methodology. Both are appropriate methods for testing reliability. Stability analysis examined changes in provider's performance over time. 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

- Clinician Group/Practice level (Palliative Care:
  - Developers report a single kappa value of 0.91
  - 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 0.91 means that the raters agreed 91.0% 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.
  - NQF guidance indicates that data element testing should be conducted for all critical data elements, although at minimum, results about the numerator, denominator, and exclusions should be provided. For this measure, only a single kappa value is reported. Although not explicitly stated, we assume that the 20 charts that were abstracted included only seriously ill patients who received specialty palliative care in an acute hospital setting and whose length of stay was at least one day, and thus this kappa value applies to the numerator. .
- Facility-level (Hospice)
  - Split-half analysis ICC=0.83.
    - The ICC reflects the percentage of variance in score results that is due to "true" or real variance between the hospices. A value of 0.7 is often regarded as a minimum acceptable reliability value.
  - Developers report a signal-to-noise ratio of 0.98.
    - A signal-to-noise analysis quantifies the amount of variation in a performance measure that is due to true differences between hospices (i.e., signal) as opposed to random measurement error (i.e., noise). Results will vary based on the amount of variation between the providers and the number of patients treated by each provider. This method results in a reliability statistic that ranges from 0 to 1 for each hospice. A value of 0 indicates that all variation is due to measurement error and a value of 1 indicates that all variation is due to real differences in hospices performance. A value of 0.7 often is regarded as a minimum acceptable reliability value.

Guidance from the Reliability Algorithm
Facility-level (Hospice):
Precise specifications (Box 1) → Empirical reliability testing with measure as specified (Box 2) → Score-level testing (Box 4) → Appropriate method (Box 5) → High certainty that measure results are reliable (Box 6a) → High

Clinician-level (Palliative):
Precise specifications (Box 1) → Empirical reliability testing with measure as specified (Box 2) → Score-level testing not conducted (Box 4) → Data elements tested (Box 8) → Appropriate method (Box 9) → High level of agreement between raters → Moderate

Questions for the Committee:

- *Are there any concerns about accurately and consistently identifying the denominator and exclusions in the hospital setting?*
- *In the 2012 evaluation of the measure, the Committee raised concerns that the numerator data may not be consistently documented. Is this still a concern?*
- *Is the test sample adequate to generalize for widespread implementation for palliative care? If not, is current testing sufficient until more data are available (assuming the developer is planning to do additional testing)?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

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**

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:

- In the prior review, the measure was tested for the palliative care (hospital) setting using face and construct validity.

Describe any updates to validity testing:

- Additional validity testing of the measure score at the facility level of analysis (for the hospice setting) was conducted using FY15 data from the Hospice Quality Reporting System (HQRP).

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

Validity testing method:

- Facility-level (Hospice):
  - Using FY15 data from the Hospice Quality Reporting System (n=3,992 hospice organizations; n=1,218,786 patient stays), the developer conducted non-parametric Spearman rank correlation analysis between this measure and 5 other hospice quality measures. They hypothesized that agencies should perform similarly on assessment processes at hospice admission and expected the resulting correlations to be high. This is an appropriate method of score-level validation.
- Clinician Group/Practice Level (Palliative Care)
  - Face validity was assessed by a group of nursing and physician stakeholders who were asked to comment on the validity, accuracy, and actionability of the measure. According to NQF guidance, the face validity of the measure score as a quality indicator may be adequate if it explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. It does appear that this testing conforms to NQF's requirements for face validity.
  - Construct validity was tested by comparing measure results for seriously ill patients seen in specialty interdisciplinary palliative care consultations (n=102) in one hospital to those who did not receive these services (n=460). The developers did not explain what they expected to find with this analysis and how it would demonstrate that the measure results reflect quality of care.

Validity testing results:

- Facility-level (Hospice)
  - Correlation results were positive and statistically significant, confirming the developer's hypothesis. However, the magnitude of the correlation was lower than they had expected. The developer states reasons for this may be due to clustering of scores, skewed distributions, and low score variability.

- Clinician Group/Practice (Palliative Care):
  - Face validity results from the stakeholder group indicated broad endorsement of the face validity of the measure.
  - Construct validity results found patients with dyspnea were more likely to be screened for dyspnea when specialty palliative care consults were in place than when they were not (100% vs 95%, p=0.016).

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *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 examined the effect of removing the previous length of stay <7 days exclusion criterion for hospice patients. Using FY15 data from the HQRP, the developers found that the removal of the criterion increased the average size of the denominator per hospice organization and had little effect on the distribution of the measure scores. The developers stated these findings supported the removal of the length of stay criterion as an exclusion criterion from the measure.
- The developer did not include information on exclusion analyses for palliative care patients.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the 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)*:

- Meaningful differences among hospices were examined using FY15 data from the Hospice Quality Reporting Program. The developers examined 95% confidence intervals to determine the proportion of hospices with results significantly different from the national hospice-level mean. Results indicated 36.7% of hospices had a quality measure score that was significantly difference from the national mean. Hospices were more likely to report scores above the national mean than below the national mean (25.2% vs 11.5%, respectively).
- Developers compared measure score results for patients in one acute care hospital who received palliative care consults to those who did not receive palliative care consults. However, this analysis does not speak to whether the measure results reflect meaningful differences between clinician groups at different acute care hospitals.

Question for the Committee:

- *Does this measure identify meaningful differences about quality between hospice facilities? Between clinician groups/practices in the hospital setting?*

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

2b7. Missing Data

- For hospices, the rate of missingness is low (0.02% - 0.12% at the patient-level). More than 90% of hospices had no missing information for the two key data elements from the Hospice Item set that make up this measure.
- The developer does not provide any information on missing data specific to palliative care.

Guidance from the Validity Algorithm
Facility-level (Hospice)
Specifications consistent with evidence (Box 1) → potential threats to validity assessed (Box2) →empirical validity testing (Box 3) → validity testing of the measure score (Box 6) →method appropriate (Box7) →moderate certainty (Box 8b)→ Moderate

Clinician group/practice level (Palliative Care)
Specifications consistent with evidence (Box 1) → potential threats to validity only somewhat assessed (Box2) → empirical validity testing (Box 3) → validity testing of the measure score (Box 6) →method not well described (Box7) →face validity systematically assessed (Box 4) → results indicate substantial agreement as to validity (Box 5)→ Moderate (assuming no concerns around exclusions, missing data, or meaningful differences).

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

Committee pre-evaluation comments

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

2a, 2b.

Reliability – Specifications

* Per measure tool:

Screening may be completed using verbal, numeric, visual analog, or rating scales designed for use with non-verbal patients.

No concerns that this measure can be consistently implemented. Each EHR or system would need documentation area for this screening and I note this was the same concern for 2012 evaluation of the measure.

Validity – Specifications

* Specifications consistent with the evidence.

Reliability – Testing

* Each EHR or system would need documentation area that this for screening was done, and if is a visual screen, is there a cue for documentation, if results negative. This was the same concern for 2012 evaluation of the measure.

* HOSPICE: 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.

Palliative Care: Two research nurse abstractors independently recorded quality measures data on a random subset of 20 seriously ill patients. Abstractors used the pre-defined operational definitions and a structured chart abstraction tool to record numerator and denominator data separately. Patients were a sub-sample of 460 seriously ill patients without specialty palliative care admitted to an acute hospital for at least 1 day to four inpatient services from February 2008 to November 2009

Validity Testing

* Adequate scope to generalize for widespread implementation: Using FY15 data from the Hospice Quality Reporting System (n=3,992 hospice organizations; n=1,218,786 patient stays)

Agree the screen is needed to provide the treatment for any seriously ill within dyspnea and that this is then an indicator of quality.

* EXCLUSION OF <7 DAY LOS: In summary, these new analyses demonstrate that removing the LOS exclusion in QM calculations, thus including all hospice patients in the denominator, is feasible and appropriate for quality of care at the patient level, and enhance completeness and statistical stability of QM reporting at the hospice level. Further, though the mean and median scores on this QM are high, hospice organizations in the 10th percentile have meaningful opportunity to improve in their timely screening of dyspnea. Since this quality measure is paired, successful screening is linked to subsequent measurement of dyspnea treatment (NQF #1639).

Threats to Validity

* The developer examined the effect of removing the previous length of stay <7 days exclusion criterion for hospice patients. Using FY15 data from the HQRP, the developers found that the removal of the criterion increased the average size of the denominator per hospice organization and had little effect on the distribution of the measure scores. The developers stated these findings supported the removal of the length of stay criterion as an exclusion criterion from the measure.

The developer did not include information on exclusion analyses for palliative care patients.

I agree with the removal of the length of stay criterion as an exclusion criterion from the measure.

For hospices, the rate of missingness is low (0.02% - 0.12% at the patient-level). More than 90% of hospices had no missing information for the two key data elements from the Hospice Item set that make up this measure.

* Over 90% of hospices did not have any admissions with missing data for these items.

Overall, we found that only a very small number of admission records for the eligible stays did not include data, i.e. coded as dash, on dyspnea screening items J2030B and J2030C. And a vast majority of hospices did not have any missing data for these items. These results indicate that missing data for these items should not have a negative impact on the validity of the QM.

Palliative care: N/A

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 are routinely collected and are incorporated in hospice providers' electronic clinical documentation system (i.e., electronic health record (EHR)). Data collection for all measures is achieved through use of the Hospice Item Set (HIS). It is unclear if the HIS is automatically populated by the EHR.
- The developer does not provide any information on potential or actual implementation challenges for palliative care.
- The developer reports cost estimates are not available. According to the Guidance Manual for Completion of the Hospice Item Set, completing an entire response for any one admission takes 20 minutes (though seven performance measures, of which this measure is only one, are collected as part of this process).
- In the 2012 endorsement of the measure, the Palliative and End-of-Life Care Steering Committee noted that a substantial data collection effort may be needed to abstract data if electronic data are not available.

Questions for the Committee:

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

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

Committee pre-evaluation comments

Criteria 3: Feasibility

* Required data elements are routinely collected and are incorporated in hospice providers' electronic clinical documentation system (i.e., electronic health record (EHR)). Data collection for all measures is achieved through use of the Hospice Item Set (HIS). It is unclear if the HIS is automatically populated by the EHR.
The developer does not provide any information on potential or actual implementation challenges for palliative care.

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

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

Accountability program details:

- This measure is included in the Hospice Quality Reporting Program (HQRP). Medicare-certified hospices are

required to submit a Hospice Item Set (HIS)-Admission record and an HIS-Discharge record for each patient admission on or after July 1st, 2015. In FY2015, 3,992 hospice organizations provided measure data for 1,218,786 patient stays. The distribution of reporting hospice agencies by geographic area is provided.

Improvement results:

- Longitudinal data for this measure 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 2012, the Measure Applications Partnership (MAP) supported the inclusion of this measure in the Hospice Quality Reporting program. Public comments generally concurred with the recommendation.
- In February 2013, the Measure Application Partnership supported the measure for inclusion in the Hospice Quality Reporting Program and the Physician Quality Reporting System (PQRS) to address a core concept (i.e., pain assessment) not addressed in the programs' measure sets. Public comments from American Academy of Hospice and Palliative Medicine (AAHMP), Center to Advance Palliative Care (CAPC), and National Coalition for Hospice and Palliative Care (NCHPC) supported the inclusion of this measure in the PQRS. CAPC and NCHPC suggested that the measure should be expanded beyond hospice patients. The measure is not currently included in the PQRS program, but it is included in the Hospice Quality Reporting Program.
- In February 2016, the Measures Application Partnership Post-Acute Care and Long-Term Care workgroup supported the continued development of a composite measure of Hospice Item Set measures, which includes NQF #1639.

***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**

Committee pre-evaluation comments

Criteria 4: Usability and Use

* This measure is not currently being publicly reported.
- This measure is included in the Hospice Quality Reporting Program (HQRP). Medicare-certified hospices are required to submit a Hospice Item Set (HIS)-Admission record and an HIS-Discharge record for each patient admission on or after July 1st, 2015. In FY2015, 3,992 hospice organizations provided measure data for 1,218,786 patient stays. The distribution of reporting hospice agencies by geographic area is provided.

* no public reporting at this time

Criterion 5: Related and Competing Measures

Related measures

- 0179: Improvement in dyspnea: Percentage of home health episodes of care during which the patient became less short of breath or dyspneic. *[facility-level outcome measure in home health setting]*
- 1638: Percentage of patients who screened positive for dyspnea who received treatment within 24 hours of screening. *[clinician-level & facility-level process measure in hospice and hospital setting*]

Harmonization

- Committee 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: Recommend making the measure specific to palliative or hospice care, not both in the same measure. Palliative care is not restricted to inpatient treatment.

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.

NQF #: 1639	**NQF Project:** Palliative Care and End-of-Life Care

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)***:**

Dyspnea screening and assessment are necessary to detect the presence of dyspnea (for which physical signs such as hypoxia and tachypnea do not clearly correlate), and to understand its severity and underlying etiology. Evidence-based treatment of dyspnea will vary with its severity and etiology, with treatment options differing for causes such as malignant pleural effusion, bulky tumor mass, congestive heart failure, anemia, COPD, among others. Additional guidelines from the American College of Physicians recommend dyspnea screening and assessment. Additional evidence includes numerous systematic reviews.

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)***:**

Dyspnea is a prevalent, distressing and functionally limiting symptom common to many serious illness conditions, and associated with risk of death.

a) Combining the results of several systematic reviews, moderate evidence supports pharmacologic treatments, including opioids for treatment of dyspnea in cancer and non-cancer diagnoses, and inhaled beta agonists for COPD.

b) Combining the results of several systematic reviews, moderate evidence supports non-pharmacologic treatments, including oxygen for hypoxic patients with cancer and non-cancer diagnoses, exercise interventions for COPD and CHF, thoracentesis for malignant pleural effusions, and nurse-led coping or relaxation interventions in cancer dyspnea.

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)*: 1. Systematic review of dyspnea treatment in palliative care (Lorenz, 2008); reported on 7 systematic reviews and 12 additional individual studies.

2. Systematic review of dyspnea management in cancer care, with evidence included for other diagnoses (Dy, 2008); reported on 25 studies of oxygen, 33 RCTs of beta agonists in COPD and 1 in "terminal illness," 18 studies of opioids, 9 of which were RCTs, and 1 meta-analysis of 36 RCTs of pleurodesis.

3. Systematic review for cancer dyspnea management (Ben-Aharon, 2008); reported on 18 randomized clinical trials.

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):* a) Multiple randomized trials of varied quality support the use of opioids for breathlessness; follow-up times are generally short, and evidence is stronger for COPD than for cancer patients.

b) Multiple randomized trials support the use of beta agonists for dyspnea in COPD

c) Several randomized trials support the use of oxygen, with mixed results and stratification showing beneficial effects for hypoxic but not for non-hypoxic patients.

d) Several small trials provide early evidence for the benefit of coping or relaxation interventions.

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

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

Benefits outweigh harms. Few harms of dyspnea screening or assessment are reported, few harms of treatment are reported.

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? Yes

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: Overall moderate quality (see Lorenz, 2008)

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

1c.12 If other, identify and describe the grading scale with definitions:

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

1c.14 Summary of Controversy/Contradictory Evidence: N/A

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

1. Luce JM, Luce JA. Management of dyspnea in patients with far-advanced lung disease. JAMA 2001; 285:1331-1337.

2. Ben-Aharon I, Gafter-Gvili A, Paul M et al. Interventions for alleviating cancer-related dyspnea: a systematic review. J Clin Oncol 2008; 26:2396-2404.

3. Roberts DK, Thorne SE, Pearson C. Cancer Nurs 1993; 16:310-320.

4. Currow DC, Ward AM, Abernethy AP. Advances in the pharmacologic management of breathlessness. Current Opin Supp Pall Care 2009; 3:103-106.

5. Bausewin C, Booth S, Gysels M et al. Non-pharmacologic interventions for breathlessness in advanced stages of malignant and nonmalignant diseases. Cochrane Database Syst Rev 2009. Apr 16; 2:CD005623.

6. Lorenz KA, Lynn J, Dy SM et al. Evidence for improving palliative care at the end of life: a systematic review. Ann Intern Med 2008; 148:147-159.

7. Dy SM, Lorenz KA, Naeim A et al. Evidence-based recommendations for cancer fatigue, anorexia, depression and dyspnea. J Clin Onc 2008; 26:3886-3895.

8. Lorenz KA, Lynn J et al. End-of-life care and outcomes. AHRQ Publication No. 05-E004-2, December 2004.

OTHER GUIDELINES:

9. Medical Services Commission. Palliative care for the patient with incurable cancer or advanced disease. Part 2: pain and symptom management. Victoria (BC): British Columbia Medical Services Commission; 2011 Sep 30. 44 p.

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

Health Care Guideline: Palliative Care for Adults

Guideline 4. Physical Aspects of Care: The physical aspects of the patient's serious illness should be an integral component of the palliative care plan. (p.21)

Guideline 2.1 Pain, other symptoms, and side effects are managed based upon the best available evidence, with attention to disease-specific pain and symptoms, which is skillfully and systematically applied.

• Regular, ongoing assessment of pain, non-pain symptoms (including but not limited to shortness of breath, nausea, fatigue and weakness, anorexia, insomnia, anxiety, depression, confusion, and constipation), treatment side effects, and functional capacities are documented through a systematic process. Validated instruments, where available, should be utilized. Symptom assessment in children and cognitively impaired patients should be performed by appropriately trained professionals with appropriate tools.

1c.17 Clinical Practice Guideline Citation: Palliative Care for Adults: McCusker M, Ceronsky L, Crone C, Epstein H, Greene B, Halvorson J, Kephart K, Mallen E, Nosan B, Rohr M, Rosenberg E, Ruff R, Schlecht K, Setterlund L. Institute for Clinical Systems Improvement. Palliative Care for Adults. Updated November 2013.p.21

1c.18 National Guideline Clearinghouse or other URL: Palliative Care for Adults: http://www.guideline.gov/content.aspx?id=47629

1c.19 **Grading of Strength of Guideline Recommendation.** Has the recommendation been graded? Yes

1c.20 **If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias:** Palliative Care for Adults: Institute for Clinical Systems Improvement (ICSI)

1c.21 **System Used for Grading the Strength of Guideline Recommendation:** Palliative Care for Adults: GRADE Metholodogy

1c.22 **If other, identify and describe the grading scale with definitions:**

1c.23 **Grade Assigned to the Recommendation:** Palliative Care for Adults: Low Quality Evidence, Strong Recommendation

1c.24 **Rationale for Using this Guideline Over Others:** We chose to use the Palliative Care for Adults guideline from the Institute for Clinical Systems Improvement (ICSI) because it is evidenced based, uses GRADE methodology, and is included in the National Guideline Clearinghouse.

In addition, the National Consensus Project for Quality Palliative Care was the first United States national guidelines development project for palliative care quality, inclusive of hospice care. This set of guidelines, along with 38 preferred practices, has been rigorously reviewed and endorsed by the National Quality Forum. Although specific investigative groups and specialty organizations have published other guidelines in pain management or hospice or palliative care practice for specific settings or populations, none have been as comprehensive or comprehensively debated, peer reviewed, or NQF endorsed.

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: Moderate1c.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 #: 1639
De.2. **Measure Title:** Hospice and Palliative Care -- Dyspnea Screening
Co.1.1. **Measure Steward:** University of North Carolina-Chapel Hill
De.3. **Brief Description of Measure:** Percentage of hospice or palliative care patients who were screened for dyspnea during the hospice admission evaluation / palliative care initial encounter.
1b.1. **Developer Rationale:** Dyspnea is prevalent and undertreated for many populations of seriously ill patients, including those patients nearing the end of life. Screening for dyspnea is necessary to determine its presence and severity, and forms the basis for treatment decision-making. Unlike pain, structured clinical assessment of the symptom is less well-defined, yet similar to pain, effective treatment is available to alleviate symptom distress.

S.4. **Numerator Statement:** Patients who are screened for the presence or absence of dyspnea and its severity during the hospice admission evaluation / initial encounter for palliative care.
S.7. **Denominator Statement:** Patients enrolled in hospice OR patients receiving hospital-based palliative care for 1 or more days.
S.10. **Denominator Exclusions:** Patients with length of stay < 1 day in palliative care.

De.1. **Measure Type:** Process
S.23. **Data Source:** Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record
S.26. **Level of Analysis:** Clinician : Group/Practice, 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:
1640:Hospice and Palliative Care -- Dyspnea Screening and Dyspnea Treatment

De.4. **IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** Part of the PEACE Measures Set
Paired with Hospice and Palliative Care – Dyspnea Treatment (percentage of patients who screened positive for dyspnea who received treatment within 24 hours of screening)

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
1639_Evidence_MSF5.0_Data.doc,1639_Evidence_3.17.16.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)*
Dyspnea is prevalent and undertreated for many populations of seriously ill patients, including those patients nearing the end of life. Screening for dyspnea is necessary to determine its presence and severity, and forms the basis for treatment decision-making. Unlike pain, structured clinical assessment of the symptom is less well-defined, yet similar to pain, effective treatment is available to alleviate symptom distress.

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.*
Hospice: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1639 - Hospice and Palliative Care Dyspnea Screening.

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 1,218,786 patient stays.

The mean score for this QM was 97.3% with a range from 0% to 100%, the median was 99.2%, the interquartile range was 2.9, and the standard deviation was 618. For this measure, 37.1% of hospices had perfect scores and 6.7% of hospices scored below 90%.

Scores by decile:
10th percentile 92.6%
25th percentile 97.1%
Median 99.2%
75th percentile 100%
90th percentile 100%

Palliative Care:

This submission to the Palliative and End-of-Life Care project updates hospice setting data for NQF #s 1634, 1637, 1638, 1639, 1641, 1647. We are currently in the process of updating palliative care data by collecting and analyzing data in multiple non-hospice settings but final analyses are not available for this submission cycle.(1) Data comes from two sources -- a multi-site study of quality of care in palliative care (R18HS022763 Implementing Best Practice in Palliative Care, PI Johnson) and (CMS Health Care Innovation Award: Increasing patient and system value with community based palliative care, PI Bull / Four Seasons Compassion for Life). We anticipate these data will become available for NQF review next year. This data will allow further updates to the evidence base for non-hospice palliative care beyond what is currently submitted.

(1) Kamal AH, Bull J, Ritchie CS, Kutner JS, Hanson LC, Friedman F, Taylor DH Jr; AAHPM Research Committee Writing Group. Adherence to Measuring What Matters measures using point-of-care data collection across settings. J Pain Symptom Manage 2016; 51:497-503.

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.**
Prevalence of dyspnea in advanced cancer ranges from 50-70%. Among COPD patients with advanced illness enrolled in the SUPPORT Study, dyspnea which was moderate to severe at least half of the time was present for at least 65% of patients throughout the 6 months preceding death.

Effective treatment for dyspnea is available, but not consistently administered. Evidence-based treatments include pharmacologic interventions such as opioids and inhaled bronchodilators, and non-pharmacologic interventions including oxygen for hypoxic patients, pulmonary rehabilitation and exercise in COPD, and drainage of pleural effusion.

1. Luce JM, Luce JA. Management of dyspnea in patients with far-advanced lung disease. JAMA 2001; 285:1331-1337.

2. Ben-Aharon I, Gafter-Gvili A, Paul M et al. Interventions for alleviating cancer-related dyspnea: a systematic review. J Clin Oncol 2008; 26:2396-2404.
3. Roberts DK, Thorne SE, Pearson C. Cancer Nurs 1993; 16:310-320.
4. Currow DC, Ward AM, Abernethy AP. Advances in the pharmacologic management of breathlessness. Current Opin Supp Pall Care 2009; 3:103-106.
5. Bausewin C, Booth S, Gysels M et al. Non-pharmacologic interventions for breathlessness in advanced stages of malignant and nonmalignant diseases. Cochrane Database Syst Rev 2009. Apr 16; 2:CD005623.
6. Lorenz KA, Lynn J, Dy SM et al. Evidence for improving palliative care at the end of life: a systematic review. Ann Intern Med 2008; 148:147-159.
7. Dy SM, Lorenz KA, Naeim A et al. Evidence-based recommendations for cancer fatigue, anorexia, depression and dyspnea. J Clin Onc 2008; 26:3886-3895.

