NATIONAL QUALITY FORUM
Driving measurable health improvements together

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
Corresponding Measures:
Measure Title: Hospice and Palliative Care – Treatment Preferences
Measure Steward: University of North Carolina-Chapel Hill
sp.02. Brief Description of Measure: Percentage of patients with chart documentation of preferences for life sustaining treatments.
1b.01. 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.
sp.12. Numerator Statement: Patients whose medical record includes documentation of life sustaining preferences
sp.14. Denominator Statement: Seriously ill patients enrolled in hospice OR receiving specialty palliative care in an acute hospital setting.
sp.16. Denominator Exclusions: There are no denominator exclusions for this measure.
Measure Type: Process
sp.28. Data Source: Assessment Data; Electronic Health Records
sp.07. Level of Analysis: Clinician: Group/Practice; Facility
IF Endorsement Maintenance – Original Endorsement Date: 02/14/2012
Most Recent Endorsement Date: 10/26/2016
IF this measure is included in a composite, NQF Composite#/title:
IF this measure is paired/grouped, NQF#/title:
sp.03. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?:

Preliminary Analysis: Maintenance of Endorsement

To maintain NQF endorsement, endorsed measures are evaluated periodically to ensure that the measures still meet 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 ***structure, process or intermediate outcome*** measure are that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured. For measures derived from patient report, evidence also should demonstrate that the target population values the measured process or structure and finds it meaningful.

The developer provides the following description for this measure:

- This is a maintenance process measure at the facility and clinician group/practice level that assesses the percentage of patients who have their preferences for life sustaining treatments documented within their chart.
- The developer provides a logic model that depicts actions that the accountable entity can take, such as eliciting and documenting patient treatment preferences, increased used of advance directives, and care planning through engaging patient values. These actions ultimately result in improved patient autonomy and control over treatments, increased patient and family satisfaction, enhanced transitions to hospice and palliative care services, and reduced emotional distress for family members.

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 2016

- The developer cited individual studies and systematic reviews examining the impact of communication quality on patient and family outcomes. The developer attested that there had been no changes in the evidence since endorsement in 2012 but provided two new guidelines during the 2016 review: the 2014 Michigan Quality Improvement Consortium guideline and the 2013 Institute for Clinical Systems Improvement (ICSI) guideline on Palliative Care for Adults.
- During the previous evaluation, the Standing Committee agreed that the updated evidence appears to be directionally the same since the last NQF endorsement evaluation in 2012. The Standing Committee accepted the prior evaluation without further discussion.

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 one new source of updated evidence for this measure:

- The developer cited the updated 2020 ICSI Palliative Care Guideline Workgroup recommendation that states that facilitation of advance care planning conversations is appropriate for all adult patients.

Regular review of goals and wishes should occur as the patient's condition or life circumstances change.

- Quality: The developer noted that the recommendation strength is strong using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) methodology system.
- Quantity: The developer noted that ICSI identified 63 supporting references in support of the updated recommendation and rated the quality of evidence as low.
- Consistency: The developer noted that a regular review of Advance Care Planning (ACP) ensures accurate documentation of the patient's treatment wishes and family understanding of all treatment options, including benefits and burdens. Although the consistency of the benefit across the 63 supporting references was not provided by the developer.

Exception to evidence

- NA

Questions for the Committee:

- *How does the Standing Committee assess the quality of the evidence for the recommendation provided by the developer, noting 1a05 