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.*
Hospice:

Racial disparity analysis. We conducted racial and ethnic disparity analyses at both the patient-stay and hospice levels. At the patient-stay level, we compared the percentage of patients who received dyspnea screening among different racial and ethnic groups. Patients were grouped by their racial and ethnic identification as follows: white non-Hispanic, black non-Hispanic, other non-Hispanic, or Hispanic. Other non-Hispanic includes patients who identify as American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, or who identify as more than one race. A Chi-square test was performed to determine if there were any statistically significant differences in dyspnea screening between groups. The lowest rate of dyspnea screening was found for patients with racial and ethnic group missing (94.8%), and the highest rate was among patients identifying as White non-Hispanic (97.6%). Differences in the rate of dyspnea screening by racial identification were found to be statistically significant ($p<0.001$).

Analyses at the hospice level examined the differences in this measure across two groups: hospices with proportions of non-white patients that are greater than or equal to the national median proportion (11.9%), and hospices with fewer non-white patients than the national median. For this analysis, white non-Hispanic patients were included in the white group, and all other racial and ethnic identifications, as described above, were grouped as non-white. We ran a Wilcoxon-Mann-Whitney test for statistical dependence between group and QM score. The results showed that the QM score was significantly different between the two groups of hospices (97.6% compared to 96.9%, $p=0.008$). Although statistically significant results were found at both the patient and hospice level, actual differences in screening rates do not seem to be clinically substantial.

Gender disparity analysis. We performed both the patient stay- and hospice-level analyses on gender disparity. At the patient-stay level, we compared the percentage of patients who received dyspnea screening between female and male patients. A Chi-square test was performed to determine if there were any statistically significant differences in dyspnea screening between groups. We found a slightly lower rate of dyspnea screening for female patients (97.26%) than for male patients (97.34%). Differences in the rate of dyspnea screening by gender were statistically significant ($p=0.016$). At the hospice level, we examined the differences in this measure across hospices with proportions of female patients that are greater than or equal to the national median proportion (55.2%). We ran a Wilcoxon-Mann-Whitney test for statistical dependence between group and QM score. The results showed that the QM score was not statistically significantly different between the two groups of hospices split by median proportion of female patients (97.3% compared to 97.2%, $p=0.96$). Although statistically significant results were found at the patient level, actual differences in screening rates do not seem to be clinically substantial.

Socioeconomic disparity analysis. We performed socioeconomic disparity analyses at both the patient-stay and hospice levels using the same methods described in previous disparity analyses. Medicaid status was used as a proxy measure of low socioeconomic status. A statistically significant ($p<0.001$) lower rate of dyspnea screening was found for non-Medicaid patients (97.2%) than for Medicaid patients (97.4%). At the hospice level, the results showed that the QM score was significantly different between the two groups of hospices split by median proportion of Medicaid patients, 21.5% (96.9% compared to 97.7%, $p<0.001$). The statistically significant results at the patient-stay level may indicate that quality of hospice care, measured by dyspnea screening, for non-Medicaid patients is lower than for Medicaid patients. The significant findings at the hospice level indicate that hospices with a smaller proportion of Medicaid patients are less likely to provide dyspnea screening to patients at admission. A potential caveat of this analysis is that the data used for this analysis may not be a reliable indicator of a patient's Medicaid eligibility status. This data is missing for nearly a quarter of patients. In addition, some of the Medicaid numbers submitted do not appear to be legitimate Medicaid numbers. However, the data used in this analysis is the only national patient-level assessment data that is currently

available in hospices. We will update this analysis as more-accurate data sources are available and accessible.

Rural-urban disparity analysis. We compared the QM score between rural and urban hospices using a Wilcoxon-Mann-Whitney test. The results showed that the QM score was significantly different between rural and urban hospices (97.6% compared to 97.1%, $p<0.001$). Although statistically significant results were found, actual differences in screening rates do not seem to be clinically substantial.

Palliative Care: Disparities data not available

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.**
Limited research has explored the nature of health disparities in the experience of dyspnea or in dyspnea management. One observational study of dyspnea in cancer patients provides evidence that dyspnea and other symptoms, in addition to minority race / ethnicity, independently predict worsened survival.

Tammemagi CM, Neslund-Dudas C, Simoff M, Kvale P. Lung carcinoma symptoms – an independent predictor of survival and an important mediator of African-American health disparity in survival. Cancer. 2004 Oct 1;101(7):1655-63.

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**
Affects large numbers, Patient/societal consequences of poor quality, Severity of illness
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.**
The Hospice and Palliative Care Dyspnea Screening measure addresses dyspnea for patients with high severity of illness and risk of death, including seriously and incurably ill patients enrolled in hospice or hospital-based palliative care. Research on care of patients nearing the end of life shows they experience high rates of physical, emotional, and spiritual causes of distress.(1,2) The National Priorities Partnership has identified palliative and end-of-life care as one of its national priorities. A goal of this priority is to ensure that all patients with life-limiting illness have access to effective treatment for symptoms such as pain and shortness of breath.(3) In 2014, 1.66 million people with life-limiting illness received hospice care.(4) In 2013, 67% of US hospitals with 50 or more beds had some form of palliative care service, up from 58.5% in 2008, and national trends show a steady expansion of these services. (5)

Dyspnea is a common symptom in serious illness, more common than pain for patients with chronic obstructive lung disease, lung cancer, cystic fibrosis, and restrictive lung diseases such as pulmonary fibrosis.(6) Unlike pain, dyspnea severity is associated with the risk of death.(7) Between 50-70% of patients with advanced lung cancer experience dyspnea near the end of life. As detailed in a recent systematic review, opioids, oxygen and non-pharmacologic nursing interventions demonstrate efficacy in randomized controlled trials of treatment for dyspnea in cancer and in other serious illness.(8,9) Unfortunately, dyspnea is often persistent and under-treated in advanced cancer and other end-stage diseases.(10)

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
1. The Writing Group for the SUPPORT Investigators. A controlled trial to improve care for seriously ill hospitalized patients. The study to understand prognosis and preferences for outcomes and risks of treatments (SUPPORT). JAMA. 1995;274:1591-1598.
2. Gade G, Venohr I, Conner D, et al. Impact of an inpatient palliative care team: a randomized control trial. J Palliat Med. 2008;11(2):180–190.
3. http://www.nationalprioritiespartnership.org/PriorityDetails.aspx?id=608
4. NHPCO Facts and figures: hospice care in America 2015 edition
http://www.nhpco.org/hospice-statistics-research-press-room/facts-hospice-and-palliative-care
5. Dumanovsky T, Augustin R, Rogers M, Lettang K, Morrison RS. The growth of palliative care in U.S. hospitals: a status report. J Pall Med. 2016; 19(1): 8-15

6. Luce JM, Luce JA. Management of dyspnea in patients with far-advanced lung disease. JAMA 2001; 285:1331-1337.
7. Olajidae O, Hanson LC, Usher BM et al. Validation of the Palliative Performance Score in the acute tertiary hospital setting. J Palliat Med 2007; 10:111-117
8. Ben-Aharon I, Gafter-Gvili A, Paul M et al. Interventions for alleviating cancer-related dyspnea: a systematic review. J Clin Oncol 2008; 26:2396-2404.
9. Lorenz KA, Lynn J, Dy SM et al. Evidence for improving palliative care at the end of life: a systematic review. Ann Intern Med 2008; 148:147-159.
10. Roberts DK, Thorne SE, Pearson C. Cancer Nurs 1993; 16:310-320.

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):*

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.)*
PEACE Hospice and Palliative Care Quality Measures: http://www.med.unc.edu/pcare/resources/PEACE-Quality-Measures

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.
We would like to propose the removal of the less than 7 day length of stay (LOS) denominator exclusion for hospice patients.

Background: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1639 - Hospice and Palliative Care Dyspnea Screening.

Six of the seven QMs exclude patient stays that are less than 7 days from the measure denominator. When the length of stay (LOS) is too short, hospices may not have enough time to complete all the clinically recommended care processes. Thus, at the time the measures were developed, technical experts recommended that short patient stays be excluded from those measure denominators

for assessing quality of care in hospices. However, no national data regarding the implications of the LOS exclusion was available to the Technical Expert Panel (TEP) at that time. Under contract to the Centers for Medicare and Medicaid Services (CMS), the Research Triangle Institute (RTI) performed descriptive analyses on 5 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 1,218,786 patient stays.

Rationale for inclusion of all hospice patients regardless of LOS: At the patient level, approximately 40% of patients were excluded based on the LOS < 7 days exclusion, thus omitting this essential quality of care standard for many seriously ill patients. Further, excluding short-stay patients omits measures of quality of care for a subset of the sickest patients entering hospice care, and those who arguably may be in greatest need of rapid symptom assessment and treatment. In addition, the original rationale for this exclusion -- allowing time for hospice providers to complete this care process -- does not appear to be necessary. Our analyses show that a large portion of dyspnea screenings were performed on day 1 of admission to hospice, demonstrating a normative standard of care includes prompt attention to symptom distress.

At the hospice level, applying different LOS exclusion criteria had an impact on denominator size. Smaller denominator size at the hospice level may generate less stable and reliable QM scores. Under the LOS < 7 days exclusion, the median number of qualifying stays in the denominator was 94 stays per hospice vs. a median number of 136 stays per hospice with no LOS exclusion. Applying the LOS exclusion resulted in some hospices with no qualifying stays, which excluded 14 hospices from QM score calculations altogether.

Applying or removing the LOS exclusion generally had little impact on the distribution of hospices' QM scores. Under the 7-day LOS exclusion, the mean score was 97.2%, the median score was 99.4%, and the score for hospices in the 10th percentile distribution was 92.3%. With no LOS exclusions, the mean score was 97.3%, the median score was 99.2%, and the score for hospices in the 10th percentile distribution was 92.6%. The impact of the different LOS criteria on the distribution of QM scores was consistent across quarters.

In summary, these new analyses demonstrate that removing the LOS exclusion in QM calculations, thus including all hospice patients in the denominator, is feasible and appropriate for quality of care at the patient level, and enhance completeness and statistical stability of QM reporting at the hospice level. Further, though the mean and median scores on this QM are high, hospice organizations in the 10th percentile have meaningful opportunity to improve in their timely screening of dyspnea. Since this quality measure is paired, successful screening is linked to subsequent measurement of dyspnea treatment (NQF #1639).

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 who are screened for the presence or absence of dyspnea and its severity during the hospice admission evaluation / initial encounter for palliative care.

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.)*
Hospice admission evaluation / initial clinical encounter for palliative care

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.
Patients who are screened for the presence or absence of dyspnea during the admission evaluation for hospice / initial encounter for hospital-based palliative care, and asked to rate its severity. Screening may be completed using verbal, numeric, visual analog, or rating scales designed for use with non-verbal patients.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Patients enrolled in hospice OR patients receiving hospital-based palliative care for 1 or more days.

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

Populations at Risk, 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)*
The Dyspnea Screening quality measure is intended for patients with serious illness who are enrolled in hospice care OR receive specialty palliative care in an acute hospital setting. Conditions may include, but are not limited to: cancer, heart disease, pulmonary disease, dementia and other progressive neurodegenerative diseases, stroke, HIV/AIDS, and advanced renal or hepatic failure.

[NOTE: This quality measure should be paired with the Dyspnea Treatment quality measure (NQF #1639) to ensure that all patients who report dyspnea are clinically considered for treatment.]

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Patients with length of stay < 1 day in palliative care.

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)*
Calculation of length of stay; discharge date is identical to date of initial encounter.

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)*
N/A

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*)

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.)*
Screened for dyspnea:

a.Step 1- Identify all patients with serious, life-limiting illness who are enrolled in hospice care or who receive specialty palliative care in an acute hospital setting

b.Step 2- Identify admission / initial encounter dates; exclude palliative care patients if length of stay is less than one day.

c.Step 3- Identify patients who were screened for dyspnea during the admission evaluation (hospice) OR during the initial encounter (palliative care)

Quality measure = Numerator: Patients screened for dyspnea in Step 3 / Denominator: Patients in Step 1 – Patients excluded in Step 2

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.
Hospice: The hospice analysis was not based on a sample. It was conducted on the entire hospice population that had admission and discharge records in the specified period of analysis.

Palliative care: consecutive sample of equal numbers of admissions + decedents beginning with a randomly selected date.

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.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Hospice: For the dyspnea screening measure, there are two items on the HIS that can include missing data – J2030B and J2030C. If missing, these items are coded as dashes and analyzed as though the process did not happen.

In order to assess how these missing data impact the validity of the dyspnea screening measure, we conducted the patient stay- and hospice-level analyses. For the patient stay-level analysis, we calculated the number and percentage of eligible patient stays for which the HIS-Admission records included a dash for each of these items. The overall rate of missing data ranged from 0.02 percent to 0.12 percent between October 2014 and September 2015. For the hospice-level analysis we calculated each hospice's percent of eligible admissions that included missing data for items J2030B and J2030C. Over 90% of hospices did not have any admissions with missing data for these items.

Overall, we found that only a very small number of admission records for the eligible stays did not include data, i.e. coded as dash, on dyspnea screening items J2030B and J2030C. And a vast majority of hospices did not have any missing data for these items. These results indicate that missing data for these items should not have a negative impact on the validity of the QM.

Palliative care: N/A

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record

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.
Hospice: Hospice analysis uses the Hospice Item Set (HIS) as the data source to calculate the quality measure.

Palliative Care: Structured medical record abstraction tool, with separate collection of denominator and numerator data

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 in attached appendix at A.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Group/Practice, Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospice, 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
1639_MeasureTesting_MSF5.0_Data.doc,1639_MeasureTestingAttachment_2.26.16.docx

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 1639	**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)***:**

Hospice: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1639 - Hospice and Palliative Care Dyspnea Screening.

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.

Palliative Care: Two research nurse abstractors independently recorded quality measures data on a random subset of 20 seriously ill patients. Abstractors used the pre-defined operational definitions and a structured chart abstraction tool to record numerator and denominator data separately. Patients were a sub-sample of 460 seriously ill patients without specialty palliative care admitted to an acute hospital for at least 1 day to four inpatient services from February 2008 to November 2009. Records eligible for sampling included all seriously ill adult patients admitted to medical and surgical intensive care, medically complex patients aged 65 and older admitted to an Acute Care of the Elderly Unit, and medical oncology patients with Stage IV carcinoma.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)***:**

Hospice:

Split-half reliability. Split-half reliability assesses the internal consistency of a QM in different samples by randomly dividing the patient stays within each hospice into two halves, and calculating correlation between the QM scores on the basis of the two randomly divided halves. In this analysis, we conducted a split-half reliability analysis on all facilities with 20 or more patient stays counted in the measure denominator, and used the Interclass Correlation (ICC) coefficients to measure the internal reliability. In general, the ICC coefficient varies between 0 and 1, where an ICC of 0 indicates no reliability and an ICC of 1 indicates perfect reliability.

Signal-to-noise analysis. If a measure is reliable, then true differences in provider performance should explain a substantial proportion of the variance in QM scores. We conducted an analysis of variance (ANOVA) to determine what proportion of total variance in the measure is attributable to differences among providers. A higher proportion indicates better reliability.

Stability analysis. Stability analysis describes the extent to which providers' performance assessed by a QM changes across time. We analyzed the change in facility scores between four consecutive quarters (Q4 2014 and Q1 2015, Q1 and Q2 2015, and Q2 and Q3 2015). The changes in facility scores are reported in standard deviations.

Palliative Care: Inter-rater reliability between the two abstractors was assessed using kappa statistics. Two research nurse abstractors independently recorded quality measures data on a random subset of 20 seriously ill patients. Abstractors used the pre-defined operational definitions and a structured chart abstraction tool to record numerator and denominator data separately.

2a2.3 **Testing Results** *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**

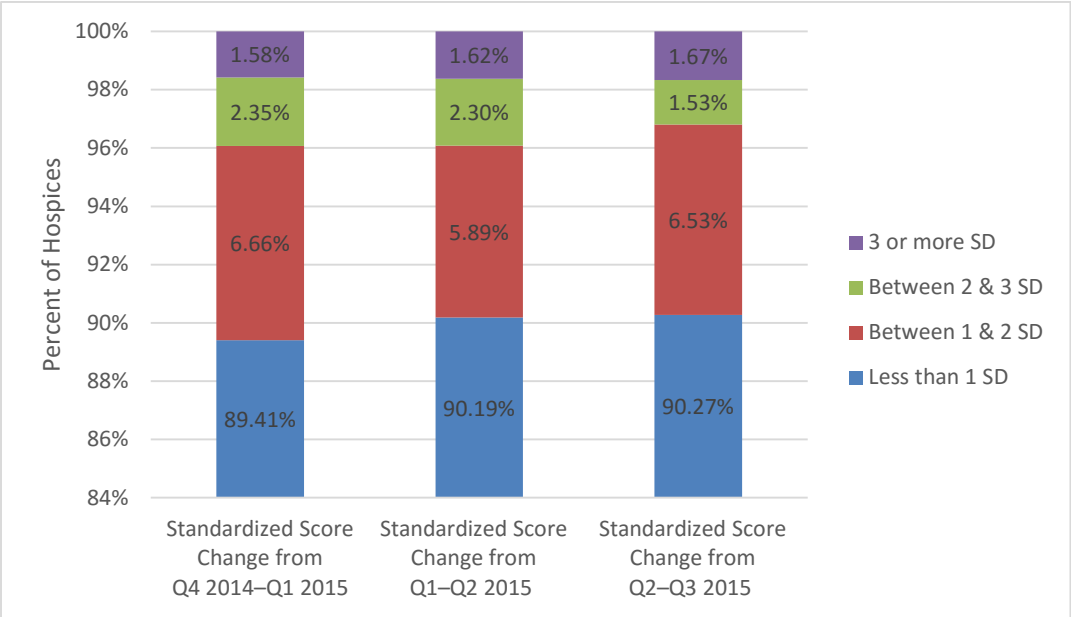
Hospice:

Split-half reliability: In general, the ICC coefficient varies between 0 and 1, where an ICC of 0 indicates no reliability and an ICC of 1 indicates perfect reliability. The ICC coefficient for this measure is 0.83, indicating high internal reliability.

Signal-to-noise analysis: The analysis results found the signal-to-noise ratio to be 0.98, indicating that about 98% of the variance in this measure is because of differences among facilities. This proportion indicates strong reliability for this measure.

Stability analysis: The results of this analysis indicated that facility scores were very stable. Approximately 90% of facilities had a change in QM score of less than one standard deviation, indicating high stability of the QM. The number of facilities with a change in QM of less than one standard deviation increased across quarters from 89.4% to 90.3%, suggesting improved reliability across time. Less than one-tenth of facilities had a change in QM between one and two standard deviations. These results indicate a measure that is generally quite stable. Figure 1 illustrates the change in facility scores between the four consecutive quarters.

Figure 1
Standardized Score Change in QM Score from Q4 2014 to Q3 2015



Palliative Care: Kappa scores range from 0 to 1 with higher scores indicating better agreement. The nurse abstractors achieved excellent inter-rater reliability for this measure: Kappa=0.91. Landis and Koch describe kappa values that range from 0.81 – 0.99 as almost perfect and Fleiss describes kappas over 0.75 as excellent.

Landis, J.R.; Koch, G.G. (1977). "The measurement of observer agreement for categorical data". Biometrics 33 (1): 159–174

Fleiss, J.L. (1981). Statistical methods for rates and proportions (2nd ed.). New York: John Wiley

2b. VALIDITY. Validity, Testing, including all Threats to Validity: 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:**

The measure focus is Dyspnea Screening, designed to pair with Dyspnea Treatment to ensure quality care processes for dyspnea. The target populations are hospice patients, and seriously ill hospitalized patients with diverse underlying diagnoses who are at high risk for palliative care clinical needs, including dyspnea.

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)***:**

Hospice: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1639 - Hospice and Palliative Care Dyspnea Screening.

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.

Palliative Care: The total patient sample size was 562. Chart abstractions were completed for 102 consecutive seriously ill patients with specialty palliative care consultation, and a random sample of 460 seriously ill patients without specialty palliative care admitted to an acute care hospital for at least 1 day to four inpatient services with high proportions of seriously ill patients from February 2008 to November 2009. Records eligible for sampling included all patients admitted to medical and surgical intensive care, medically complex patients aged 65 and older admitted to a Geriatric Evaluation Unit, and medical oncology patients with Stage IV carcinoma. Because palliative care domains become even more relevant closer to death, patients dying in hospital were over-sampled to ensure a final ratio of 1 decedent to 1 live discharge. Consistent with oversampling of decedent records, 55% of these patients died in hospital.

The age of the patients ranged from 16 to 99 years, with the mean age 61. Patients were predominantly Caucasian (65%), with smaller subgroups who were African American (24%) and Hispanic / Latino (4%) The most common life-limiting diagnoses were infections (37%), cancer (34%), pulmonary (29%), and neurologic diseases (21%).

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

Hospice:

Correlations. Providers should perform similarly on QMs that reflect the quality of similar care processes. Thus, a common strategy used to evaluate validity of a QM is to examine the relative performance of the provider on QMs that are conceptually related, either due to some core competency manifested across a stay or for processes that occur in close proximity. We conducted nonparametric Spearman rank correlation analysis among all seven HQRP QMs, which address care processes around hospice admission that are clinically recommended or required in the hospice Conditions of Participation. We used nonparametric methods since the data are heavily skewed (most providers perform well, while a small proportion have lower scores). Since all of the seven current QMs measure desirable care assessment processes upon hospice admission, higher positive Spearman correlation coefficients among these QMs will demonstrate higher validity of the QM set. Statistically significant correlations, indicated by p-values of the Spearman correlation coefficients less than 0.05, also support the validity of the QM set.

Missing data. Missing data represent a potential threat to the validity of an HIS item, which in turn may harm the validity of the QM that relies on the item. For the dyspnea screening measure, there are two items on the HIS that can include missing data, which are coded as dashes – J2030B and J2030C. In order to assess how these missing data impact the validity of the dyspnea screening measure, we conducted the patient stay- and hospice-level analyses.

Palliative Care sample:

Face validity of PEACE quality measures for hospital-based specialty palliative care was addressed using stakeholder review and feedback. Investigators prepared data reports in a summary format with detailed operational definitions, and led a 1-hour discussion with nursing and physician leaders from each service group – MICU, SICU, Acute Care for the Elderly (Geriatrics), Oncology, and Palliative Care. The discussion included feedback of quality measure data, response to questions and critiques, and eliciting stakeholder feedback about the validity and actionability of this data for the care of their patients. Stakeholders were specifically asked to comment on the accuracy of the data as a reflection of current care practices, and their highest priority area for future quality improvement.

Construct validity was tested by comparing the PEACE quality measures for patients seen by specialty interdisciplinary palliative care consultants to those not receiving specialty palliative care services.

CITATION: Schenck AP, Rokoske FS, Durham DD, et al. The PEACE Project: identification of quality measures for hospice and palliative care. J Palliat Med 2010; 13:1451-1459.

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)*:

Hospice:

Correlations: Table 1 presents the Spearman correlation coefficients. The p-values for all the Spearman correlation coefficients are significant ($p < 0.001$). The significant positive correlations between every pair of QMs indicate that high-performing hospices in one area also provided high-quality care in other areas at hospice admission. Overall, the correlations between the QMs are low to moderate. The clustering of QM scores, skewed distributions, and low variability across all seven QMs may affect the level of correlations between QMs.

Table 1
Correlation of Hospice QMs, Percentile Ranking

Quality measure	NQF #1647 (modified) Beliefs/Values Addressed	NQF #1634 Pain Screening	NQF #1637 Pain Assessment	NQF #1639 Dyspnea Screening	NQF #1638 Dyspnea Treatment	NQF #1617 Bowel Regimen
NQF #1641 Treatment Preferences	0.64***	0.37***	0.21***	0.44***	0.28***	0.25***
NQF #1647 (modified) Beliefs/Values Addressed		0.42***	0.17***	0.43***	0.31***	0.25***
NQF #1634 Pain Screening			0.17***	0.41***	0.28***	0.15***
NQF #1637 Pain Assessment				0.17***	0.08***	0.20***
NQF #1639 Dyspnea Screening					0.30***	0.25***
NQF #1638 Dyspnea Treatment						0.34***

NOTE: The correlation is on the basis of each hospice's percentile ranking on the QMs.
*** indicates significant correlation at $p < 0.001$

Overall, NQF #1639, Dyspnea Screening, has significant positive correlations with the other QMs, indicating hospices providing higher-quality care in this area also performed better in other areas at hospice admission.

Missing data. For the patient stay-level analysis, we calculated the number and percentage of eligible patient stays for which the HIS-Admission records included a dash for each of these three items. The overall rate of missing data ranged from 0.02 percent to 0.12 percent between October 2014 and September 2015. For the hospice-level analysis we calculated each hospice's percent of eligible admissions that included missing data for items J2030B and J2030C. Over 90% of hospices did not have any admissions with missing data for these items.

Overall, we found that only a very small number of admission records for the eligible stays did not include data, i.e. coded as dash, on dyspnea screening items J2030B and J2030C. And a vast majority of hospices did not have any missing data

for these items. These results indicate that missing data for these items should not have a negative impact on the validity of the QM

Palliative Care sample:

Face Validity: Stakeholder discussions provided broad endorsement of face validity, with some considerations for specific patient populations. Medical oncologists endorsed the face validity of these quality measures, but favored quality measures endorsed by oncology professional organizations.

Construct Validity: Screening for dyspnea was nearly universal for all seriously ill patients, but was more consistently done by specialty palliative care providers (100% vs 95%, p=0.016). Patients with dyspnea were likely to receive some form of treatment within 24 hours, with or without the addition of specialty palliative care (96% vs 93%, p=NS).

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)***:**

Hospice: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1639 - Hospice and Palliative Care Dyspnea Screening.

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.

Palliative Care: N/A

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

Hospice: Under contract to the Centers for Medicare and Medicaid Services (CMS), the Research Triangle Institute (RTI) performed descriptive analyses on 5 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.

Palliative Care: N/A

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

Hospice: At the patient level, approximately 40% of patients were excluded based on the LOS < 7 days exclusion, thus omitting this essential quality of care standard for many seriously ill patients. Further, excluding short-stay patients omits measures of quality of care for a subset of the sickest patients entering hospice care, and those who arguably may be in greatest need of rapid symptom assessment and treatment. In addition, the original rationale for this exclusion -- allowing time for hospice providers to complete this care process -- does not appear to be necessary. Our analyses show that a large portion of dyspnea screenings were performed on day 1 of admission to hospice, demonstrating a normative standard

of care includes prompt attention to symptom distress.

At the hospice level, applying different LOS exclusion criteria had an impact on denominator size. Smaller denominator size at the hospice level may generate less stable and reliable QM scores. Under the LOS < 7 days exclusion, the median number of qualifying stays in the denominator was 94 stays per hospice vs. a median number of 136 stays per hospice with no LOS exclusion. Applying the LOS exclusion resulted in some hospices with no qualifying stays, which excluded 14 hospices from QM score calculations altogether.

Applying or removing the LOS exclusion generally had little impact on the distribution of hospices' QM scores. Under the 7-day LOS exclusion, the mean score was 97.2%, the median score was 99.4%, and the score for hospices in the 10th percentile distribution was 92.3%. With no LOS exclusions, the mean score was 97.3%, the median score was 99.2%, and the score for hospices in the 10th percentile distribution was 92.6%. The impact of the different LOS criteria on the distribution of QM scores was consistent across quarters.

In summary, these new analyses demonstrate that removing the LOS exclusion in QM calculations, thus including all hospice patients in the denominator, is feasible and appropriate for quality of care at the patient level, and enhance completeness and statistical stability of QM reporting at the hospice level.

Palliative Care: N/A

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)***:**

N/A

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

N/A

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)***:**

N/A

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

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)***:**

Hospice: CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281).

Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1639 - Hospice and Palliative Care Dyspnea Screening.

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.

Palliative Care: The total patient sample size was 562. Chart abstractions were completed for 102 consecutive seriously ill patients with specialty palliative care consultation, and a random sample of 460 seriously ill patients without specialty palliative care admitted to an acute care hospital for at least 1 day to four inpatient services with high proportions of seriously ill patients from February 2008 to November 2009. Records eligible for sampling included all patients admitted to medical and surgical intensive care, medically complex patients aged 65 and older admitted to a Geriatric Evaluation Unit, and medical oncology patients with Stage IV carcinoma. Because palliative care domains become even more relevant closer to death, patients dying in hospital were over-sampled to ensure a final ratio of 1 decedent to 1 live discharge. Consistent with oversampling of decedent records, 55% of these patients died in hospital.

The age of the patients ranged from 16 to 99 years, with the mean age 61. Patients were predominantly Caucasian (65%), with smaller subgroups who were African American (24%) and Hispanic / Latino (4%) The most common life-limiting diagnoses were infections (37%), cancer (34%), pulmonary (29%), and neurologic diseases (21%).

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

Hospice: Confidence interval analysis. We examined proportions of hospices with the QM scores that are significantly different from the national hospice-level mean. If a high proportion of hospices have a measure score significantly different from the mean, the QM can identify facilities with different levels of performance. For this analysis, statistical significance was determined using 95 percent confidence intervals: a hospice's QM score was significantly different from the national mean if the national mean was not included within the hospice's 95 percent confidence interval. High-performing facilities should have scores that are significantly below average, and low-performing facilities should be significantly above average.

Palliative Care: Construct validity was tested by comparing the PEACE quality measures for patients seen by specialty interdisciplinary palliative care consultants to those not receiving specialty palliative care services. Percentage of patients with and without specialty palliative care for whom the quality measure was met was compared for difference using chi-square statistics.

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)*:

Hospice:

The mean score for this QM was 97.3% with a range from 0% to 100%, the median was 99.2%, the interquartile range was 2.9, and the standard deviation was 6.1. For this QM, 37.1% of hospices had perfect scores and 6.7% of hospices scored below 90%.

Scores by decile:

10th percentile 92.6%

25th percentile 97.1%

Median 99.2%

75th percentile 100%

90th percentile 100%

Across all hospices, 36.7% had a QM score that is significantly different than the national mean. Hospices were more likely to report scores above the national mean than below the national mean (25.2% vs 11.5%, respectively, overall). The QM is able to identify those hospices that are performing well (higher than the national mean) and those that are performing less well (lower than the national mean).

Palliative Care:

Seriously ill patients with palliative care sample: 100% met quality measure

Seriously ill patients without palliative care: 95% (p=0.016)

Dyspnea screening for patients with serious illness in hospitals is not consistently achieved; access to specialty palliative care providers demonstrates strong evidence this quality measure can be improved in practice.

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)***:**

N/A

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

N/A

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)***:**

N/A

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)***:**

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

Disparities in dyspnea screening and treatment have not been well characterized in the hospice and palliative care population. Future research with larger sample sizes can be used to test for differential performance by race / ethnicity and by gender.

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)*

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

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.

Missing data. Missing data represent a potential threat to the validity of an HIS item, which in turn may harm the validity of the QM that relies on the item. For the dyspnea screening measure, there are two items on the HIS that can include missing data, which are coded as dashes – J2030B and J2030C.

In order to assess how these missing data impact the validity of the dyspnea screening measure, we conducted the patient stay- and hospice-level analyses. For the patient stay-level analysis, we calculated the number and percentage of eligible patient stays for which the HIS-Admission records included a dash for each of these items. The overall rate of missing data ranged from 0.02 percent to 0.12 percent between October 2014 and September 2015. For the hospice-level analysis we calculated each hospice's percent of eligible admissions that included missing data for items J2030B and J2030C. Over 90% of hospices did not have any admissions with missing data for these items.

Overall, we found that only a very small number of admission records for the eligible stays did not include data, i.e. coded as dash, on dyspnea screening items J2030B and J2030C. And a vast majority of hospices did not have any missing data for these items. These results indicate that missing data for these items should not have a negative impact on the validity of the QM. There was no missing data for the elements needed to calculate this subset of dyspnea measures for the Palliative Care samples.

Record abstraction does not require collection of unique patient identifiers and thus protects confidentiality. Timing of data collection can be concurrent with admission / initial encounter care, or can be retrospective based on medical record sampling.

Costs have not been formally estimated; medical record abstraction or electronic capture of the elements of a dyspnea screen and treatment will have more modest cost compared to survey data.

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)*.
N/A

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)
Public Reporting	Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Hospice Item Set-Hospice Quality Reporting Program https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Hospice-Quality-Reporting/Hospice-Item-Set-HIS.html Quality Improvement (Internal to the specific organization) Voluntary use by UNC Palliative Care Program http://www.med.unc.edu/pcare/resources/PEACE-Quality-Measures

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

Name of program and sponsor: Hospice: HIS/CMS; Palliative Care: Voluntary use by PC organizations

Purpose- Hospice: Quality reporting for hospice; Palliative Care: Internal quality improvement

Geographic area and number and percentage of accountable entities and patients included – United States and all accountable entities and 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**

a. Due to insufficient longitudinal data, we are currently unable to discuss progress.

b. Analyses encompassed 3,922 hospice organizations and 1,218,786 patient stays across the United States.
Region:
South: 39.3%
West: 25.1%
Midwest: 23.1%
Northeast: 11.3%
Territories: 0.94%
Unknown: 0.25%

Urban/rural status:
Urban: 75.8%
Rural: 23.9%
Unknown: 0.31%

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.**
We only have one year of data to report, which is not enough to show trends over time.

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.**
None

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.)
This measure is part of the NPCRC Key Palliative Measures Bundle. Refer to the NPCRC cover letter and table of bundle measures for description of the selection and harmonization of the Key Palliative Measures Bundle.

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 **Attachment:** Appendix_A.1_NQF_1639.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** University of North Carolina-Chapel Hill
Co.2 **Point of Contact:** Laura, Hanson, lhanson@med.email.unc, 919-843-4096-
Co.3 **Measure Developer if different from Measure Steward:** University of North Carolina- Chapel Hill
Co.4 **Point of Contact:** Laura, Hanson, lhanson@med.email.unc, 919-843-4096-

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.
The Carolinas Center for Medical Excellence PEACE Project Technical Expert Panel

The PEACE project team convened a 14-member Technical Expert Panel (TEP) of nationally recognized experts with extensive experience in the following areas: medical or nursing expertise in hospice and palliative care, methods and instrumentation, and quality improvement. Using criteria provided by the CCME study team, TEP members rated each potential quality measure on four criteria: importance, scientific soundness, feasibility and usability.

Mary Ersek, PhD, RN, Research Associate Professor, Swedish Medical Center- Pain Research Department, Seattle, WA

Betty R. Ferrell, PhD, FAAN, Research Scientist, City of Hope National Medical Center, Duarte, CA
Sean Morrison, MD, Mount Sinai Medical Center, NY, NY
Richard Payne, MD, Director, Duke Institute on Care at the End of Life, Duke Divinity School, Durham, NC
Chris Feudtner, MD, PHD, MPH, Children's Hospital of Philadelphia, Philadelphia, PA
Karen Steinhauser, PhD, Research Health Scientists, Center for Health Services Research in Primary Care, Durham VA Medical Center and Duke University, Durham, NC
Joan M. Teno, MD, Professor of Community Health and Medicine, Center for Gerontology and Health Care Research, Brown University, Providence, RI
Melanie Merriman, PhD, MBA, Touchstone Consulting, North Bay Village, FL
Sydney Dy, MD, MSc, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
David Casarett, MA, MD, Assistant Professor, Division of Geriatrics, Institute on Aging and Center for Bioethics, University of Pennsylvania School of Medicine and NHPCO Board of Directors
Judi Lund-Person, Vice President, Division of Quality, National Hospice and Palliative Care Organization, Washington, DC
Jean Kutner, MD, MSPH, Associate Professor, University of Colorado Health Sciences Center, Denver, CO
Lin Simon, Analyst, National Hospice and Palliative Care Organization, Washington, DC
Karen Pace, NAHC

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2010

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

Ad.4 **What is your frequency for review/update of this measure?** 3 years or as required

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

Ad.6 **Copyright statement:**

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:** This measure is part of the NPCRC Key Palliative Measures Bundle. Refer to the NPCRC cover letter and table of bundle measures for description of the selection and harmonization of the Key Palliative Measures Bundle.

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 #: 1647
Measure Title: Beliefs and Values - Percentage of hospice patients with documentation in the clinical record of a discussion of spiritual/religious concerns or documentation that the patient/caregiver did not want to discuss.
Measure Steward: University of North Carolina-Chapel Hill
Brief Description of Measure: This measure reflects the percentage of hospice patients with documentation of a discussion of spiritual/religious concerns or documentation that the patient/caregiver/family did not want to discuss.
Developer Rationale: Spiritual care is an essential domain of hospice and palliative care, according to national guidelines based on strong evidence from patient and family survey research. One of the unique aspects of hospice care involves a true interdisciplinary approach providing care for both the physical and psychosocial and spiritual needs of the patient and caregiver, and access to spiritual care if needed. Discussion of spiritual concerns is the core of a rigorous assessment of spiritual care needs and is essential to assuring that these needs are met. This measure will help agencies improve processes for addressing spiritual/religious concerns for patients and families receiving hospice care.

Numerator Statement: Patients whose medical record includes documentation that the patient and/or caregiver was asked about spiritual/existential concerns within 5 days of the admission date.
Denominator Statement: Seriously ill patients 18 years of age or older enrolled in hospice.
Denominator Exclusions: Testing has only been done with the adult population; thus patients younger than 18 are excluded.

Measure Type: Process
Data Source: Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record
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 provides a rationale for the relationship of this process of care (document of a discussion of spiritual/religious concerns) to patient outcomes (i.e., spiritual care needs are addressed).
- The developer cites a National Consensus Project guideline and an NQF-endorsed Preferred Practice as evidence for the measure, although they state that formal studies examining this process of care do not exist.
- During its 2012 evaluation of the measure, additional data from a retrospective study was presented to the Palliative/End-of-Life (EOL) Committee. This study showed that patients whose records documented a conversation of their spiritual or religious concerns demonstrated improvement in overall spiritual distress as opposed to those whose records did not document this conversation. It is unclear whether this is the Deyta study referred to in the submission. While the Committee noted the lack of evidence for this measure but agreed that the benefits of the measure outweigh the risks.

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:

- Although the developer attested that there have been no changes in the evidence since it was endorsed in 2012, a new guidelines is cited.
    - 2013 ICSI guideline Palliative Care for Adults. The pertinent recommendation from this graded guideline states that "A spiritual assessment should be an integral part of the palliative care plan. (Low Quality Evidence, Strong Recommendation)".

Exception to evidence

- Because the evidence for this measure is only on expert opinion, it is insufficient to meet NQF's criterion for evidence. However, an exception to the evidence criterion is allowed if the Committee agrees that empirical evidence is not needed to hold providers accountable for the measure and that other outcome measures or evidence-based measures are not available or feasible at this time.

Guidance from the Evidence Algorithm

Process measure, not based on systematic review and grading (Box 3) → No empirical evidence submitted (Box 7) → No related outcome measures exist at this time (Box 10) → Systematic assessment of expert opinion (guidelines) (Box 11) → If Committee agrees it is OK/beneficial to hold providers accountable for performance in the absence of empirical evidence of benefits to patients → rate as INSUFFICIENT WITH EXCEPTION

Questions for the Committee:

- *Are there, or could there be, performance measures of a related health outcome, OR evidence-based intermediate clinical outcomes, intervention/treatment?*
- *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*
- *Does the SC agree that it is acceptable (or beneficial) to hold providers accountable without empirical evidence?*

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

Rationale: No evidence linking documentation spiritual/religious concerns to improved patient outcomes.

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

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

- The developer provided data from the Hospice Quality Reporting Program (HQRP) for FY15. The data were

collected from 3,922 hospice organizations and approximately 1.2 million patient stays.

- The developer reported that 29.4% of hospices had perfect scores and 23.4% of hospices scored below 90%.

Mean	92.2%
Range	0% - 100%
10th percentile	78.6%
25th percentile	90.8%
Median	97.7%
75th percentile	100%
90th percentile	100%

Disparities

- Using the same dataset described above:

Subgroup	**Hospice-level**	
	Score	**Wilcoxon-Mann-Whitney test p-value**
Race		
Non-white ≥ national median of 11.9%	93.4%	p<0.001
Non-white < national median of 11.9%	91.1%	
Sex		
Female ≥ national median of 55.2%	92.3%	p=0.79
Female < national median of 55.2%	92.2%	
Medicaid status (proxy for SES)		
Medicaid patients >= national median of 21.5%	90.2%	p<0.001
Medicaid patients < national median of 21.5%	93.9%	
Geographic location		
Rural	94.5%	p<0.001
Urban	91.5%	

Questions for the Committee:

- *Is there a gap in hospice care (of spiritual/religious discussions) that warrants a national performance measure?*

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.

* Process Measure Maintenance: This measure reflects the percentage of hospice patients with documentation of a discussion of spiritual/religious concerns or documentation that the patient/caregiver/family did not want to discuss.
No Quality, Quantity or Consistency of Evidence.
Clinical Practice Guideline and Expert Opinion.
The developer cites a National Consensus Project guideline and an NQF-endorsed Preferred Practice as evidence for the measure, although they state that formal studies examining this process of care do not exist.

- During its 2012 evaluation of the measure, additional data from a retrospective study was presented to the Palliative/End-of-Life (EOL) Committee. This study showed that patients whose records documented a conversation of their spiritual or religious concerns demonstrated improvement in overall spiritual distress as opposed to those whose records did not document this conversation. It is unclear whether this is the Deyta study referred to in the submission. While the Committee noted the lack of evidence for this measure but agreed that the benefits of the measure outweigh the risks.

Update: o 2013 ICSI guideline Palliative Care for Adults. The pertinent recommendation from this graded guideline states that "A spiritual assessment should be an integral part of the palliative care plan. (Low Quality Evidence, Strong Recommendation)".

* No evidence presented.
* This is a direct measure of the percentage of hospice patients over 18 yrs old who have documentation in the clinical record of a discussion of spiritual/religious concerns or documentation that the patient/caregiver did not want to discuss these concerns within 5 days of admission to hospice. The measure is supported by several guidelines in including NCP and NQF but lacks a research literature including any systematic reviews or QQC data and this reviewer is unaware of any systematic review of expert opinion in this area. The developers do report a new guideline however it does not change the basic thrust of the guidelines already presented. There are multiple studies which suggests that patients want to discuss these concerns with their medical teams even when they do not specifically request them and that having these discussions improves patient and family satisfaction and decreases use of aggressive care at EOL. This measure was originally approved as insufficient evidence with exception. The recommendation of the staff is to continue that exception which seems reasonable and appropriate especially given that this is the only spiritual care measure currently approved by the committee. As is already true with 1641, it would seem appropriate to investigate extending this measure to palliative care patients.
* The measure relates to the process of relieving distress by asking a patient about their spiritual beliefs within 5 days of entry into a hospice program. The measure is based on an expert opinion that such conversations are of benefit to patients. This measure tracks the occurance of such patient engagement.
* This is a maintenance of endorsement therefore less emphasis is placed on evidence. This measure was last endorsed in 2012. There have been no changes in the evidence since the measure was last evaluated.

1b.
* Preliminary rating for opportunity for improvement: Moderate
* Yes, average performance is above 90%;no subgroups below 90%.
* The developer provides data on the normally investigated groups for a disparity analysis. While there are significant differences statistically, like with measure 1641. the overall compliance percentages are sufficiently and so uniformly high that the clinical significance of these differences is questionable. The question for this measure as with 1641 is whether it has topped out in this population. Clearly the high compliance is due to its inclusion in the Hospice Quality Reporting Program database where reporting is required for all hospices. The question then is would CMS be committed to keeping this measure if NQF made it inactive. Clearly, this data is at least potentially useful and required to drive further investigation into what patient's spiritual needs are and how to meet them. If it were dropped, this reviewer has some concern that the development of that evidence would be impaired and the inclusion of spiritual needs in care planning would be neglected.
* Yes, performance data on the measure was provided. The data indicates that 29.4% of hospices had perfect scores and 23.4% of hospices scored below 90%. Yes, data on the measure by population subgroups was provided. Racial analysis, gender disparity analysis, socioeconomic disparity analysis, and rural-urban disparity analysis were performed. From the racial analysis, it was found that Black non-Hispanic patients had the lowest rate of documentation of spiritual/religious beliefs (90.6%) and White non-Hispanic (92.4%) had the highest rate. The gender analysis did not demonstrate statistically significant differences between the rate of documentation between female patients and the national median. For the socioeconomic disparity analysis, it was found that a statistically significant lower rate of documentation of a discussion of spiritual/religious belief was found for non-Medicaid patients (90.6%) than for Medicaid patients (91.7%). Lastly, the rural-urban data analysis indicated that there is a statistically significant difference between rural (94.5%) and urban (91.5%) hospices.
* Performance data was provided. 23.4 % of hospices reporting scored below 90% while 29.4 % had perfect scores. The measure demonstrates a moderate rating for opportunity for improvement using data from the HQRP for FY15. Using the same data set there was no significant evidence for disparity

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 sources specified in the submission form include EHRs and other electronic clinical data. **Specifications:** • This measure is specified for the facility level of analysis in the hospice setting. A higher score indicates better

quality.

- The numerator (patient whose medical record includes documentation that patient and/or caregiver was asked about spiritual/existential concerns within 5 days of admission date) and denominator (patients enrolled in hospice) are identified from the Hospice Item Set, a standardized, patient-level dataset used in the Medicare Hospice Quality Reporting Program.
- The developer provides examples of discussions that would meet the meet the measure.
- Patients younger than 18 are excluded from the measure.
- A calculation algorithm is provided.

Questions for the Committee :

- *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:

- Data element validity testing was conducted by comparing agency-abstracted data to that abstracted by a research study abstractor (considered the gold standard). NQF guidance indicates that if data element validity is demonstrated, additional data element reliability testing is not required.

Describe any updates to testing

- Score-level testing for the hospice setting included split-half reliability, signal-to-noise, and stability analyses at the facility level.

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:

- For data element testing, see method(s) of validity testing, below.
- The developers also report conducting a test-retest bivariate correlation to assess the consistency of the measure from one period of time (2009) to another time (2010). However, NQF does not consider analysis of data across time to be an appropriate method of testing the reliability of the measure score.
- In updated testing, the reliability of the measure score was assessed using FY15 data from the Hospice Quality Reporting Program (n=3,992 hospice organizations and n=1,218,786 patient stays). Two types of reliability testing were conducted: the first via a signal-to-noise analysis and the second using a split-sample (or "test-retest") methodology. Both are appropriate methods for testing reliability. Stability analysis examined changes in provider's performance over time. As noted earlier, 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:

- The reliability testing results were:
    - Split-half analysis ICC=0.94.
        - The ICC reflects the percentage of variance in score results that is due to "true" or real variance between the hospices. A value of 0.7 is often regarded as a minimum acceptable reliability value.

- Developers report a signal-to-noise ratio of 0.99.
  - A signal-to-noise analysis quantifies the amount of variation in a performance measure that is due to true differences between hospices (i.e., signal) as opposed to random measurement error (i.e., noise). Results will vary based on the amount of variation between the providers and the number of patients treated by each provider. This method results in a reliability statistic that ranges from 0 to 1 for each hospice. A value of 0 indicates that all variation is due to measurement error and a value of 1 indicates that all variation is due to real differences in hospices performance. A value of 0.7 often is regarded as a minimum acceptable reliability value.

Guidance from the Reliability Algorithm

Precise specifications (Box 1) → Empirical reliability testing with measure as specified (Box 2) → Score-level testing (Box 4) → Appropriate method (Box 5) → High certainty that measure results are reliable (Box 6a) → High

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

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**

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:

- Data element validity testing was conducted by comparing agency-abstracted data to that abstracted by a research study abstractor (considered the gold standard). This is an appropriate method of data element validation.

Describe any updates to validity testing:

- Validity of the measure score was conducted using FY15 data from the Hospice Quality Reporting System (HQRP).

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

Validity testing method:

- Data element testing:
  - The developers compared agency-abstracted data to that abstracted by a research study abstractor

(considered the gold standard). Ten records per agency were examined, but the total number of agencies/records was not reported. The developers noted that testing was conducted "*on the entire data collection tool used in the AIM Project on which the measures are based*".

- Score-level testing:
  - Using FY15 data, from the Hospice Quality Reporting System (n=3,992 hospice organizations; n=1,218,786 patient stays), the developer conducted non-parametric Spearman rank correlation analysis between this measure and 5 other hospice quality measures. They hypothesized that agencies should perform similarly on assessment processes at hospice admission and expected the resulting correlations to be high. This is an appropriate method of score-level validation. **NOTE** that the results table indicates that measure "NQF #1647 (modified)" was included in this testing. The developers do not say how this measure was modified for testing. NQF requires that testing be conducted for the measure as specified, so this testing may not be sufficient.
  - Face validity was assessed by proving quarterly reports to agencies participating in the study. After revising measure abstraction tools, etc. based on agency feedback, the agencies were given a second opportunity to comment on the feasibility and accuracy of the measure. Additional feedback was sought from 2 advisory/expert panels. According to NQF guidance, the face validity of the measure score as a quality indicator may be adequate if it explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. This testing does not appear to conform to NQF's requirements for face validity.

Validity testing results:

- Data element testing:
  - Developers report a single kappa value of 0.795 (95% CI 0.79-0.80, agency range: .70-.90).
  - 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 0.795 means that the raters agreed 79.5% of the time over and above what would be expected by chance alone. According to the Landis and Koch classification, this represents "substantial" agreement.
  - NQF guidance indicates that data element testing should be conducted for all critical data elements, although at minimum, results about the numerator, denominator, and exclusions should be provided. For this measure, only a single kappa value is reported. It is unclear which data element(s) this statistic refers to, as the developer describes it as a kappa value "for all categorical variables".
- Score-level testing:
  - Correlation results were positive and statistically significant, confirming the developer's hypothesis. However, the magnitude of the correlation was lower than they had expected. The developer states reasons for this may be due to clustering of scores, skewed distributions, and low score variability.
  - The developer does not provide results from the face validity assessment.

Questions for the Committee:

- *How was this measure modified for testing? Should the testing results be accepted, as they appear not to reflect the measure as specified?*
- *Are the test samples adequate to generalize for widespread implementation?*
- *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 examined the effect of excluding patients less than 18 years of age. Using FY15 data from the HQRP, the developers found this exclusion has a minimal impact on the size of the measure denominator and the distribution of the measure score.

Questions for the Committee:

○ *Are the exclusions consistent with the evidence?*

○ *Are any patients or patient groups inappropriately excluded from the 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)*:

- Meaningful differences among hospices were examined using FY15 data from the Hospice Quality Reporting Program. The developers examined 95% confidence intervals to determine the proportion of hospices with results significantly different from the national hospice-level mean. Results indicated 61.8% of hospices had a quality measure score that was significantly difference from the national mean. Hospices were more likely to report scores above the national mean than below the national mean (43% vs 18.8%, respectively).

Question for the Committee:

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

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

2b7. Missing Data

- For hospices, the rate of missingness is low (0.02% - 0.08% at the patient-level). More than 95% of hospices had no missing information for the key data element from the Hospice Item set that make up this measure.

Guidance from the Validity Algorithm

Specifications consistent with evidence (Box 1) → potential threats to validity assessed (Box2) →empirical validity testing of both the measure score and of the data elements, but unclear if with measure as specified (Box 3) → Insufficient

If measure score tested as specified, validity testing of the measure score (Box 6) →method appropriate (Box7) →moderate certainty (Box 8b)→ Moderate

If measure score not tested as specified, testing conducted at the data element level (Box 10) → critical data elements tested with appropriate method, but results are not clear (Box 11) → Insufficient OR Moderate **if** results are for the measure numerator

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

Rationale: Need to understand how the measure was modified for score-level testing. Results of data element validity are unclear, but fine if results are for the measure numerator. Face validity assessment does not appear to conform to NQF requirements.

Committee pre-evaluation comments

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

2a, 2b.

Reliability – Specifications

* Elements clearly defined by the specific examples provided, as well as what would not count.
* Very straightforward. No concerns.
* All the data elements seem to be clearly defined. The algorithm is provided and seems clear. Thus it would seem that this measure could be consistently implemented.
* Question: Is it clearly defined for providers what constitutes a spiritual assessment for this measure in order for it to be consistently implemented?
* this measure produces consistent and credible results about the quality of care when implemented. the data elements are clearly defined

This measure can be consistently implemented because the data is part of the EMR and identified from the Hospice Item Set in the Medicare HQRP; The steps in the calculation algorithm are clear

Validity – Specifications

* Specifications consistent with evidence per developer.There is no evidence, and this is based on a guideline. This is by exception.

* Not enough evidence to show inconsistency.
* The specifications are consistent with the evidence.
* There may not be enough evidence as to what the intended target population finds meaningful as this is based on an expert opinion what not direct data of the relief provided through the consultation to a large group of patients vs. not receiving the consultation.
* Specifications are consistent with the evidence. No inconsistencies with what the target audience finds valuable.

Reliability – Testing
* Adequate scope. High reliability documented.
* Results of testing demonstrate high degree of reliability; signal-to-noise and split-sample analysis provided.
* Both the signal-noise ratio and split half analysis results were very positive indicating sufficient reliability so that differences in performance can be identified and indicating strong reliability for this measure. The sample size was quite large and appears more than sufficient,
* Yes, the measure was reliability tested with an adequate scope to generalize for widespread implementation and with an appropriate method. The split-half analysis ICC result was 0.94, which exceeds the minimum acceptable reliability value. The signal-to-noise ratio is 0.99, which also exceeds the minimum acceptable reliability value.
* Data element validity testing was conducted by comparing agency abstracted data to that abstracted by a research study abstractor . NQF guidance indicates that if data element validity is demonstrated additional data element reliability testing is not required. Score level testing for the hospice setting was completed. Reliability testing results were above the minimum reliability values. Preliminary testing for reliability is high

Validity Testing
* Validity of the measure score was conducted using FY15 data from the Hospice Quality Reporting System (HQRP)
Score level testing: NOTE that the results table indicates that measure "NQF #1647 (modified)" was included in this testing. The developers do not say how this measure was modified for testing. NQF requires that testing be conducted for the measure as specified, so this testing may not be sufficient.
Face Validity: This testing does not appear to conform to NQF's requirements for face validity.
Rating: Insufficient.
* High degree validity; gold standard method using nurse researchers.
* The major issue with the validity testing is that the developers' results report indicates that the testing was done on a modified version of the measure but no details of the modification are given. If the modification is not true to the original measure, this testing does not meet NQF guidelines and the testing data needs to be graded as "insufficient". If the modification is consistent with the original measure, the data is likely graded as "moderate". The scores on the testing were positive and statistically significant. The face validity was tested however the developers do not specify whether the scores can be used to distinguish good from poor quality care as specified by NQF guidelines. The data element testing reports only one Kappa Score. However if this represents the scores for each of the elements as the developers suggest, the agreement is substantial.
* Yes, the measure was validity tested with adequate scope to generate for widespread implementation and with an appropriate method. The measure has significant positive correlations with the other quality metrics, which indicates that hospices providing higher-quality care in this area performed better in other areas at hospice admission.

Question: The measure was modified for testing. We need to know how it was modified and the implications of that on the validity testing.
* Validity testing demonstrates that the measure data elements are correct and the measure score correctly reflects quality of care and identifies differences in quality
testing was conducted at both the data element and score levels . The score level testing was done utilizing a modification of the measure which may invalidate the study.

Threats to Validity
* The developer examined the effect of excluding patients less than 18 years of age. Using FY15 data from the HQRP, the developers found this exclusion has a minimal impact on the size of the measure denominator and the distribution of the measure score.
For hospices, the rate of missingness is low (0.02% - 0.08% at the patient-level). More than 95% of hospices had no missing information for the key data element from the Hospice Item set that make up this measure.
* Only exclusion is based on age less than 18; agree that this would not threaten validity.
* The developers tested the threat to validity of excluding persons under 18 years of age and found minimal impact. This exclusion is appropriate and no other groups are excluded. The measure does appear to be able to identify meaningful differences in performance and there was minimal missing data.

There is no risk adjustment.
* Minors are excluded from the measure. The basis for exclusion is not provided.
Question: why are minors excluded? Is there a different intervention that is more appropriate?

The measure is not risk-adjusted.

Unclear whether this measure identifies meaningful differences in quality of hospice providers or hospice care.

No, missing data does not constitute a threat to the validity of the measure since the rate of missingness is low (0.02% - 0.08%). According to the developers, more than 95% of hospices had no missing information for the key data elements from the Hospice Item set that make up this measure.
* No missing data, no risks 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.

- All required data elements are available in electronic form, abstracted from a hospice record. The developers report that [then] current program participants had reported difficulties related to the ability to query this information out of their EHR. Data collection for all measures is achieved through use of the Hospice Item Set (HIS). It is unclear if the HIS is automatically populated by the EHR.
- The developer reports cost estimates are not available. According to the Guidance Manual for Completion of the Hospice Item Set, completing an entire response for any one admission takes 20 minutes (though seven performance measures, of which this measure is only one, are collected as part of this process).

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

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

Committee pre-evaluation comments
Criteria 3: Feasibility

* All required data elements are available in electronic form, abstracted from a hospice record. The developers report that [then] current program participants had reported difficulties related to the ability to query this information out of their EHR. Data collection for all measures is achieved through use of the Hospice Item Set (HIS). It is unclear if the HIS is automatically populated by the HER
* Available through electronic submission to CMS but not clear whether the data are derived from EHR.
* All of the data elements are routinely generated and available in electronic form. While this measure would seem to have obvious uses in care delivery, the developers do not report any such use. That use would seem essential to the rationale for having and continuing this measure.
* Developers reported that current program participants had difficulties related to their ability to query this information out of their EHR. This measure is currently used in hospice quality reporting to CMS.
* No concerns, current EHR and data sets are collected and feasibility is high as the protocols are already in place and widely utilized

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

Publicly reported? ☐ **Yes** ☒ **No**

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

Accountability program details:

- This measure is included in the Hospice Quality Reporting Program (HQRP). Medicare-certified hospices are required to submit an Hospice Item Set (HIS)-Admission record and an HIS-Discharge record for each patient admission on or after July 1st, 2015. In FY2015, 3,992 hospice organizations provided measure data for 1,218,786 patient stays. The distribution of reporting hospice agencies by geographic area is provided.

Improvement results:

- Longitudinal data for this measure 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 2016, the Measures Application Partnership Post-Acute Care and Long-Term Care workgroup encouraged the continued development of a composite measure of Hospice Item Set measures, which includes this measure.
- In February 2013, the Measure Application Partnership supported the measure for inclusion in the Hospice Quality Reporting Program, and the Physician Quality Reporting System (PQRS) to address a core concept (i.e., pain assessment) not addressed in the programs' measure sets. Public comments from American Academy of Hospice and Palliative Medicine (AAHMP), Center to Advance Palliative Care (CAPC), and National Coalition for Hospice and Palliative Care (NCHPC) supported the inclusion of this measure in the PQRS. CAPC and NCHPC suggested that the measure should be expanded beyond hospice patients. The measure is not currently included in the PQRS program, but it is included in the Hospice Quality Reporting Program.

***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**

Committee pre-evaluation comments

Criteria 4: Usability and Use

* Current use not publicly reported.
This measure is included in the Hospice Quality Reporting Program (HQRP). Medicare-certified hospices are required to submit an Hospice Item Set (HIS)-Admission record and an HIS-Discharge record for each patient admission on or after July 1st, 2015. In FY2015, 3,992 hospice organizations provided measure data for 1,218,786 patient stays.
* Part of hospice quality reporting system.
* The do not seem to be any related or competing measures. The developers do report some data that suggests that more patients who had a documented discussion had improved spiritual distress scores as opposed to patients who did not have this discussion. The developers report that they do not yet have sufficient longitudinal data to report on improvement of use. The developers report no unexpected findings or unintended consequences.

This measure does clearly have the potential to further the ability to meet patients' spiritual needs which has been shown to have significant positive consequences for overall quality of care. It would be anticipated that the discussions

resulting from this measure would raise patient satisfaction. The measure's further testing would also set the stage for expansion of this measure into the general palliative care population. Given that the literature suggests that most patients want to have this discussion and do not find it a burden, it would seem that the benefits of this measure outweigh any potential unintended consequences.
* This measure is not currently being publicly reported; however, it is being used in an accountability program. The measure is included in the Hospice Quality Reporting Program (HQRP). The developers did not report any unintended consequences.
* this measure is currently utilized in an accountability program:it is included in the HQRP. Medicare certified hospices are required to submit an HIS record and an HIS Discharge record for each patient admission as of July 1, 2015. The measure is not being publicly reported. Longitudinal data is not yet available. the developer did not report any unintended consequences. I believe this measure provide performance results that can be used to further the goal of high quality healthcare that reflects provision of spiritual care, an essential domain of hospice and palliative care

Criterion 5: Related and Competing Measures

Related or competing measures

- None

Harmonization

- N/A

Pre-meeting public and member comments

Name: Michele Galioto, RN, MSN

Organization: Oncology Nursing Society

Comment: Recommend making the measure specific to palliative or hospice care, not both in the same measure. Palliative care is not restricted to inpatient treatment.

Consider reworking measures 1634 & 1637 to be a singular, stronger measure related to screening for pain.

Consider reviewing measure 209 along with measures 1634 & 1637 to strengthen measures for pain assessment and intervention.

Consider incoporating recommended intervals for screening as the current measure indicates one assessment but one screening is not sufficient in this setting. Perhaps "at each patient encounter" is more appropriate?

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.

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 1647	**NQF Project:** Palliative Care and End-of-Life Care

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)***:**

As mentioned previously, a true interdisciplinary approach providing care for both the physical and psychosocial and spiritual needs of the patient and caregiver is a unique aspect of hospice care. A discussion of spiritual concerns is the core of a rigorous assessment of spiritual care needs and is essential to assuring that these needs are met.

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

Clinical Practice Guideline

Other

Expert Opinion

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)***:**

In order to address hospice patients' spiritual distress, the hospice team must discuss any spiritual/religious concerns with the patient.

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)***:** No other known formal studies.

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):* Because there are no known studies on this topic, there is a low level of certainty regarding net benefit.

1c.7 **Consistency of Results <u>across Studies</u>** *(Summarize the consistency of the magnitude and direction of the effect):* No other known studies for comparison.

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

Because there are no known studies on this measure, there is a low level of certainty regarding net benefit.

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:** Because there are no known studies on this measure, grading has not been performed.

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

1c.12 **If other, identify and describe the grading scale with definitions:** Grading has not been completed.

1c.13 **Grade Assigned to the Body of Evidence:** N/A

1c.14 **Summary of Controversy/Contradictory Evidence:** Comparative data is limited throughout the industry for this measure. Data for this measure comes solely from participation in Deyta´s proprietary system, Quality Navigator, however could be obtained from other sources. Participants include hospices with varied characteristics for a representative sample of hospices in the industry: for profit and not-for-profit, single and large multi-location agencies, small (ADC < 50) to very large (> 1000), representing multiple regions of the country, use of an EHR and those with paper documentation.

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

N/A

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

Health Care Guideline: Palliative Care for Adults

Guideline 8. Spiritual Aspects of Care: A spiritual assessment should be an integral part of the palliative care plan.

National Consensus Project Guidelines (2009)

Guideline 5.1: Spiritual and existential dimensions are assessed and responded to based upon the best available evidence, which is skillfully and systematically applied.

National Consensus Project and National Quality Forum Framework and Preferred Practices for Palliative and Hospice Care

Preferred Practice 20: Develop and document a plan based on an assessment of religious, spiritual, and existential

concerns using a structured instrument, and integrate the information obtained from the assessment into the palliative care plan.

1c.17 **Clinical Practice Guideline Citation:** McCusker M, Ceronsky L, Crone C, Epstein H, Greene B, Halvorson J, Kephart K, Mallen E, Nosan B, Rohr M, Rosenberg E, Ruff R, Schlecht K, Setterlund L. Institute for Clinical Systems Improvement. Palliative Care for Adults. Updated November 2013.p.26

1c.18 **National Guideline Clearinghouse or other URL:** http://www.guideline.gov/content.aspx?id=47629

1c.19 **Grading of Strength of Guideline Recommendation.** Has the recommendation been graded? Yes

1c.20 **If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias:** Institute for Clinical Systems Improvement (ICSI)

1c.21 **System Used for Grading the Strength of Guideline Recommendation:** GRADE Methodology

1c.22 **If other, identify and describe the grading scale with definitions:**

1c.23 **Grade Assigned to the Recommendation:** Low Quality Evidence, Strong Recommendation

1c.24 **Rationale for Using this Guideline Over Others:** We chose to use the Palliative Care for Adults guideline from the Institute for Clinical Systems Improvement (ICSI) because it is evidenced based, uses GRADE methodology, and is included in the National Guideline Clearinghouse.

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

1c.25 Quantity: Low 1c.26 Quality: Low1c.27 Consistency: Low

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 #: 1647
De.2. **Measure Title:** Beliefs and Values - Percentage of hospice patients with documentation in the clinical record of a discussion of spiritual/religious concerns or documentation that the patient/caregiver did not want to discuss.
Co.1.1. **Measure Steward:** University of North Carolina-Chapel Hill
De.3. **Brief Description of Measure:** This measure reflects the percentage of hospice patients with documentation of a discussion of spiritual/religious concerns or documentation that the patient/caregiver/family did not want to discuss.
1b.1. **Developer Rationale:** Spiritual care is an essential domain of hospice and palliative care, according to national guidelines based on strong evidence from patient and family survey research. One of the unique aspects of hospice care involves a true interdisciplinary approach providing care for both the physical and psychosocial and spiritual needs of the patient and caregiver, and access to spiritual care if needed. Discussion of spiritual concerns is the core of a rigorous assessment of spiritual care needs and is essential to assuring that these needs are met. This measure will help agencies improve processes for addressing spiritual/religious concerns for patients and families receiving hospice care.

S.4. **Numerator Statement:** Patients whose medical record includes documentation that the patient and/or caregiver was asked about spiritual/existential concerns within 5 days of the admission date.
S.7. **Denominator Statement:** Seriously ill patients 18 years of age or older enrolled in hospice.
S.10. **Denominator Exclusions:** Testing has only been done with the adult population; thus patients younger than 18 are excluded.

De.1. **Measure Type:** Process
S.23. **Data Source:** Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record
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?** N/A

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
1647_Evidence_MSF5.0_Data.doc,1647_Evidence_3.17.16-635938216838986957.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)*
Spiritual care is an essential domain of hospice and palliative care, according to national guidelines based on strong evidence from patient and family survey research. One of the unique aspects of hospice care involves a true interdisciplinary approach providing care for both the physical and psychosocial and spiritual needs of the patient and caregiver, and access to spiritual care if needed. Discussion of spiritual concerns is the core of a rigorous assessment of spiritual care needs and is essential to assuring that these needs are met. This measure will help agencies improve processes for addressing spiritual/religious concerns for patients and families receiving hospice care.

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.*
CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1647 - Percentage of hospice patients with documentation in the clinical record of a discussion of spiritual/religious concerns or documentation that the patient/caregiver did not want to discuss.

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.

The mean score for this QM was 92.2% with a range from 0% to 100%, the median was 97.7%, the interquartile range was 9.2, and the standard deviation was 13.6. For this QM, 29.4% of hospices had perfect scores and 23.4% of hospices scored below 90%.

Scores by decile:
10th percentile 78.6%
25th percentile 90.8%
Median 97.7%
75th percentile 100%
90th percentile 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.**
N/A

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.*
Racial disparity analysis. We conducted racial and ethnic disparity analyses at both the patient-stay and hospice levels. At the patient-stay level, we compared the percentage of patients with documentation of a discussion of spiritual/religious concerns or documentation that the patient/caregiver did not want to discuss among different racial and ethnic groups. Patients were grouped by their racial and ethnic identification as follows: white non-Hispanic, black non-Hispanic, other non-Hispanic, or Hispanic. Other non-Hispanic includes patients who identify as American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, or who identify as more than one race. A Chi-square test was performed to determine if there were any statistically significant differences in spiritual/religious concerns between groups. The lowest rate of documentation was found for Black non-Hispanic patients (90.6%), and the highest rate was among patients identifying as White non-Hispanic (92.4%). Differences in the rate of documentation of spiritual/religious beliefs by racial identification were found to be statistically significant ($p<0.001$).

Analyses at the hospice level examined the differences in this measure across two groups: hospices with proportions of non-white patients that are greater than or equal to the national median proportion (11.9%), and hospices with fewer non-white patients than the national median. For this analysis, white non-Hispanic patients were included in the white group, and all other racial and ethnic identifications, as described above, were grouped as non-white. We ran a Wilcoxon-Mann-Whitney test for statistical dependence between group and QM score. The results showed that the QM score was significantly different between the two groups of hospices

(93.4% compared to 91.1%, p<0.001). Although statistically significant results were found actual differences in rates do not seem to be clinically substantial.

Gender disparity analysis. We performed both the patient stay- and hospice-level analyses on gender disparity. At the patient-stay level, we compared the percentage of patients with documentation of a discussion of spiritual/religious concerns or documentation that the patient/caregiver did not want to discuss between female and male patients. A Chi-square test was performed to determine if there were any statistically significant differences in documentation of a discussion of spiritual/religious concerns between groups. We found a slightly higher rate of documented discussion for female patients (92.1%) than for male patients (92.0%). Differences in the rate of documentation of a discussion of spiritual/religious concerns by gender were statistically significant (p=0.032). At the hospice level, we examined the differences in this measure across hospices with proportions of female patients that are greater than or equal to the national median proportion (55.2%). We ran a Wilcoxon-Mann-Whitney test for statistical dependence between group and QM score. The results showed that the QM score was not statistically significantly different between the two groups of hospices split by median proportion of female patients (92.3% compared to 92.2%, p=0.79).

Socioeconomic disparity analysis. We performed socioeconomic disparity analyses at both the patient-stay and hospice levels using the same methods described in previous disparity analyses. Medicaid status was used as a proxy measure of low socioeconomic status. A statistically significant (p<0.001) lower rate of documentation of a discussion of spiritual/religious concerns was found for non-Medicaid patients (90.6%) than for Medicaid patients (91.7%). The highest rate of documented discussion was seen for patients with Medicaid status missing (95.6%). At the hospice level, the results showed that the QM score was significantly different between the two groups of hospices split by median proportion of Medicaid patients, 21.5% (90.2% compared to 93.9%, p<0.001). The statistically significant results at the patient-stay level may indicate that quality of hospice care, measured by documentation of a discussion of spiritual/religious concerns or documentation that the patient/caregiver did not want to discuss, for non-Medicaid patients is lower than for Medicaid patients. The significant findings at the hospice level indicate that hospices with a smaller proportion of Medicaid patients are less likely to document discussions of spiritual/religious concerns or that the patient/caregiver did not want to discuss. A potential caveat of this analysis is that the data used for this analysis may not be a reliable indicator of a patient's Medicaid eligibility status. This data is missing for nearly a quarter of patients. In addition, some of the Medicaid numbers submitted do not appear to be legitimate Medicaid numbers. However, the data used in this analysis is the only national patient-level assessment data that is currently available in hospices. We will update this analysis as more-accurate data sources are available and accessible.

Rural-urban disparity analysis. We compared the QM score between rural and urban hospices using a Wilcoxon-Mann-Whitney test. The results showed that the QM score was statistically significantly different between rural and urban hospices (94.5% compared to 91.5%, p<0.001). Although statistically significant results were found actual differences in rates do not seem to be clinically substantial.

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.
N/A

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
Affects large numbers, Patient/societal consequences of poor quality, Other
1c.2. If Other: Data are captured as part of the Hospice Item Set Hospice Quality Reporting Program requirement.

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.
Hospice care is an increasingly important piece of the healthcare continuum, both from the number of patients served and the financial benefits (reducing costs associated with end-of-life care and re-hospitalizations for home health care and hospitals). According to NHPCO Facts and Figures (2014), over 1.6 million patients received services from approximately 6100 hospice throughout the United States, and serves approximately 44% of people who die in the US.(1)

Spiritual care has been shown to be a critical element of quality of life at the end of life. Evidence comes from rigorous survey studies of patients living with serious life-limiting illnesses, and family caregivers and healthcare providers who care for them. (2-5) This measure is in accordance with the Clinical Practice Guidelines for Quality Palliative Care, Guidelines 5.1, and the National Quality Forum-endorsed preferred practices #20.(6)

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

1. NHPCO Facts and figures: hospice care in America 2015 edition
http://www.nhpco.org/hospice-statistics-research-press-room/facts-hospice-and-palliative-care

2. Cohen SR, Mount BM, Tomas JJN, Mount LF. Existential well-being is an important determinant of quality of life. Cancer 1996; 77:576-86.

3. Steinhauser KE, Christakis NA, Clipp EC, McNeilly M, McIntyre L, Tulsky JA. Factors considered important at the end of life by patients, family, physicians, and other care providers. JAMA 2000 Nov 15;284(19):2476-82.

4. Boston P, Bruce A, Schrieber R. Existential suffering in the palliative care setting: an integrated literature review. J Pain Symptom Manage. 2011 Mar;41(3):604-18. Epub 2010 Dec 8.

5. Puchalski C, Ferrell B, Virani R, Otis-Green S, Baird P, Bull J, Chochinov H, Handzo G, Nelson-Becker H, Prince-Paul M, Pugliese K, Sulmasy D. Improving the quality of spiritual care as a dimension of palliative care: the report of the Consensus Conference. J Palliat Med. 2009 Oct;12(10):885-904.

6. Clinical Practice Guidelines for Quality Palliative Care – 3rd Edition National Consensus Project
National Consensus Project for Quality Palliative Care. (2013). Clinical Practice Guidelines for Quality Palliative Care, Third Edition.
http://www.nationalconsensusproject.org/NCP_Clinical_Practice_Guidelines_3rd_Edition.pdf

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):*

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.med.unc.edu/pcare/resources/PEACE-Quality-Measures

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)*
Attachment **Attachment:** QNAV CPD - Sample-634425372974245559.pdf

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 measure specifications.

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 medical record includes documentation that the patient and/or caregiver was asked about spiritual/existential concerns within 5 days of the admission date.

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.)*
Numerator criteria must be met within 5 days of hospice admission date.

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.
Examples of a discussion may include asking about patient's need for spiritual or religious support, questions about the cause or meaning of illness or death. Other examples include discussion of God or a higher power related to illness, or offer of a spiritual resource including a chaplain. Discussion of spiritual or religious concerns may occur between patient and/or family and clergy or pastoral worker or patient and/or family and member of the interdisciplinary team.

This item is meant to capture evidence of discussion and communication. Therefore, documentation of patient's religious or spiritual affiliation by itself does not count for inclusion in numerator.

Data are collected via chart review. Criteria are:
1) evidence of a discussion about spiritual/religious concerns, or
2) evidence that the patient, and/or family declined to engage in a conversation on this topic.

Evidence may be found in the initial screening/assessment, comprehensive assessment, update assessments within 5 days of admission to hospice, visit notes documented by any member of the team, and/or the spiritual care assessment.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Seriously ill patients 18 years of age or older enrolled in hospice.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, 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)*
This quality measure is intended for patients with serious illness who are enrolled in hospice care. Conditions may include, but are not limited to: cancer, heart disease, pulmonary disease, dementia and other progressive neurodegenerative diseases, stroke, HIV/AIDS, and advanced renal or hepatic failure.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Testing has only been done with the adult population; thus patients younger than 18 are excluded.

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)
N/A

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)*
N/A

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*)

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.)*
Step 1- Identify all patients with serious, life-limiting illness who were discharged from hospice care during the designated reporting period.

Step 2- Exclude patients who are less than 18 years of age.

Step 3- Identify patients with documented discussion of spiritual/religious concerns or documentation that the patient/family did not want to discuss spiritual/religious concerns.

Quality measure = Numerator: Patients with documented discussion or who responded they did not want to discuss in Step 3 / Denominator: patients in Step 1 – Patients excluded in Step 2

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.
N/A

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.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Hospice: For this measure, there is only one item on the HIS that can include missing data – F3000B. If missing, these items are coded as dashes and analyzed as though the process did not happen.

In order to assess how these missing data impact the validity of the measure, we conducted the patient stay- and hospice-level analyses. For the patient stay-level analysis, we calculated the number and percentage of eligible patient stays for which the HIS-Admission records included a dash for this item. The overall rate of missing data ranged from 0.02 percent to 0.08 percent between October 2014 and September 2015. For the hospice-level analysis we calculated each hospice's percent of eligible admissions that included missing data for F3000B. Over 95% of hospices did not have any admissions with missing data for these items.
Overall, we found that only a very small number of admission records for the eligible stays did not include data, i.e. coded as dash, on F3000B. And a vast majority of hospices did not have any missing data for these items. These results indicate that missing data for these items should not have a negative impact on the validity of the QM.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record

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.
The Hospice Item Set (HIS) is the data source used to calculate the quality measure.

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 in attached appendix at A.1

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)*
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
1647_MeasureTesting_MSF5.0_Data.doc,1647_MeasureTestingAttachment_2.26.16.docx

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 1647	**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)***:**

CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1647 Beliefs/values addressed.

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.

Inter-rater reliability was conducted on this measure in two independent studies, the PEACE Project and the AIM Project.

The PEACE Project assessed inter-rater reliability using two research nurse abstractors who independently recorded quality measures data on a random subset of 20 seriously ill patients. Abstractors used the pre-defined operational definitions and a structured chart abstraction tool to record numerator and denominator data separately. Inter-rater reliability between the two abstractors was assessed using kappa statistics. The nurse abstractors achieved excellent inter-rater reliability for this measure with Kappa=1.0

The AIM Project conducted reliability on the entire data collection tool used in the AIM Project on which the measures are based. Inter-rater reliability between IPRO's medical record abstractor "the gold standard" and each agency's abstractor

was calculated using alpha=0.05 and power=0.8 and a preset value of kappa as 0.8. A sample size of 10 clinical records per agency was required to detect a kappa test statistic of 0.8 or greater. A convenience sample of clinical records from discharged patients who met the inclusion criteria was utilized.

We used percent agreement to test the reliability for dates and we conducted a kappa test on all categorical variables. Responses that had the same value in the quality measure calculations were collapsed into one value when appropriate (e.g., no, not documented, and unable to determine).

Inter-rater reliability was assessed between "the gold standard" abstractor and each agency's abstractor. Data from all 10 records were pooled and each agency was analyzed against the gold standard.

The kappa test statistic for all categorical variables was 0.795 (95% CI 0.79-0.80) (agency range: .70-.90), indicating substantial agreement. There is no reason to believe that achieved reliability for the data items contained within this measure would be substantially different than other categorical items. In fact during structured interviews and evaluations site abstractors noted that abstraction of this item was easier to conduct than most other items.

Deyta, LLC has been capturing data for this measure in the Quality Navigator since December 2008. In addition to enabling individual hospices compare and trend their own performance, comparative benchmarking is available for the more than 100 hospice agencies participating in this measure. Patient-level data from 13,435 records was used for the testing for 2009 and 2010.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)***:**

Split-half reliability. Split-half reliability assesses the internal consistency of a QM in different samples by randomly dividing the patient stays within each hospice into two halves, and calculating correlation between the QM scores on the basis of the two randomly divided halves. In this analysis, we conducted a split-half reliability analysis on all facilities with 20 or more patient stays counted in the measure denominator, and used the Interclass Correlation (ICC) coefficients to measure the internal reliability. In general, the ICC coefficient varies between 0 and 1, where an ICC of 0 indicates no reliability and an ICC of 1 indicates perfect reliability.

Signal-to-noise analysis. If a measure is reliable, then true differences in provider performance should explain a substantial proportion of the variance in QM scores. We conducted an analysis of variance (ANOVA) to determine what proportion of total variance in the measure is attributable to differences among providers. A higher proportion indicates better reliability.

Stability analysis. Stability analysis describes the extent to which providers' performance assessed by a QM changes across time. We analyzed the change in facility scores between four consecutive quarters (Q4 2014 and Q1 2015, Q1 and Q2 2015, and Q2 and Q3 2015). The changes in facility scores are reported in standard deviations.

A test-retest bivariate correlation was used to assess the consistency of the measure from one period of time (2009) to another time (2010).

Please refer to 2a2.1 for a description of the analytic methods used for the AIM and PEACE data testing.

2a2.3 Testing Results *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**

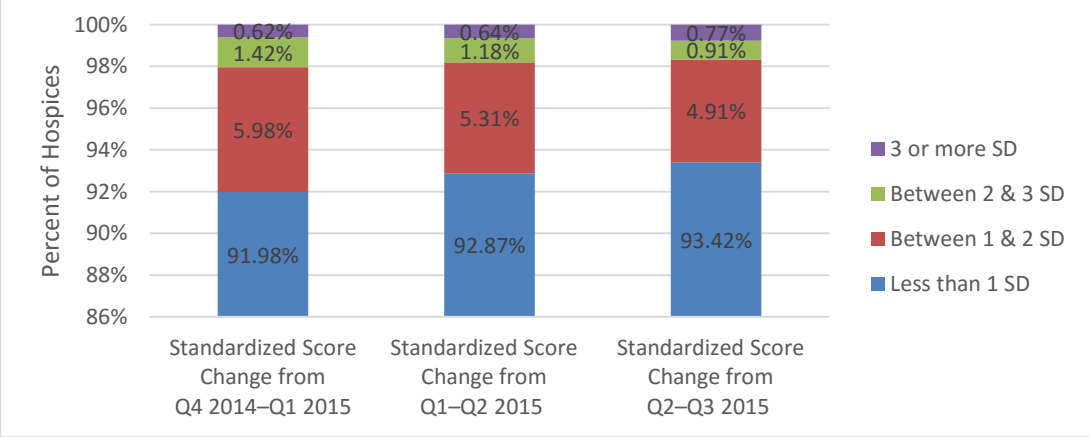
Split-half reliability: In general, the ICC coefficient varies between 0 and 1, where an ICC of 0 indicates no reliability and

an ICC of 1 indicates perfect reliability. The ICC coefficient for this measure is 0.94, indicating high internal reliability.

Signal-to-noise analysis: The analysis results found the signal-to-noise ratio to be 0.99, indicating that about 99% of the variance in this measure is because of differences among facilities. This proportion indicates strong reliability for this measure.

Stability analysis: The results of this analysis indicated that facility scores were very stable. Slightly more than 90% of facilities had a change in QM score of less than one standard deviation, indicating high stability of the QM. The number of facilities with a change in QM of less than one standard deviation increased across quarters from 92% to 93.4%, suggesting improved reliability across time. Less than one-tenth of facilities had a change in QM between one and two standard deviations. These results indicate a measure that is generally quite stable. Figure 1 illustrates the change in facility scores between the four consecutive quarters.

Figure 1
Standardized Score Change in QM Score from Q4 2014 to Q3 2015



Pearson Correlation: 0.026

Correlation is significant at the 0.01 level (2-tailed)

Significance (2-tailed): 0.004

Please refer to 2a2.1 for a description of the testing results from the AIM and PEACE data testing.

.

2b. VALIDITY. Validity, Testing, including all Threats to Validity: 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:**

This measure captures data on whether or not a discussion of spiritual care needs or concerns was documented, or if there was a refusal to discuss. In order for a spiritual care screening or assessment to be performed, a discussion between hospice staff and the patient/caregiver must occur.

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)***:**

CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1647 - Hospice and Palliative Care Beliefs/values addressed.

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.

Deyta, LLC has been capturing data for this measure in the Quality Navigator since December 2008. In addition to enabling individual hospices compare and trend their own performance, comparative benchmarking is available for the more than 100 hospice agencies participating in this measure. Patient-level data from 13,435 records was used for the testing for 2009 and 2010.

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

Correlations. Providers should perform similarly on QMs that reflect the quality of similar care processes. Thus, a common strategy used to evaluate validity of a QM is to examine the relative performance of the provider on QMs that are conceptually related, either due to some core competency manifested across a stay or for processes that occur in close proximity. We conducted nonparametric Spearman rank correlation analysis among all seven HQRP QMs, which address care processes around hospice admission that are clinically recommended or required in the hospice Conditions of Participation. We used nonparametric methods since the data are heavily skewed (most providers perform well, while a small proportion have lower scores). Since all of the seven current QMs measure desirable care assessment processes

upon hospice admission, higher positive Spearman correlation coefficients among these QMs will demonstrate higher validity of the QM set. Statistically significant correlations, indicated by p-values of the Spearman correlation coefficients less than 0.05, also support the validity of the QM set.

Missing data. Missing data represent a potential threat to the validity of an HIS item, which in turn may harm the validity of the QM that relies on the item. For this measure, there is only one item on the HIS that can include missing data, which is coded as dashes – F3000B. In order to assess how these missing data impact the validity of the measure, we conducted the patient stay- and hospice-level analyses.

PEACE Project: Construct validity was tested by comparing the PEACE quality measures for patients seen by specialty interdisciplinary palliative care consultants to those not receiving specialty palliative care services.

AIM Project: The AIM Project used the following methods to conduct face validity as follows: Following the first three quarters of data collection, participating agencies were each given quarterly reports for the measure based on the analysis of each agency's data and the aggregate project data. Agencies were then given opportunities to provide feedback (via written evaluations, conference call, best practice learning sessions, or individual correspondence) on whether they thought the data matched their actual practices. Agencies were asked to review results with their clinical staff and to review a subset of records and report to IPRO any discrepancies between the results and actual practice. Based on this feedback, revisions of the data abstraction tool and data dictionary were made and presented to the agencies to determine accuracy, feasibility, and to be sure the questions/items/answers represented actual practice. Additionally extensive feedback was sought from both the Hospice AIM Technical Advisory Panel and the Palliative Care Technical Expert Panel.

Because the Quality Navigator tool uses retrospective data collection approach and is the first hospice quality improvement instrument developed for data collection on this measure, we only conducted face validity testing of the measure. Based on discussions with participants in this measure, the agencies are able to capture data for this measure and indicate if a discussion of spiritual care needs was documented in the chart.

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)***:**

Correlations: Table 1 presents the Spearman correlation coefficients. The p-values for all the Spearman correlation coefficients are significant ($p < 0.001$). The significant positive correlations between every pair of QMs indicate that high-performing hospices in one area also provided high-quality care in other areas at hospice admission. Overall, the correlations between the QMs are low to moderate. The clustering of QM scores, skewed distributions, and low variability across all seven QMs may affect the level of correlations between QMs.

Table 1
Correlation of Hospice QMs, Percentile Ranking

Quality measure	NQF #1647 (modified) Beliefs/Values Addressed	NQF #1634 Pain Screening	NQF #1637 Pain Assessment	NQF #1639 Dyspnea Screening	NQF #1638 Dyspnea Treatment	NQF #1617 Bowel Regimen
NQF #1641 Treatment Preferences	0.64***	0.37***	0.21***	0.44***	0.28***	0.25***
NQF #1647 (modified) Beliefs/Values Addressed		0.42***	0.17***	0.43***	0.31***	0.25***

NQF #1634 Pain Screening	0.17***	0.41***	0.28***	0.15***
NQF #1637 Pain Assessment		0.17***	0.08***	0.20***
NQF #1639 Dyspnea Screening			0.30***	0.25***
NQF #1638 Dyspnea Treatment				0.34***

NOTE: The correlation is on the basis of each hospice's percentile ranking on the QMs.

*** indicates significant correlation at $p < 0.001$

Overall, this measure has significant positive correlations with the other QMs, indicating hospices providing higher-quality care in this area also performed better in other areas at hospice admission. The QM having the strongest correlation with this measure is NQF #1641, Treatment Preferences ($p = 0.64$). We expect the strong correlation between these two QMs because they both address the competency of the hospice to solicit the patient's preferences.

Missing data: For the patient stay-level analysis, we calculated the number and percentage of eligible patient stays for which the HIS-Admission records included a dash for each of these three items. The overall rate of missing data ranged from 0.02 percent to 0.08 percent between October 2014 and September 2015. For the hospice-level analysis we calculated each hospice's percent of eligible admissions that included missing data for F3000B. Over 95% of hospices did not have any admissions with missing data for these items.

Overall, we found that only a very small number of admission records for the eligible stays did not include data, i.e. coded as dash, on F3000B. And a vast majority of hospices did not have any missing data for these items. These results indicate that missing data for these items should not have a negative impact on the validity of the QM.

PEACE Project results for Construct Validity: Hypothesizing that specialty palliative care providers will be better trained to screen for spiritual concerns, data demonstrates this quality measure is more often met for patients with (64%) vs. without (40%, $p<0.01$) specialty palliative care added.

AIM Project supports that the items contained within this measure were valid.

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)***:**

CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1647 - Hospice and Palliative Care Beliefs/values addressed.

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.

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

We examined the effect of excluding patients less than 18 years of age. At the patient-stay level we examined the proportion of patient stays that are excluded from the denominator. At the hospice level we examined the distribution of hospice-level exclusion rates. At the QM level, we examined the distribution of hospice QM scores with and without applying the age exclusion.

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

From October 2014 – September 2015, very few stays (approximately 0.3%) were excluded from the denominator because the patients were less than 18 years old at hospice admission. Among these patients, 88.8% had documented discussion of spiritual/religious concerns within 2 days of admission, compared to 92% among patients who were at least 18 years old. This suggests that non-pediatric patients are slightly more likely to receive the desired care process compared to pediatric patients.

Across hospices, the mean proportion of stays for patients < 18 years of age was 0.36% and the median was 0%; over 75% of hospices had no pediatric patients and in more than 99% of hospices, less than 5% of patients were pediatric patients. These findings suggest that the age exclusion only affects a small proportion of hospices nationwide, and that for most hospices with pediatric patients, the impact of the exclusion on denominator size is minimal.

The impact of the age exclusion on the QM score distribution is relatively small. Without the exclusion, the mean and median scores among hospices with at least one patient < 18 years of age were 93.4% and 97.4%, respectively. With the age exclusion, the mean and median scores were 93.4% and 97.5%, respectively.

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)*:

N/A - This measure is not risk adjusted

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

N/A

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):

N/A

2b4.4 **If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment:** This measure applies to all hospice patients. There are not any variables that would impact whether or not the practice/process should be performed.

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)*:

CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the Hospice Quality Reporting Program (HQRP) in the FY 2014 Hospice Wage Index final rule (78 FR 48234–48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014. The HIS collects data to calculate seven quality measures (QMs)—six are QMs endorsed by the National Quality Forum (NQF) and one is a modified NQF-endorsed QM. One of these QMs is #1647 – Beliefs/values addressed.

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.

Deyta, LLC has been capturing data for this measure in the Quality Navigator since December 2008. In addition to enabling individual hospices compare and trend their own performance, comparative benchmarking is available for the more than 100 hospice agencies participating in this measure. Patient-level data from 13,435 records was used for the testing for 2009 and 2010. Additional data continues to be collected throughout 2011, but has not been included in testing or comparison at this time.

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

Confidence interval analysis. We examined proportions of hospices with the QM scores that are significantly different from the national hospice-level mean. If a high proportion of hospices have a measure score significantly different from the mean, the QM can identify facilities with different levels of performance. For this analysis, statistical significance was determined using 95 percent confidence intervals: a hospice's QM score was significantly different from the national mean if the national mean was not included within the hospice's 95 percent confidence interval. High-performing facilities should have scores that are significantly below average, and low-performing facilities should be significantly above average.

By simple percentile rankings of results for agencies, variation in performance is clearly evident. (Refer to 2b5.3. Results below)

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)*:

The mean score for this QM was 92.2% with a range from 0% to 100%, the median was 97.7%, the interquartile range was 9.2, and the standard deviation was 13.6. For this QM, 29.4% of hospices had perfect scores and 23.4% of hospices

scored below 90%.

Scores by decile:

10th percentile 78.6%

25th percentile 90.8%

Median 97.7%

75th percentile 100%

90th percentile 100%

Across all hospices, 61.8% had a QM score that is significantly different than the national mean. Hospices were more likely to report scores above the national mean than below the national mean (43% vs 18.8%, respectively, overall). The QM is able to identify those hospices that are performing well (higher than the national mean) and those that are performing less well (lower than the national mean).

2009	2010	
Mean	68.6%	63.7%
10th percentile	20.0%	10.6%
25th percentile	47.0%	38.2%
Median	78.2%	73.6%
75th percentile	92.15%	90.9%
90th percentile	100%	97.0%

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)***:**

No other known data sources are available for this measure.

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

N/A

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)***:**

N/A

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 - This

measure is not stratified.

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

N/A

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)*

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

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.

While this measure can be built such that it becomes part of a proprietary tool or documentation system, it is not a proprietary measure. Any hospice could easily incorporate this measure into their QAPI program. The challenge identified by current participants relates to the ability to query this information out of their EHR. Other participants have been easily able to incorporate this single measure into their already existing chart review/audit processes.

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)*.

None

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)
Public Reporting	Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Hospice Item Set-Hospice Quality Reporting Program https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Hospice-Quality-Reporting/Hospice-Item-Set-HIS.html

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

• Name of program and sponsor: Hospice: HIS/CMS
• Purpose: Hospice: Quality reporting for hospice
• Geographic area and number and percentage of accountable entities and patients included: United States and all accountable entities and 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**

a. Due to insufficient longitudinal data, we are currently unable to discuss progress.

b. Analyses encompassed 3,922 hospice organizations and 1,218,786 patient stays across the United States.

Region:
South: 39.3%
West: 25.1%
Midwest: 23.1%
Northeast: 11.3%
Territories: 0.94%

Unknown: 0.25%

Urban/rural status:
Urban: 75.8%
Rural: 23.9%

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.**
We only have one year of data to report, which is not enough to show trends over time.

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.**
None

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.)
No known competing measures exist.

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 **Attachment:** Appendix_A.1_NQF_1647.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** University of North Carolina-Chapel Hill
Co.2 **Point of Contact:** Laura, Hanson, lhanson@med.unc.edu, 919-843-9096-
Co.3 **Measure Developer if different from Measure Steward:** University of North Carolina-Chapel Hill
Co.4 **Point of Contact:** Laura, Hanson, lhanson@med.unc.edu, 919-843-9096-

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.
As this and other measures were being developed, we sought input from several academic researchers, hospice providers and other representatives in the industry.

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2008
Ad.3 **Month and Year of most recent revision:** 02, 2016
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?** 02, 2016

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

Ad.8 **Additional Information/Comments:** Data is also available on this quality measure from Deyta LLC, a quality measurement and analytics corporation, derived from their Quality Navigator data system for hospices. Deyta's Quality Navigator offers clients the ability to capture, track and trend data on the intensity of spiritual distress – a measurable outcome related to spiritual care. Deyta conducted an analysis of a subset of their 2009 and 2010 records that were used for the initial measure submission to determine whether patients who had documentation of a spiritual discussion were more likely to improve their spiritual distress scores (distress score measured as no distress, mild, moderate, severe and improvement defined as a decrease of at least one level in distress score) than patients who did not have documentation of spiritual discussion. Data were retrospectively abstracted by trained personnel at each of the agencies and submitted to Deyta's system for trending and reporting. Patient records were abstracted from agencies that utilize both spiritual care measures. Records were excluded due to patients reporting that they were not experiencing any spiritual distress or due to missing data (patient was unable to report at least one spiritual distress score while on service). A total of 652 records were included in the analysis.

Findings revealed that a greater proportion of patients who had documentation of a spiritual discussion (63%) showed improvement in their spiritual distress score than patients who did not have a documented spiritual discussion (48%). These findings help support the importance of the spiritual discussion and the existence of a relationship between the occurrence of the spiritual discussion and reduction in spiritual distress scores.

Documentation of spiritual discussion; Improvement in spiritual distress score: 330 (63%)
Documentation of discussion; No improvement in distress score: 194 (37%)
No documentation of discussion; Improvement in distress score: 62 (48%)
No documentation of discussion; No improvement in distress score: 62 (52%)

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 #: 1617
Measure Title: Patients Treated with an Opioid who are Given a Bowel Regimen
Measure Steward: RAND Corporation/UCLA
Brief Description of Measure: Percentage of vulnerable adults treated with an opioid that are offered/prescribed a bowel regimen or documentation of why this was not needed
Developer Rationale: Reduction in opioid-induced constipation has the potential to result in improved opioid medication compliance, reduction in patient discomfort, and improved quality of life.

Numerator Statement: Patients from the denominator that are given a bowel regimen or there is documentation as to why this was not needed
Denominator Statement: Vulnerable adults who are given a prescription for an opioid
Denominator Exclusions: Non-hospice outpatients who are already taking an opioid at the time of the study period opioid prescription

Measure Type: Process
Data Source: Paper Medical Records
Level of Analysis: Clinician : Group/Practice, Clinician : Individual, Facility, Health Plan

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 or **Summary of prior review in [year]**

- The developer provided a rationale for the relationship of this process of care (bowel regimen for opioid

patients) and patient outcomes by linking the reduction in opioid-induced constipation to potential improvements in medication compliance, reduction in patient discomfort, and improved quality of life.

- The developer cited selected 2 clinical practice guidelines to support the measure. These include:
  - 2002 AGS Panel on Persistent Pain in Older Persons: *"A prophylactic bowel regimen should be initiated at the commencement of persistent opiate therapy."* (***evidence grade 1A***)
  - 2009 American Pain Society-American Academy of Pain Medicine guideline. *Recommendation 8.1: Clinicians should anticipate, identify, and treat common opioid-associated adverse effects (**strong recommendation, moderate-quality evidence**).*
- The developer also cited a systematic review by Lorenz, et al. (2007) and an article by Schenck, et al. (2010).

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:**

Exception to evidence – N/A

Guidance from the Evidence Algorithm

Process measure based on systematic review and grading of evidence (Box 3) → Summary of quantity, quality, and consistency of data not provided (Box 4) → Grade indicates high and moderate quality evidence and strong recommendation → Moderate

Questions for the Committee:

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

- The developer provided updated information on gap from two relatively recent articles:
  - Hanson, 2012: Data from 254 seriously ill patients hospitalized between December 2007 and March 2010 – 44%
  - Walling, 2013: 2008 national sample of veterans with advanced cancer – 52% in 472 outpatients; 71% in 303 inpatients

Disparities

- No data on disparities were provided.

Questions for the Committee:

- *Although data from two newer studies were provided, the results are more than 5 years old. Do these results demonstrate that there is still a gap in care that warrants a national performance measure?*
- *This measure is included in the CMS Hospice Reporting Program and is collected through the Hospice Item Set. Do any members know the current performance rate in the hospice setting?*
- *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 current data on performance is provided.

Committee pre-evaluation comments

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

1a.

* This measure evaluates an important healthcare process, prescribing a bowel regimen (including medication or diet change) to support patients at high risk of constipation due to a new start of an opioid. Timely prevention of constipation is important to support opioid medication compliance and promote quality of life. Constipation in vulnerable adults is common, and a significant performance gap persists.
* Having read through the evidence originally provided by the developer and having sought guidance from the Evidence Algorithm, I do not feel another discussion is necessary nor that we need to vote again. In regard to a choice of Pass or No Pass, I choose pass.
*This is a process measure presented for maintenance. The developer presents the rationale that the reduction in opioid induced constipation has the potential to result in improved opioid medication compliance and reduction in patient discomfort and improved quality of life. The evidence to support the measure is direct and has received systematic review. evidence has been graded. Two clinical practice guidelines were cited to support the measure of high and moderate quality and strong recommendation. The evidence algorithm gives a graded rating of moderate.
No new evidence since last review is reported. I support that the evidence basis has not changed and there is no need for repeat discussion and vote by the committee on evidence

1b.

* Two studies show 50% or less performance on this measure from consultative palliative care. Data from the Hospice Item Set are still not released.

Data on disparities is not presented.
* Although there is plenty of evidence to argue that the disparity begins with who receives opioids for pain, I am not aware of evidence that disparities exist in the area of patients receiving opioids and not receiving a bowel regimen. The data provided was not new and so we cannot say for certain that there are documented disparities; therefore, I rate the performance gap as insufficient
*Performance data was supplied information on gap from two relatively recent articles. No data on disparities was provided. I am not aware of any disparities existing in this area of healthcare. as no current data on performance was provided the preliminary rating for opportunity for improvement was rated as "insufficient"
This measure is included in the CMS Hospice Reporting Program and is collected throught the hospice Item Set. current performance rate in the hospice setting will be accessed and provided at the inperson meeting of the SC in May of 2016

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): Paper medical records

Specifications:

- This measure is specified for the following levels of analysis: individual clinicians, clinician group/practices, facilities, and health plans. A higher score indicates better quality.
- Care settings for the measure include clinician offices/clinics, hospices, and hospitals.
- The numerator includes patients given a bowel regimen (offer/prescription of a laxative, stool softener, or high-fiber supplement or diet within 24 hours of opioid prescription OR those with documentation as to why it wasn't needed.
- The denominator includes vulnerable adults who are older than 17 who have been prescribed an opioid as an outpatient, hospital inpatient, or hospice patient; "vulnerable" patients 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
    - Receiving hospice care
- Exclusions include non-hospice outpatients who are already taking an opioid at the time of the measurement period opioid prescription.
- A calculation algorithm is provided.
- Data are collected via a medical record abstraction tool, although this tool was not provided.
- This measure is not risk-adjusted.

Questions for the Committee :

- *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:

- Previous reliability testing included inter-rater reliability testing of data from inpatient, outpatient, and hospice patients.

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 three reliability studies, but noted the methodology of only two (inter-rater reliability and data "re-abstractions"). The developer provided details about the sample size used in the testing for only two of the studies (ACOVE3: 47 inpatient decedents; ASSIST: 39 inpatient and outpatient decedents). The developer provided information about which data elements were tested for only two of the three studies (ACOVE3 and ASSIST).
- In the PEACE study cited, reliability of the data elements was assessed using data obtained from the medical charts for 20 seriously ill patients without specialty palliative care admitted to an acute care hospital. Inter-rater reliability between two nurse abstractors was calculated. This is an appropriate method for demonstrating data element reliability. (This testing was described for other measures being evaluated for re-endorsement).

Results of reliability testing:

- PEACE study
    - One kappa value reported: 0.86.
    - 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 1.0 means that the raters agreed 100% 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.
    - NQF guidance indicates that data element testing should be conducted for all critical data elements, although at minimum, results about the numerator, denominator, and exclusions should be provided. Only a single kappa value was reported. Although not explicitly stated, we assume that the 20 charts that were abstracted included only seriously ill patients who received specialty palliative care in an acute hospital setting and whose length of stay was at least one day, and thus this kappa value applies to the numerator.
- ACOVE3 study
    - Denominator percent agreement=98%
    - Numerator kappa=0.64
- ASSIST study
    - Denominator kappa=0.87
    - Numerator kappa=0.86

Guidance from the Reliability Algorithm

Precise specifications (Box 1) → Empirical reliability testing with measure as specified (Box 2) → Score-level testing not conducted (Box 4) → Data elements tested (Box 8) → 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 moderate 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**

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:

- In the prior evaluation, the developer cited several face validity assessments of the measure.

Describe any updates to validity testing:

- No updated testing was provided.

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 developer states "*Validity of the process-outcome link was explicitly evaluated by the ACOVE, ACOVE3, ASSIST, and PEACE expert panels that reviewed the relevant literature and used a modified Delphi panel of voting on the validity of the measure*".

Validity testing results:

- The developer did not provide actual results of the various face validity assessments.

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:

- No information on exclusions was presented.

Questions for the Committee:

- *Are the exclusions (non-hospice outpatients who are already taking an opioid at the time of the study period opioid prescription) consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

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)*:

- No information on meaningful differences was presented.

Question for the Committee:

- *Does the Committee have any evidence concerning the ability of this measure to identify meaningful differences in quality of care for the various entities assessed by the measure and in the various settings specified for the measure?*

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

2b7. Missing Data

- No information on missing data was presented (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 exclusions, 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: Developers did not provide results of the face validity assessments and did not demonstrate that the measure is able to identify meaningful differences between providers. 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

* The major issue here, as also expressed through public comment, is the narrowness of the denominator to "vulnerable elders" who are defined by the measure stewards as those receiving inpatient care, hospice care, near the end of life (<6 months) or have advanced cancer or over age 75. Truly, ALL patients with a serious illness on longacting or regular use of short acting opioids should be given a bowel regimen, regardless of system of care (e.g. hospice), prognosis (e.g. <6 months) or disease (advanced cancer). These limitations in eligibility have largely been solved by the expansion of the scope of palliative and end of life care to include all serious illnesses, independent of stage or prognosis. Further, the use of as-needed or occasional opioids for pain may not warrant regular use of a bowel regimen. Truly, longacting or regular use of short acting is what is the focus here.
* I believe that extracting the data from paper medical records is likely problematic due to the unorganized nature of the document and the move towards EHRs. Also, the developers did not report on their extraction tool. The calculation algorithm is clear but I do not see the implementation of the tool being easily done due to the data source being paper medical records.
*Data elements in the measure are clearly defined. Measure data was extracted utilizing paper medical records, implementation of the EHR will make abstraction and data collection more feasible. This measure can be consistently implemented in most settings but will be more challenging in setting where paper records continue to be utilized such as LTC, or in certain outpatient settings. The calculation algorithm is clear

Validity – Specifications

* Expansion of measure eligibility to all patients with serious illness receiving an opioid should be explored.
*Specifications are consistent with the evidence
specifications are consistent with the values that the target audience finds meaningful

Reliability – Testing

* Newer reliability testing was not presented. Older, but recent testing demonstrated kappa scores near 90% in representative populations.
* I have concerns about the size of the test samples - very small - and so would hesitate to generalize for widespread implementation. I do think that the results demonstrate sufficient reliability so differences can be identified. I do not think we need to re-discuss and re-vote on reliability. Rating is Moderate.
*Testing was completed at the data element level but not the score
There reliability studies were cited. the PEACE study was utilized in reliability testing and demonstrated denominator percent agreement of 98%.
Guidance from the reliability algorithm provide a graded score of moderate
Although test samples not entirely adequate to generalize for widespread implementation, in my opinion there is sufficient reliability to identify differences in performance. prior testing demonstrated moderate reliability and the committee does not need to re-discuss or vote on reliability

Validity Testing

* Validity testing is older but adequate. Three separate validity evaluations demonstrate this.
* Face and expert panel validity are weak tests and those were the method used. I think that score from the measure is an indicator of quality. I also think the Committee needs to do a re-vote.
*face value validity and score were tested. Per the developer, 'validity of the process-outcome link was explicitly evaluated by the ACOVE and ACOVE3, ASSIST and PEACE expert panels that reviewed the relevant literature and used a modified Delphi panel of voting on the validity of the measure.
Results demonstrate sufficient validity so that conclusion about quality can be made and the score from this measure as specified is an indicator f quality. Theree is no need to re-vote on validity assuming that threats to validitywere adequately assessed

Threats to Validity
* Exclusions - inappropriately narrow

Adjustment - none needed

Meaningful differences - data is not presented. The benchmark for this measure is truly 100% performance, so anything less would be considered a meaningful gap.

Comparability of scores - N/A

Missing data: The measure is an either/or. There should not be issues with missing data.
* No missing data was reported but it is likely that in situations where data is gathered from paper records, there will be missing data, which I feel would pose a threat to the measure's validity. Overall, I would say that the rating for validity is insufficient.
*No information on missing data was presented.
the validity algorithm provides a preliminary rating of "insufficient because empirical validity testing was not conducted and results of face validity were not provided. if this information is provided the measure is eligible for a moderate rating.

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, abstracted from a hospice or palliative care clinical record. The developer noted that exclusions for the measure are not yet available in electronic medical record applications. Data collection for all measures is achieved through use of the Hospice Item Set (HIS). It is unclear if the HIS is automatically populated by the EHR.
- Cost estimates are not available. According to the Guidance Manual for Completion of the Hospice Item Set, completing an entire response for any one admission takes 20 minutes (though seven performance measures, of which this measure is only one, are collected as part of this process).
- In the 2012 endorsement of the measure, the Palliative and End-of-Life Care Steering Committee described data elements as ”easily collected”.

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

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

Committee pre-evaluation comments
Criteria 3: Feasibility

* The measurement specifications highlight paper-based data collection. Collection in EHR's would require e-specifications that outline which structured data fields and answer choices would sufficiently report for the measure.
* I think the data elements are likely routinely gathered and used, and easily gotten through the Hospice Item Set. That said, it is not clear whether electronic records are included in the Set and that does make a difference. I would assume that the data elements are available in electronic form. My rating of feasibility is moderate.
*The required data elements are routinely generated and used during care delivery and are available in the EHR . preliminary rating for feasibility was 'moderated"

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
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 Hospice Quality Reporting Program (HQRP) is a pay-for-reporting program. NQF#1617 is part of the Hospice Item Set (HIS), a report submitted to CMS for every admission and discharge from a Medicare-certified hospice. The HIS is used to calculate seven quality measures, including NQF#1617.

Improvement results

- Longitudinal data for these measures for the hospice setting is 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 2016, the Measures Application Partnership Post-Acute Care and Long-Term Care workgroup supported the continued development of a composite measure of Hospice Item Set measures, which includes NQF #1617.
- In February 2013, the Measure Application Partnership supported the measure for inclusion in the Hospice Quality Reporting Program, and the Physician Quality Reporting System (PQRS) to address a core concept (i.e., pain assessment) not addressed in the programs' measure sets. Public comments from American Academy of Hospice and Palliative Medicine (AAHMP), Center to Advance Palliative Care (CAPC), and National Coalition for Hospice and Palliative Care (NCHPC) supported the inclusion of this measure in the PQRS. CAPC and NCHPC suggested that the measure should be expanded beyond hospice patients. The measure is not currently included in the PQRS program, but it is included in the Hospice Quality Reporting Program.

***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**

Committee pre-evaluation comments

Criteria 4: Usability and Use

* Currently not publicly reported outside of Hospice reporting. That data is not yet available. As a process measure, there is no assurance that patients are taking the medications prescribed (or can fill them or afford them).
* The performance results could be used to further the goal of high-quality efficient healthcare. The research is clear about the issue of opioids and the potential for constipation and, worse, impactions of the bowels. This measure has the potential to provide data on 1) whether healthcare professionals are giving sufficient attention to this issue and 2) that they are doing so for all vulnerable groups (gets to the issue of disparities. There are no potential unintended consequences reported at this point and I cannot foresee any so I feel the benefits will outweigh any unintended consequences. I rate usability and use as Moderate.
*Currently not being publicly reported.

There is current accountability use and planned in the future.
the HQRS is a pay for reporting program. This measure is part of the hospice Item Set, a report submitted to CMS for every admission and discharge from a Medicare- certified hospice. the HIS is used to calculate seven quality measures including this one.
there are no unexpected finding or unintended consequences identified.
This measure is extremely valuable in furthering the goal of hih quality effieicne healthcare because of the ramifications of not address the issue which this measure assesses and the measure maintains diligence in the hospice area and opportunity to do so in other settings

Criterion 5: Related and Competing Measures

Related measures

- The definition of 'vulnerable adults' is harmonized with another RAND measure (1626: Patients Admitted to ICU who Have Care Preferences Documented)

Harmonization

- N/A

Pre-meeting public and member comments

Name: Michele Galioto, RN, MSN
Organization: Oncology Nursing Society
Comment: Is the term "vulnerable adult" the best descriptor?

Excluding non-hospice patients already taking an opioid at the time of study would likely exclude the majority of people with cancer; would be in favor of removing this exclusion.

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.

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 1617 **NQF Project:** Palliative Care and End-of-Life Care

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)***:**

Measure focus is on the process of bowel prophylaxis initiated at the time of opioid prescription with the intent of improved outcomes with regard to opioid compliance, pain control, lower incidence of constipation, and consequential improved quality of life.

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

Clinical Practice Guideline

Selected individual studies (rather than entire body of evidence)

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, the clinical effect of the care process on opiate use is clear, as reflected in clinical guidelines recommending constipation prophylaxis.

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 <u>across Studies</u>** *(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? Yes

1c.10 **If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias:** AGS Panel on Persistent Pain in Older Persons

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

1c.12 **If other, identify and describe the grading scale with definitions:**

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

1c.14 **Summary of Controversy/Contradictory Evidence:**

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

See also 1a.4

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.

Schenck AP, Rokoske FS, Durham DD, et al. The PEACE project: identification of quality measures for hospice and palliative care. J Pall Care 2010;13(12):1451-1459

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

Constipation is one of the most common opioid-related adverse effects. Most patients develop some degree of constipation after opioid initiation or dose increases, and resolution of constipating effects of opioids often does not occur with continued exposure. In older adults or other patients with additional reasons to develop constipation, we recommend routinely considering initiation of a bowel regimen before the development of constipation. Though most evidence is anecdotal, bowel regimens including increased fluid and fiber intake, stool softeners, and laxatives are often effective. There is insufficient evidence to recommend oral opioid antagonists to prevent or treat opioid-induced bowel dysfunction in persons with CNCP, though randomized trials suggest some potential benefits over placebo. (Chou 2009, page 121)

A prophylactic bowel regimen should be initiated at the commencement of persistent opiate therapy. (AGS, page S217)

1c.17 Clinical Practice Guideline Citation: Chou R, Fanciullo GJ, Fine PG, Adler JA, Ballantyne JC, Davies P, Donovan MI,

Fishbain DA, Foley KM, Fudin J, Gilson AM, Kelter A, Mauskop A, O´Connor PG,

Passik SD, Pasternak GW, Portenoy RK, Rich BA, Roberts RG, Todd KH, Miaskowski C; American Pain Society-American Academy of Pain Medicine Opioids Guidelines Panel. Clinical guidelines for the use of chronic opioid therapy in chronic noncancer pain. J Pain. 2009;10:113-30.

AGS Panel on Persistent Pain in Older Persons. The management of persistent pain in older persons. J Am Geriatr Soc. 2002 Jun;50(6 Suppl):S205-24.

1c.18 National Guideline Clearinghouse or other URL:

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? Yes

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

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

1c.22 If other, identify and describe the grading scale with definitions:

1c.23 Grade Assigned to the Recommendation: 1A

1c.24 Rationale for Using this Guideline Over Others: No contradictory guidelines

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

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
1617_Evidence_MSF5.0_Data-635278463467243641.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)*
Reduction in opioid-induced constipation has the potential to result in improved opioid medication compliance, reduction in patient discomfort, and improved quality of life.

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.*
Edits in brackets []

N, % measure performance
Assessing Care of Vulnerable Elders (ACOVE)(Wenger 2003): Community dwelling vulnerable elders. N=46, 0%

ACOVE3 (Walling 2010): Inpatients who died during admission. N=460, 61%

Assessing Symptoms Side Effects and Indicator of Supportive Treatment (ASSIST) (Dy 2011): Cancer Center and hospitalized patients with advanced cancer. N=39, 51%

ACOVE3/Tufts (unpublished data): Community dwelling elders >74 years old. N=48, 17%

[Seriously ill, hospitalized patients (Hanson 2012): N=254, 44%]

[National sample of veterans with advanced cancer (Walling 2013): N=472 outpatients, 52%; N=303 inpatients, 71%]

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.**
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 Pall Med 2011;14(4):451-457.

[Hanson LC, Rowe C, Wessell K, Caprio A, Winzelberg G, Beyea A, Bernard SA. Measuring palliative care quality for seriously ill hospitalized patients. J Palliat Med. 2012;15:798-804.]

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

[Walling AM, Tisnado D, Asch SM, Malin JM, Pantoja P, Dy SM, Ettner SL, Zisser AP, Schreibeis-Baum H, Lee M, Lorenz KA. The quality of supportive cancer care in the veterans affairs health system and targets for improvement. JAMA Intern Med. 2013;173:2071-9.]

Wenger NS, Solomon DH, Roth CP, et al. The quality of medical care provided to vulnerable community-dwelling older patients. Arch Intern Med 2003;139:740-47.

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.*
None available

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 of 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**
Affects large numbers, A leading cause of morbidity/mortality, 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.**
Opioids are commonly used in the management of moderate to severe pain, and constipation is a common adverse effect. (Myotoku 2010; Tuteja 2010; Pappagallo 2001) A systematic review evaluating the extent and management of opioid-related side effects in both cancer and non-cancer patients indicated that tolerance is not developed to opioid-induced constipation and confirmed the need for prophylaxis. (McNicol 2003) Risk of constipation is further aggravated by immobility and dehydration in older people with pain. The American Pain Society and American Geriatrics Society as well as expert consensus opinion recognize the frequency of constipation with opioid use and the necessity for prophylactic therapy. (APS 2005; RANO 2002; AGS 2002; APS 2002; Weiner 2001; Davis 2003; Etzioni 2007; Dy 2008) A study of 194,017 emergency department visits made by 76,759 cancer patients in the final 6 months of life revealed that 3,392 visits were made for constipation. (Barbera 2010)

The denominator for this measure includes vulnerable adults as this was the population in which it was tested. However, the literature cites constipation as a significant problem in adults of all ages and including those taking opioids for chronic non-cancer pain as well. (Tuteja 2010) A Cochrane systematic review of 26 studies of patients at least 18 years old taking opioids for at least 6 months for non-cancer pain revealed gastrointestinal complaints (e.g., constipation, nausea, dyspepsia) as the most commonly reported side effect. (Noble 2010)

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
AGS Panel on Persistent Pain in Older Persons. The management of persistent pain in older persons. J Am Geriatr Soc 2002;50(6 Suppl):S205-24

American Pain Society (APS). Guideline for the management of cancer pain in adults and children. 2005

American Pain Society (APS). Guideline of the management of pain in osteoarthritis, rheumatoid arthritis, and juvenile chronic arthritis. 2002.

Barbera L, Taylor C, Dudgeon D. Why do patients with cancer visit the emergency department near the end of life? Can Med Assoc J 2010;182(6):563-569

Davis MP, Srivastava M. Demographics, assessment and management of pain in the elderly. Drugs Aging 2003;20(1):23-57

Dy SM, Asch SM, Naeim A, et al. Evidence-based standards for cancer pain. J Clin Oncol 2008;26(23):3879-3885

Etzioni S, Chodosh J, Ferrell BA, et al. Quality indicators for pain management in vulnerable elders. JAGS 2007;55:S403-S408

McNicol E, Horowicz-Mehler N, Fisk RA et al. Management of opioid side effects in cancer-related and chronic noncancer pain: a systematic review. J Pain 2003;4(5):231-56

Myotoku M, Nakanishi A, Kanematsu M, et al. Reduction in opioid side effects by prophylactic measures of palliative care team may result in improved quality of life. J Pall Care 2010;13(4):401-406

Noble M, Treadwell JR, Tregear SJ, et al. Long-term opioid treatment for chronic noncancer pain. Cochrane Database Sys Rev 2010;(1):CD006605

Pappagallo M. Incidence, prevalence, and management of opioid bowel dysfunction. Am J Surg 2001;182(5A Suppl):11s-8s

Registered Nurses Association of Ontario (RNAO). Assessment and management of pain. 2002. (Nursing Best Practice Guideline: Shaping the Future of Nursing)

Tuteja AK, Biskupiak J, Stoddard GJ, et al. Opioid-induced bowel disorders and narcotic bowel syndrome in patients with chronic non-cancer pain. Neurogastroenterol Motil 2010;22:424-e96

Weiner DK, Hanlon JT. Pain in nursing home residents: management strategies. Drugs Aging 2001;18(1):13-29

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, Gastrointestinal (GI)

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

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 from the denominator that are given a bowel regimen or there is documentation as to why this was not needed

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.)*
Within 24 hours of opioid prescription.

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.
Patients from the denominator given a bowel regimen (or one is already in place) defined as an offer/prescription of a laxative, stool softener, or high fiber supplement/diet OR documentation of why such a bowel regimen is not needed.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Vulnerable adults who are given a prescription for an opioid

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
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 >17 years old prescribed an opioid as:
- An inpatient
- A hospice patient (inpatient or outpatient)
- A non-hospice outpatient in patients who are not already taking an opioid

"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
- Patients receiving hospice care in any setting

Saliba D, Elliott M, Rubenstein LZ, et al. The vulnerable elders survey: a tool for identifying vulnerable older people in the community. J Amer Geriatr Soc 2001;48:1691-1699

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Non-hospice outpatients who are already taking an opioid at the time of the study period opioid prescription

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 are prescribed an opioid in the outpatient setting are excluded if they are NOT hospice patients AND at the time of the opioid prescription that occurred during the study period, they were already taking an opioid. This exclusion does NOT apply to inpatients or to hospice patients treated in any setting. Non-hospice outpatients who are prescribed an opioid who may have been on an opioid in the past, but are not taking an opioid at the time of the study period opioid prescription are NOT excluded.

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.)*
Note that edits placed in brackets []
1. Identify vulnerable adults with a prescription for an opioid. For inpatients, identify ALL patients with an order for [standing (not prn)] opioid treatment on admission or during the hospitalization. For hospice patients, identify ALL patients with an order for opioid treatment on admission or during the episode of hospice care. For outpatient non-hospice patients, identify patients with a "new" prescription for an opioid. "New" prescription for a non-hospice outpatient means that the patient is not already taking an opioid.

2. Include only patients who are vulnerable (age >74, VES-13 score >2, or poor prognosis/terminally ill, advanced cancer, patients receiving hospice care).

3. Look for documentation within 24 hours of opioid prescription for a prescription for a laxative, stool softener, or high fiber supplement/diet OR documentation as to why such a regimen was not needed.

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)*

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.
None

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.

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

N/A

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)*
Clinician : Group/Practice, Clinician : Individual, Facility, Health Plan

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic, Hospice, 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
1617_MeasureTesting_MSF5.0_Data-635278463467243641.doc,1617_MeasureTesting_MSF5.0_Data-635278463467243641-635948597859744862.doc

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=460), 47 reliability re-abstractions: Eligibility 98% agreement (unable to calculate kappa); specified care kappa=0.64.

ASSIST (Dy 2011) inpatient/outpatient decedents (n=39): Overall eligibility kappa=0.87; overall specified care kappa=0.86

PEACE (personal communication from Dr. Laura Hanson): Hospice and palliative care patients. Inter-rater reliability kappa=0.86

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 Pall Med 2011;14(4):451-457

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

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

2b. VALIDITY. Validity, Testing, including all Threats to Validity: 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:**

Populations tested reflect similar populations in the cited evidence (cancer patients, inpatients, patients with non-cancer

pain).

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)***:**

Validity of the process-outcome link was explicitly evaluated by the ACOVE, ACOVE3, ASSIST, and PEACE 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

Schenck AP, Rokoske FS, Durham DD, et al. The PEACE project: identification of quality measures for hospice and palliative care. J Pall Med 2010;13(12):1451-1459

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

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)***:**

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)***:**

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)***:**

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)***:**

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)*

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.**

[We are developing an e-measure that has not been implemented yet. The main reason for this is that the exclusions are not yet available in electronic medical record applications.]

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 above, 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)*.

[None]

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 with Benchmarking (external benchmarking to multiple organizations) Quality Improvement (Internal to the specific organization)	Public Reporting CMS Hospice Item Set https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Hospice-Quality-Reporting/Hospice-Item-Set-HIS.html

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

CMS Hospice Item Set for accountability for national use

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**

Data not yet available from the CMS HIS to track improvement. Later study data show better performance (see 1b.2)

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.

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.

None

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.)
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.
No appendix **Attachment:**

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** RAND Corporation/UCLA

Co.2 **Point of Contact:** Neil, Wenger, nwenger@mednet.ucla.edu, 310-794-2288-
Co.3 **Measure Developer if different from Measure Steward:** RAND Corporation/UCLA
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:**

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 #: 1625
Measure Title: Hospitalized Patients Who Die an Expected Death with an ICD that Has Been Deactivated
Measure Steward: RAND Corporation
Brief Description of Measure: Percentage of hospitalized patients who die an expected death from cancer or other terminal illness and who have an implantable cardioverter-defibrillator (ICD) in place at the time of death that was deactivated prior to death or there is documentation why it was not deactivated
Developer Rationale: Given that ICDs may cause increased symptoms at the very end of life and that the timing of approaching death can be identified for patients with incurable cancers and other end-stage illness, this issue should be addressed in patients where death is expected in order to reduce suffering toward end of life. There is also the potential for an impact on provider behavior through greater awareness of the issues ICDs pose to patients at the end of life.

Numerator Statement: Patients from the denominator who have their ICDs deactivated prior to death or have documentation of why this was not done
Denominator Statement: Patients who died an expected death who have an ICD in place
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 provided a rationale for the relationship of this process of care (deactivating an ICD in terminally ill patients) and patient outcomes by linking deactivation to reduction of stress to patients and caregivers and maximizing comfort near the end of life.
- The developers offered two sources of evidence for this measure:
  - A 2008 systematic review to identify the evidence supporting high-quality clinical practices for information and care planning for cancer patients.
  - A undated clinical practice guideline statement from Newcastle, North Tyneside and Northumberland: "*When ICD therapy is offered, it is good practice to also discuss that a time will come when switching off the device becomes the best option for the patient, i.e. as the end of natural life approaches. (page 4)*". This guideline was not graded. Note that NQF staff were unable to locate the guideline.
- At the time of the 2012 evaluation, the developer noted "*there is no study showing that outcomes are better for patients if an ICD is deactivated prior to an expected death*".
- During its 2012 evaluation of the measure, the Palliative/End-of-Life (EOL) Committee agreed that there was additional evidence and clinical knowledge supporting the measure that had not yet been incorporated into clinical guidelines.

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:

Even though the developers have attested that there is no new evidence, NQF staff conducted a brief search of the recent literature and found several articles, including two consensus statements, a systematic reviews on patient and provider attitudes on deactivation, and a systematic review of barriers to deactivation, as follows:

- http://www.heartrhythmjournal.com/article/S1547-5271(10)00408-X/pdf
- http://europace.oxfordjournals.org/content/europace/12/10/1480.full.pdf
- http://www.ncbi.nlm.nih.gov/pubmed/21926561
- http://www.ncbi.nlm.nih.gov/pubmed/22645402
- http://www.ncbi.nlm.nih.gov/pubmed/25597981
- http://www.ncbi.nlm.nih.gov/pubmed/23731284
- http://www.ncbi.nlm.nih.gov/pubmed/21943937
- http://www.ncbi.nlm.nih.gov/pubmed/23546173

Exception to evidence – N/A

Guidance from the Evidence Algorithm

Process measure supported by expert opinion and a non-graded guideline (Box 3) → Other empirical evidence available (Box 7) → Unclear if articles reflect the full body of evidence (Box 8) → If Committee agrees there is a high certainty that benefits clearly outweigh undesirable effects → Moderate (if includes all studies in the body of evidence), otherwise LOW

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?*
- *Do you know of other evidence that either supports or does not support this measure?*

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.

- The developer provided information on gap from two studies:
  - Data from 719 veterans with advanced cancer in 2008: only one died an expected death with an ICD in place, and it was deactivated prior to death
  - Data from 496 decedents hospitalized at least 3 days between 2005-2006: 25% had ICD deactivated

Disparities

- No data on disparities were provided.

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? Do you have other evidence demonstrating that a gap in case still exists?*
- *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.

Committee pre-evaluation comments

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

1a.
* This is a process measure that documents discontinuation of an ICD when death is imminent of documentation of a conversation if the patient/caregiver opts to continue. Discontinuation is the care process that allows patients to not actively die while receiving shocks, which can be painful and distressful for caregivers to observe. Recent guidelines from the Heart Rhythm Society supported by AAHPM and others support these recommendations. There is a strong ethical imperative for this measure.
* There is evidence that not having an ICD device switched off can lead to a stressful death. Terminal patients whose ICD device is not switched off and who have dangerous heart rhythms receive painful the shocks from the defibulator. Since this is a lifesaving device and the person is terminal, then deactivating it has been found to reduce the stress of waiting for a shock to occur. The literature is abundant and clear on this issue and the evidence is directly applicable to the process of care being measured. I would rate the evidence as moderate (using the algorithm).
* Evidence to support measure provided in 2012. There are no changes to the evidence. The available evidence is tangential as it pertained to only cancer patients with an ICD and another study recommended discussing deactivation with pts at time of insertion but is not well documented as to source. In 2012 the developer noted that there is no study showing that the outcome is better if the device is deactivated. NQF staff have identified recent literature including consensus statements. Relationship of the measure to outcome is supported. Rating for evidence supported by expert opinion is moderate.

1b.
* Data are too few to adequate conclude the existence of a performance gap. Disparities are not known.
* I am not sure if there is evidence of disparities in terms of deactivation; however, I do know that there is a disparity in terms of racial minorities being provided this treatment option (see cite below). With the information available, though, I would rate the opportunity for improvement as insufficient.

Casale, J. C., Wolf, F., Pei, Y., & Devereux, R. B. (2016). Socioeconomic and ethnic disparities in the use of biventricular pacemakers in heart failure patients with left ventricular systolic dysfunction. Ethnicity & disease, 23(3), 275-280
* Performance data provided but more than 5 years old and do not demonstrate that there is a gap in care. No Disparities in care were demonstrated

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): Paper medical records
Specifications:

- This measure is specified for the facility level of analysis for use in the hospital setting. A higher score indicates better quality.
- The numerator includes patients who have their ICDs deactivated prior to death or have documentation of why this was not done.
- The denominator includes adult patients who died an expected death who have an ICD in place. The hospitalization has to last at least 3 days. Expected death is defined as:
  - physician documentation at least 3 days before death that the patient´s illness was terminal or that the patient had a grave prognosis
  - patient was receiving comfort care
  - patient was receiving hospice care
  - patient had a life-threatening disease
  - patient was expected to die
- There are no exclusions to the measure.
- A calculation algorithm is provided.

Questions for the Committee :

- *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:

- Inter-rater reliability for a sample of 47 inpatient decedents

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:

- Reliability of the data elements was assessed using data obtained from the medical charts for 47 inpatient decedents (a 10% sample of 496 patients, 12 of whom had an ICD in place). Inter-rater reliability between two nurse abstractors was calculated. This is an appropriate method for demonstrating data element reliability.

Results of reliability testing

- Results from the testing indicated that both nurses agreed that none of the 47 patients were eligible for the measure. Apparently, none of the 12 patients with an ICD were included in the sample. Thus, inter-rater reliability testing for the measure numerator was not done.

Guidance from the Reliability Algorithm
Precise specifications (Box 1) → Empirical reliability testing with measure as specified (Box 2) → Score-level testing not conducted (Box 4) → data element level testing conducted (Box 8) → Denominator data element tested, but not the numerator data element (Box 9) → Insufficient

Questions for the Committee:
- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Preliminary rating for reliability: ☐ **High** ☐ **Moderate** ☐ **Low** ☒ **Insufficient**
Rationale: No testing conducted for the numerator data element.

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**
Specification not completely consistent with evidence

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:
- The ASSIST expert panel conducted a face validity assessment using the modified Delphi method.

Describe any updates to validity testing:
- No updated testing was provided.

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 ASSIST expert panel conducted a face validity assessment using the modified Delphi method.

Validity testing results:
- The developers do not provide the results of this expert panel review.

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?*

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)*:

- No data on meaningful differences were provided.

Question for the Committee:

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

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

2b7. Missing Data

- No information on missing data was presented (although this may not be an issue given data are collected from paper medical records).

Specifications consistent with evidence (Box 1) → potential threats to validity not assessed (Box2) → Insufficient

If no concerns around exclusions, 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

* There is essentially no inter-rater reliability data on this measure.
* The data elements are clearly defined and I have no concerns about this measure being implemented
* Reliability not consistently demonstrated with the specifications and rating for reliability is insufficient

Preliminary validity rating is insufficient

Measure can be implemented but transfers between place of service could make documentation inconsistent

* Reliability not consistently demonstrated with the specifications and rating for reliability is insufficient
Preliminary validity rating is insufficient
Measure can be implemented but transfers between place of service could make documentation inconsistent

Validity – Specifications

* The eligibility criteria (denominator) may be too narrow. Over half of persons in the US do not die in the hospital (e.g home hospice, long term care) where this measure does not apply (as it is restricted to hospital environments). Further, hospitalizations must be 3 days or longer to be included, which may be overly prescriptive. Much of the data presented by the stewards regarding prevalence of ICD's does not apply here due to dying at home, and not in hospital.
* The measure specifications are consistent with the evidence.
* Specifications are consistent with values of the target population

preliminary validity rating is insufficient due to lack of evidence- empirical validity testing not conducted

Reliability – Testing

* Reliability testing for paper-based abstraction is inadequate. Discontinuation of ICD's is not easily found in structured data fields in an EMR.
* Based on the reliability data provided, I do not feel this test sample is adequate for widespread implementation nor is their demonstration of sufficient reliability so that differences in performance can be identified. The sample data do demonstrate that ICDs can reliably be identified in the charts of hospitalized patients. I would rate reliability as

insufficient
* preliminary rating for reliability is insufficient and results do not demonstrate sufficient reliability

Validity Testing
* Validity testing comes from the measure stewards, but is also supported by several other medical societies. The harms of continuing ICD's during the active dying phase are well known and understood by consensus.
* I have concerns regarding validity because only face and expert panel validity were established and both are weak tests of validity. With the current information, I cannot agree that the score from this measure as specified is an indicator of quality.
* I agree that the score from this measure is an indicator of quality but was not adequately tested with appropriate method to demonstrate validity

Threats to Validity
* Exclusions - leaves out non-hospitalized patients, this is problematic

Risk adjustment - no presented

Meaningful differences - this is an all/none measure, performance below 100% would highlight an opportunity for improvement

Comparability - this could easily be compared across individual clinicians

Missing data - this is of high concern as these data elements are not easily found in medical documentation
* No exclusions, no risk adjustment, no data on meaningful difference and no information on missing data. At this point I would rate validity as insufficient
* no concerns around missing data identified, paper records utilized- missing data not a threat to the 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.

- As of the prior NQF evaluation, the developer reported that none of the data elements were included in electronic sources.

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**
Rationale: Requires paper medical record abstraction, and because not in use, there is no information on ease of implementation.

Committee pre-evaluation comments
Criteria 3: Feasibility

* The required data elements (e.g. d/c of an ICD) are not routinely captured in structured data (for an EMR) nor is there an easily-identifiable component of a paper chart where this would be recorded.
* The issue is that the developer reported that none of the data elements are included in EHR. Based on the literature, I do think the elements are generated and used in the hospital when someone is terminal; however, it would be burdensome to abstract the data from a paper record. I would rate this measure low on feasibility.
* Measure currently requires paper medical record abstraction and because not in use there is no information on the ease of implementation. Preliminary rating for feasibility is low

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 measure was included as a potential metric for inclusion in a publicly-reported dashboard to measure the quality of palliative care services by the California Department of Health Care Services. However, no additional information was provided regarding a final decision about use of this measure in this program.

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) supported the inclusion of this measure in the PQRS. However, this measure has not been included in the PQRS program.

***Questions for the Committee**:*

- *NQF guidelines require performance measures to 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

* The measure is not currently publicly reported or used in a reimbursement system. This is three years post NQF endorsement, which is troublesome.
* Currently, the rating for this measure's usability and use would be Low. The measure has not used in an accountability program, although the measure was consider for use by the California Department of Health Care Services. I think that in order for the measure to retain endorsement, there would need to be a clear path towards use in accountability programs or public reporting.
* Not currently being utilized > three years after implementation of the measure. There is planned use which has not been implemented , supported by CAPC for inclusion of the measure in the PQRS and supported by the measure Application Partnership in the PQRS. Preliminary rating is low. No current data on performance or improvement in

performance over time, No unintended consequences

Criterion 5: Related and Competing Measures
Related or competing measures • None **Harmonization** • N/A

Pre-meeting public and member comments

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)***:**

Measure is focused on addressing deactivating ICDs in terminally ill hospitalized patients who die an expected death with the intent of minimizing stress to patients and caregivers and maximizing end of life comfort.

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

Clinical Practice Guideline

Selected individual studies (rather than entire body of evidence)

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 study showing that outcomes are better for patients if an ICD is deactivated prior to an expected death, however this is recognized as a good clinical practice and specifically recommended in practice guidelines.

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)*:**

Also see 1a.4

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):3896-902.

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

When ICD therapy is offered, it is good practice to also discuss that a time will come when switching off the device becomes the best option for the patient, i.e. as the end of natural life approaches. (page 4)

1c.17 **Clinical Practice Guideline Citation:** Clinical Guidelines. Guideline Number: NoT 19. Newcastle, North Tyneside and.

Northumberland Guidelines (Adopted) on. ICD Deactivation Policy. www.northoftyne.nhs.uk/...guidelines/...deactivation%20of%20implantable%20cardioverter%20defibrillator

1c.18 **National Guideline Clearinghouse or other URL:**

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 developer's assessment of the quantity, quality, and consistency of the body of evidence?

1c.25 Quantity: Moderate **1c.26** Quality: High**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 #: 1625
De.2. **Measure Title:** Hospitalized Patients Who Die an Expected Death with an ICD that Has Been Deactivated
Co.1.1. **Measure Steward:** RAND Corporation
De.3. **Brief Description of Measure:** Percentage of hospitalized patients who die an expected death from cancer or other terminal illness and who have an implantable cardioverter-defibrillator (ICD) in place at the time of death that was deactivated prior to death or there is documentation why it was not deactivated
1b.1. **Developer Rationale:** Given that ICDs may cause increased symptoms at the very end of life and that the timing of approaching death can be identified for patients with incurable cancers and other end-stage illness, this issue should be addressed in patients where death is expected in order to reduce suffering toward end of life. There is also the potential for an impact on provider behavior through greater awareness of the issues ICDs pose to patients at the end of life.

S.4. **Numerator Statement:** Patients from the denominator who have their ICDs deactivated prior to death or have documentation of why this was not done
S.7. **Denominator Statement:** Patients who died an expected death who have an ICD in place
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
1625_Evidence_MSF5.0_Data.doc,1625_Evidence_MSF5.0_Data-635948666901510272.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)*

Given that ICDs may cause increased symptoms at the very end of life and that the timing of approaching death can be identified for patients with incurable cancers and other end-stage illness, this issue should be addressed in patients where death is expected in order to reduce suffering toward end of life. There is also the potential for an impact on provider behavior through greater awareness of the issues ICDs pose to patients at the end of life.

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.*
In a national study of 719 Veterans diagnosed with advanced cancer (665 died during study period), only one died an expected death with an ICD in place and it was deactivated prior to death.

In a decedent sample of 496 hospitalized patients who died during an admission of at least 3 days duration, 12 patients with an ICD in place died an expected death. In only 3 of these cases (25%) was deactivation of the ICD addressed.

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

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):1057-63.

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.*
None available

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.**
ICDs were designed to prevent sudden death. For patients who are expected to die, an ICD constitutes an inappropriate intervention that will promote suffering without achieving a patient goal. Most hospices admit patients with ICDs (Goldstein 2010) and these devices can present care problems and interfere with patient comfort. Problems include 1) ICD discharges that are both physically painful and emotionally distressing, 2) risk of the patient´s caregivers receiving an accidental shock, and 3) increased risk of ICD shock in terminally ill patients due to electrolyte disturbances, hypoxia, and heart failure. (McGeary 2009; Goldstein 2004; Kolata 2002). In a cross-sectional survey of 900 randomly selected hospices, 97% admitted patients with ICDs, and 58% reported that in the past year, a patient had been shocked. (Goldstein 2010) A retrospective cohort study questioned next of kin about discussions with patients who had ICDs in place and died of any cause. Only 27 out of 100 patients had had a discussion of whether or not the ICD should be deactivated. (Goldstein 2004) Additionally, in a small sample (n=12) of one-on-one interviews with physicians, almost every one

agreed that conversations about ICD deactivation should occur, but they acknowledged that they rarely did this. (Goldstein 2007) As the number of elderly patients increases, clinicians are likely to care for an increasing number of elderly patients with ICDs. (Sherazi 2008)

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

Goldstein N, Carlson M, Livote E, et al. Brief communication: Management of implantable cardioverter-defibrillators in hospice: a nationwide survey. Ann Intern Med 2010;153:296-299

Goldstein NE, Mehta D, Teitelbaum E, et al. "it´s like crossing a bridge" complexities preventing physicians from discussing deactivation of implantable defibrillators at the end of life. J Gen Intern Med 2007;23(Suppl 1):2-6

Goldstein NE, Lampert R, Bradley E, et al. Management of implantable cardioverter defibrillators in end-of-life care. Ann Intern Med 2004;141(11):835-8

Kolata G. Extending life, defibrillators can prolong misery. NY Times (Print) 2002;A1,18

McGeary A, Eldergill A. Medicolegal issues arising when pacemaker and implantable cardioverter defibrillator devices are deactivated in terminally ill patients.

Sherazi S, Daubert JP, Block RC, et al. Physicians´ preferences and attitudes about end-of-life care in patients with an implantable cardioverter-defibrillator. Mayo Clin Proc 2008;83(10):1139-1141

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, Cardiovascular, Cardiovascular : Congestive Heart Failure, Cardiovascular : Ischemic Heart Disease, Coronary Artery Disease

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

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)*
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 from the denominator who have their ICDs deactivated prior to death 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.)*
During hospitalization ending in an expected death

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.
Documentation in the medical record that the ICD was deactivated or documentation of a discussion of deactivation of the ICD with the patient or documentation of why ICD deactivation was not done.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Patients who died an expected death who have an ICD in place

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
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)*
Hospitalizations of adult patients of at least 3 days duration that ended in an expected death. Expected death is defined as physician documentation at least 3 days before death that the patient´s illness was terminal or that the patient had a grave prognosis, was receiving comfort care, was receiving hospice care, had a life-threatening disease, or was expected to die.

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)*
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)*

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 adult hospitalizations of at least 3 days duration that ended in patient death
2. Identify from the medical record patients who had an ICD in place
3. Identify from physician documentation patients who were noted to have had an expected death at least 3 days prior to death
4. Determine if the ICD was deactivated prior to death or documentation noted an attempt to discuss ICD deactivation with the patient or surrogate or other documentation addressing why the ICD was not deactivated.

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)*

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.
None

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.

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

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
1625_MeasureTesting_MSF5.0_Data.doc

NQF #: 1625	**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)***:**

In a decedent sample of 496 patients who died during an admission of at least 3 days duration, 12 patients with an ICD in place died an expected death.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)***:**

Two independent nurses abstracted a 10% reliability sample of the 496 cases. Identification of eligibility for a quality indicator and whether the care process was provided to satisfy the quality indicator were compared.

2a2.3 Testing Results *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**

In the 47 reliability records, there was 100% agreement that none satisfied the quality measure eligibility statement. Indicator had good agreement but there was not enough variability to calculate a kappa. These data suggest that ICDs can be reliably identified in a sample of hospitalized, seriously ill patients.

2b. VALIDITY. Validity, Testing, including all Threats to Validity: 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:**

Population tested is similar to populations in the cited evidence (hospice patients; patients who died an expected death).

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)***:**

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

Validity of the process-outcome link was explicitly evaluated by the ASSIST expert panel that reviewed the relevant literature and used a modified Delphi panel method of voting on the validity of the measure. (Lorenz 2009)

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):943-952

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)***:**

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)***:**

While the number of eligible patients for this indicator is small, the potential impact on the patient of failed performance is large. Given the importance of this measure, any failure in performance should be considered significant.

We have not evaluated multiple samples with this measure to identify differences in performance.

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. Evaluated only using medical records so far, although this measure could be implemented using an electronic health record in the future.

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.**

The data elements could be easily specified within an electronic record.

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 above, 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?*)

Early in field for implementing advance care planning and palliative care accountability. This measure was listed as a draft potential palliative care process measure on a draft set of measures dated June 1, 2015 for Medical by Department of Health Care Services. http://coalitionccc.org/2015/06/department-of-health-care-services-seeks-comments-on-quality-measures-for-medi-cals-pilot-palliative-care-service/

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 was listed as a draft potential palliative care process measure on a draft set of measures dated June 1, 2015 for Medical by Department of Health Care Services. http://coalitionccc.org/2015/06/department-of-health-care-services-seeks-comments-on-quality-measures-for-medi-cals-pilot-palliative-care-service/

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.
Because docuementation of a conversation about the deactivation decision satisfies the measure, there is little likelihood of unintended consequences.

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.

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.)
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:** Karl, Lorenz, karl.lorenz@va.gov, 310-478-3711-43523

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.
ASSIST project expert panel members and Advisory Board as listed below:

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, Division 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:** 2010

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

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:**

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: Moderate1c.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

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, including all Threats to Validity: 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:**

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 #: 1628
Measure Title: Patients with Advanced Cancer Screened for Pain at Outpatient Visits
Measure Steward: RAND Corporation
Brief Description of Measure: Adult patients with advanced cancer who are screened for pain with a standardized quantitative tool at each outpatient visit
Developer Rationale: Routine, ongoing accurate assessment of pain is the basis for guiding providers in developing pain management interventions and adjusting those interventions over time. A quantitative standardized tool allows for consistent and comparable measurements. Adequate pain control is a major factor in maximizing quality of life for advanced cancer patients.

Numerator Statement: Outpatient visits from the denominator in which the patient was screened for pain (and if present, severity noted) with a quantitative standardized tool
Denominator Statement: Adult patients with advanced cancer who have at least 1 primary care or cancer-related/specialty outpatient visit
Denominator Exclusions: None (other than those patients noted in 2a1.7. who did not survive at least 30 days after cancer diagnosis)

Measure Type: Process
Data Source: Electronic Clinical Data, Electronic Clinical Data : Registry, Paper Medical Records
Level of Analysis: Facility, Health Plan, Integrated Delivery System

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 provides a rationale for the link between the process of care (pain screening) and patient outcomes (effective pain management).
- The developer noted that the systematic reviews pertained to cancer pain management, stating that "*a central topic of this review focused on the need for regular pain screening as a means of providing information to guide pain management*". They also noted that the reviews found that "*pain screening is a necessary but not sufficient intervention in itself to improve quality of pain care*".
- During its 2012 evaluation of the measure, the Palliative/End-of-Life (EOL) Committee noted that pain assessment is a standard of care and well-documented and agreed that inadequate management as an outpatient is more likely to lead to increased healthcare costs than poor management as an inpatient.

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:

- Although systematic reviews were cited, they were not graded and the developer did not provide the quantity, quality, and consistency of the body of evidence linking screening for pain to improved patient outcomes. Thus, it is insufficient to meet NQF's criterion for evidence. However, an exception to the evidence criterion is allowed if the Committee agrees that empirical evidence is not needed to hold providers accountable for the measure and that other outcome measures or evidence-based measures are not available or feasible at this time.

Guidance from the Evidence Algorithm

Process measure but mostly tangential evidence (Box 3) → Evidence not graded (Box 7) → An outcome measure for pain exists, but is limited to hospice only (Box 10) → Systematic assessment of expert opinion (Box 11) → If Committee agrees it is OK/beneficial to hold providers accountable for performance in the absence of empirical evidence of benefits to patients → rate as INSUFFICIENT WITH EXCEPTION

Questions for the Committee:

- *Does the Committee agree there is no change in the evidence since the last evaluation?*

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

Rationale: No evidence links screening for pain to improved patient outcomes.

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

Maintenance measures – increased emphasis on gap and variation

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

- The developer provides performance data from ASSIST studies and from an unpublished report from the Veterans Health Administration. The sample sizes ranged from 467 to 9,485 and measure results ranged from 36% to 79%.

Disparities

- The developer does 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 screening for pain in outpatient cancer patients.

Committee pre-evaluation comments

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

1a.

* The developers have not provided updated evidence related to this process measure. Furthermore, the point made by AAHPM regarding that the denominator is narrow due to inclusion only of hospice patients rather than all terminally ill patients is important to consider. Using the Algorithm, I rate Insufficient with exception.
*unable to access documented with submitted evidence. Evidently contained SR s QOC or grading. RAND reports no change in evidence base since 2012 (really?) Not an outcome or PRO measure.

1b.

* The developers do provide performance data but they do not provide any evidence of a gap. Data on the measure by population subgroups was not provided. Without data it is difficult to determine that a national performance measure is warranted. I rate the opportunity for improvement as insufficient.
*Performance gaps easily identified by PUBMED search, including studies published since 2012. No data on new studies provided.

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 sources specified for this measure include paper medical records, EHRs, and registry (the latter for case assertainment).

Specifications:

- This measure is specified for the facility, health plan, and integrated delivery system levels of analysis. A higher score indicates better quality.
- The care setting for the measure is clinician offices/clinics.
- The numerator includes patients screened for pain with a quantitative standardized tool. Screening may be completed using verbal, numeric, visual analog, rating scales designed for use with nonverbal patients, or other standardized tools.
- The denominator includes adult patients with advanced cancer (i.e., Stage IV) who have at least 1 primary care or cancer related/specialty outpatient visit; this includes any oncology (medical, surgical, radiation) visit or chemotherapy infusion .
- Patients who do not survive at least 30 days after cancer diagnosis are excluded from the measure.
- A calculation algorithm is provided.

Prior evaluation

- The Committee questioned limiting the measure to Stage IV cancer patients; however, they acknowledged that testing was conducted only for this subgroup of cancer patients.

Questions for the Committee :

- *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 one ASSIST study but did not describe the actual testing methodology. They noted that testing was conducted using data from 467 outpatient cancer visits.

Results of reliability testing:

- The developers report an overall eligibility (denominator) kappa value of 0.87 and a specific care (numerator) kappa value of 0.86.
  - 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. Kappa values of 0.87 and 0.86 means that the raters agreed 87% and 86% of the time, respectively, over and above what would be expected by chance alone. According to the Landis and Koch classification, these results 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:

- *Is the test sample 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**

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:

- The ASSIST and ACOVE expert panels conducted a face validity assessment using the modified Delphi method.

Describe any updates to validity testing

- No updated testing was provided.

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 ASSIST and ACOVE expert panels conducted a face validity assessment using the modified Delphi method.

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:

- Developers do not provide any information on the number of patients who do not survive at least 30 days after cancer diagnosis.

Questions for the Committee:

○ *Are the exclusions consistent with the evidence?*
○ *Should patients who survive <30 days post-diagnosis be excluded from the measure?*
○ *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

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 low performance results 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.

Guidance from the Validity Algorithm

Specifications consistent with evidence (Box 1) → potential threats to validity not assessed (Box2) → Insufficient

If no concerns around exclusions, 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

* The data elements are clearly defined, as is the calculation algorithm. I do not have concerns about the likelihood that the measure could be consistently implemented.

*Variability among permitted instruments -- verbal, numeroic, visual analog rating scales all permitted -- introduces potential threat to reliability. Will be difficult to implement consistently if a centers use different instruments. No updated reliability data provided.

Validity – Specifications

* The specifications are consistent with the available evidence.

Reliability – Testing

* Although the developer did not provide updated information on reliability testing, the original reported inter-rater reliability results were good. I do not see a need to re-discuss and re-vote on reliability based on these earlier results. I do question whether reliability was tested sufficiently in scope to generalize for widespread implementation. Rating for reliability is moderate.

Validity Testing

* I do not think that the results demonstrate sufficient validity as the developers used face validity testing only, which is a very weak test of validity - they also did not provide results for this testing. I feel that the Committee should re-vote on the validity of the measure. I also do not feel validity was tested sufficiently in scope to generalize for widespread implementation.

*modified Delphi for face validity in 2012. No updates provided.

Threats to Validity

* I do not understand the rationale the exclusion of patients who did not survive at least 30 days after a cancer diagnosis and so have concerns about this. No evidence is provided regarding meaningful differences about quality and so it is difficult to determine if the measure does identify meaningful differences about quality. The developer also does not provide information on missing data. Therefore, I rate validity as insufficient.

*no data on missing data provided.

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.
- In 2014, the Measure Applications Partnership conditionally supported NQF#1628 for inclusion in the PPS-Exempt Cancer Hospital Quality Reporting (PCHQR) program, citing concerns with readiness for implementation. MAP noted the measure involves repeated screenings, and expressed concern that the measure would be costly and burdensome to implement. The measure is not currently implemented in the PCHQR program.
- In the 2012 endorsement of the measure, the Palliative and End-of-Life Care Steering Committee concurred that increasing use of EHRs would mitigated commenters' concern about data collection burden.

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 feel the required data elements are routinely generated and used during care delivery and that they are available in electronic form. I do not think it would be burdensome for medical abstractors to obtain date from a medical record. My concern about the data collection strategy would be the time factor for staff conducting an assessment each visit. For feasibility I give a rating of Moderate.
*paper chart abstraction is time-consuming. Easier if integrated into EMR

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 measure was included as a potential metric for inclusion in a publicly-reported dashboard to measure the quality of palliative care services by the California Department of Health Care Services. However, no additional information was provided regarding a final decision about use of this measure in this program.

Improvement results

- Longitudinal data for these measures are not 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: N/A

***Questions for the Committee**:*

- *NQF guidelines require performance measures to 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

* The measure is not currently being publicly reported although it has been considered for inclusion in a publicly-reported dashboard nor has it been in at least one accountability program three years after endorsement. I think the measure should only retain endorsement if a clear path to use in accountability programs or public reporting is established. The performance results can be used to further the goal of high-quality efficient healthcare by insuring that advanced cancer patients are having their pain assessed and re-assessed. The drawbacks are that 1) not all patients have access to facilities and clinics such that their pain can be assessed; and 2) even if they have access, there may be few outpatient visits and so how is there pain being assessed without the visits? I do think, though that the benefits of the measure outweigh any unintended consequences. I rate usability as low because it does not appear to be in use or plans made for future use.
*Unclear if being used in CA accountability program. Not currently being used in PCHQR program

Criterion 5: Related and Competing Measures

Related and competing measures

- Other measures of pain screening, assessment, or outcome are specified for different settings or levels of analysis:
  Skilled Nursing Facility:
  - 0677: Percent of Residents Who Self-Report Moderate to Severe Pain (Long-Stay)),
  - 0675: The Percentage of Residents on a Scheduled Pain Medication Regimen on Admission Who Self-Report a Decrease in Pain Intensity or Frequency (Short-stay))

  Home Health:
  - 0523: Pain Assessment Conducted
  - 0524: Pain Interventions Implemented

  Clinician:
  - 0420: Pain Assessment and Follow-Up

  Hospice and hospital palliative care:
  - 1634: Hospice and Palliative Care -- Pain Screening
  - 1637: Hospice and Palliative Care -- Pain Assessment
- 0383: Oncology: Plan of Care for Pain – Medical Oncology and Radiation Oncology (paired with 0384) reports the rate of care plans to address pain in cancer patients. While the measures capture different clinical processes, NQF#0383 was cited by the Measure Applications Partnership as potentially duplicative in a review of measures considered for inclusion in the PPS-Exempt Cancer Hospital Quality Reporting program.

Harmonization

- Due to differences in care settings, target populations, and levels of analysis, the Committee may not be asked

to select a best-in-class measure. However, 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: Recommend suggesting assessment tools

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)***:**

This measure focuses on the necessity of a standardized quantitative screen for pain (and if present, severity) with each contact to provide the input required to effectively manage advanced cancer pain over time.

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

Selected individual studies (rather than entire body of evidence)

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)***:**

A systematic review was conducted of cancer pain management. A central topic of this review focused on the need for regular pain screening as a means of providing information to guide pain management.

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):*
Pain screening was a necessary but not sufficient intervention in itself to improve quality of pain care.

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

No harms identified. Consistent benefit is related to effective pain management.

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? Yes

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: The body of evidence was graded by the RAND Corporation Evidence-based Practice Center (EPC) ASSIST team.

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

1c.12 **If other, identify and describe the grading scale with definitions:** Randomized controlled trial; non-randomized controlled trial, 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:** There was no controversy regarding pain screening in itself. Studies that showed increased quality of pain care also included feedback of quantitative pain measurement.

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

Dy SM, Asch SM, Naiem A, et al. Evidence-based standards for cancer pain management. J Clin Oncol 2008;26(23):3879-3885

Lorenz KA, Lynn J, Dy S, et al. Quality measures for symptoms and advance care planning in cancer: a systematic review. J Clin Oncol 2006;24:4933-4938

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:**

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 developer's assessment of the quantity, quality, and consistency of the body of evidence?

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

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
1628_Evidence_MSF5.0_Data.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)*
Routine, ongoing accurate assessment of pain is the basis for guiding providers in developing pain management interventions and adjusting those interventions over time. A quantitative standardized tool allows for consistent and comparable measurements. Adequate pain control is a major factor in maximizing quality of life for advanced cancer patients.

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, % Performance
Assessing Symptoms Side Effects and Indicators of Supportive Treatment (ASSIST)(Walling 2013): Advanced cancer outpatient encounters in a VA National Sample, N=657, 58.1%

N, % performance
Assessing Symptoms Side Effects and Indicators of Supportive Treatment (ASSIST)(Dy 2011): Advanced cancer outpatient encounters at a comprehensive cancer center, N=467, 79%

ASSIST (Malin 2010): Advanced cancer outpatient encounters at VA facility, 36%

Veterans Health Administration (VHA) (VHA unpublished report): Advanced lung cancer outpatient encounters at VA facility: N=9485, 70%

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 Pall Med 2011;14(4):451-457

Malin JL, O´Neill SM, Asch SM, et al. Quality of supportive care for patients with advanced cancer in a VA medical center. J Pall Med 2011;14(5):1-5

VHA, Executive Summary: The Quality of VHA Lung Cancer Care. VHA Office of Quality and Performance Special Study, November 2010. (Unpublished)

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.*
None available yet

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**
Affects large numbers, 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.**
Although increasing pain assessment has been shown not to be sufficient for improving pain outcomes (Mularski 2006; Morrison 2006), increased routine pain assessment was found to be a key component of RCTs and quality improvement interventions that have been successful for improving pain in a systematic review of cancer pain quality measures and evidence supporting their use. (Lorenz 2006; Dy 2008) Routine pain assessment is recommended by numerous organizations and pain guidelines, including National Comprehensive Cancer Network. (NCCN 2006) Routine measurement of pain is a necessary first step in pain management. Without asking patients, clinicians´ assessments of pain are usually inaccurate, and correlation was worst for patients with severe pain. (Purcell 2003) Discrepancies between patients and physicians in perceptions of pain severity are predictive of inadequate management. (Cleeland 1994) Without regular screening for pain, many patients with significant pain do not have pain documented in the medical record and do not receive analgesia. (Rhodes 2001) A study of 76,759 patients who died of cancer revealed that they made 36,600 emergency department visits in the last 6 months of life, and the most common reason for these visits was abdominal pain. (Barbera 2010)

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
Barbera L, Taylor C, Dudgeon D. Why do patients with cancer visit the emergency department near the end of life? Can Med Assoc J 2010;182(6):563-569

Cleeland CS, Gonin R, Hatfield AK, et al. Pain and its treatment in outpatients with metastatic cancer. N Engl J Med 1994;330(9):592-596

Dy SM, Asch SM, Naeim A, et al. Evidence-based standards for cancer pain management. J Clin Oncol 2008;26(23):3879-38885

Lorenz KA, Lynn J, Dy S, et al. Quality measures for symptoms and advance care planning in cancer: a systematic review. J Clin Oncol 2006;24(30):4933-4938

Morrison RS, Meler DE, Fischberg D, et al. Improving the management of pain in hospitalized adults. Arch Intern Med 2006;166(9):1033-1039

Mularski RA, White-Chu F, Overbay D, et al. Measuring pain as the 5th vital sign doe not improve quality of pain management. J Gen Intern Med 2006;21(6):607-612

NCCN: Adult Cancer Pain. http://www.nccn.org/professionals/physician_gls/PDF/pain.pdf Accessed December 4, 2006

Purcell W, Grossman S, Carson K. High outpatient pain scores identify patients at high risk for inpatient hospital admission. Proc Am

Soc Clin Oncol 2003;22:737

Rhodes DJ, Koshy RC, Waterfield WC, et al. Feasibility of quantitative pain assessment in outpatient oncology practice. J Clin Oncol 2001;19(2):601-508

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

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)
Attachment:

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

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.
Outpatient visits from the denominator in which the patient was screened for pain (and if present, severity noted) with a quantitative standardized tool

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.)*
At the time of outpatient visit(s)

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.
Pain screening with a standardized quantitative tool during the primary care or cancer-related/specialty outpatient visit(s). Screening may be completed using verbal, numeric, visual analog, rating scales designed for use with nonverbal patients, or other standardized tools.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Adult patients with advanced cancer who have at least 1 primary care or cancer-related/specialty outpatient visit

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
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)*
Adult patients with Stage IV cancer who are alive 30 days or more after diagnosis and who have had at least 1 primary care visit or cancer-related/specialty outpatient visit. Cancer-related visit = any oncology (medical, surgical, radiation) visit, chemotherapy infusion

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
None (other than those patients noted in 2a1.7. who did not survive at least 30 days after cancer diagnosis)

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 patients at least 18 years of age with Stage IV cancer

2. Identify patients who have had at least 1 primary care or cancer-related visit. Exclude patients who are not alive 30 or more days after diagnosis.

3. For each applicable visit, determine if a screening for pain was performed using a quantitative standardized tool.

4. Performance score = number of visits with standardized quantitative screening for pain/total number of outpatient visits

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)*

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.

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.

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.
Electronic Clinical Data, Electronic Clinical Data : Registry, 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.
Patients were identified via the testing organizations´ cancer registries.
At one institution, outpatient pain vital sign scores were extracted electronically from the patient EHR.
At other institutions, quantitative pain scores were collected via medical record abstraction.

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)*

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility, Health Plan, Integrated Delivery System

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic
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
1628_MeasureTesting_MSF5.0_Data.doc

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 1628	**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)***:**

ASSIST (Dy 2011) outpatient cancer visits (n=467): Overall eligibility kappa=0.87; overall specific care kappa=0.86

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

2b. VALIDITY. Validity, Testing, including all Threats to Validity: 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)***:**

Validity of the process-outcome link was explicitly evaluated by the ASSIST and ACOVE expert panels that reviewed the relevant literature and used a modified Delphi panel of voting on the validity of the measure (Lorenz 2009, Etzioni 2007)

Etzioni S, Chodosh J, Ferrell BA, et al. Quality indicators for pain management in vulnerable elders. J Amer Geriatr Soc 2007;55S(53):5403-5408

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

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)***:**

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)***:**

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)*:

This measure provides a broader range of performance than is typical for pain screening. Performance data revealed many settings of care where routine quantitative pain assessment with a standardized tool is not being done.

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)***:**

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)*:

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.**

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

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.**

As pain scores are increasingly documented consistently in medical records where this data can be obtained electronically and/or ability to capture screening information through natural language processing in medical notes improves, reliability and validity of these approaches can be tested compared to medical record abstraction.

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 above, 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?*)

Early in field for implementing advance care planning and palliative care accountability. This measure was listed as a potential palliative care process measure on a draft set of measures dated June 1, 2015 for MediCal by Department of Health Care Services. http://coalitionccc.org/2015/06/department-of-health-care-services-seeks-comments-on-quality-measures-for-medi-cals-pilot-palliative-care-service/

Improvement

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 was listed as a draft potential palliative care process measure on a draft set of measures dated June 1, 2015 for Medical by Department of Health Care Services. http://coalitionccc.org/2015/06/department-of-health-care-services-seeks-comments-on-quality-measures-for-medi-cals-pilot-palliative-care-service/

Improvement

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**

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 primary palliative care, implementation of measures such as this are key to improvement.

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.**
none known

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.

Measures 0677, 0675, 0523, and 0524 apply to nursing home and home health care settings and are, therefore, not competing with the proposed measure.

It is unclear exactly what the scope of measure 0420 is, however it appears to be directed at ancillary, non-physician professionals. It is unclear what "initiation of therapy" is referring to. The measure´s endorsement is time limited (endorsed July 31, 2008)

Measure 0384 (paired with 0383) also has a time-limited endorsement (endorsed July 31, 2008). This measure targets only patients

who are currently receiving chemotherapy or radiation therapy, and by definition, excludes some patients with advanced cancer who are not receiving this type of treatment. The proposed measure targets patients with Stage IV cancer and includes more venues of care than the existing measure where it would be applied (primary care and all cancer-related outpatient visits). This is in keeping with the reality that pain and pain control becomes a central focus for patients with late-stage cancer, and regular pain assessment should occur in multiple outpatient care settings. The developers propose that measure 0383 be limited to patients with Stage I-III cancer and endorse the proposed measure which targets Stage IV cancer patients.

Proposed measure 1634: Hospice and Palliative Care - Pain Screening: Proposed measure 1634 targets patients with serious conditions who are entering hospice or hospital-based palliative care. The measure proposed here targets a sub-population (advanced cancer). However, the setting and timing of 1634 is hospice/palliative care admission and is a one-time screen. 1628 focuses on pain screening at all outpatient visits. Although the 2 measures focus on different venues of care (and 1 is a time measure and the other every visit), they are completely harmonized in content.

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:** Karl, Lorenz, karl.lorenz@va.gov, 310-478-3711-43523

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:** 2010

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

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:**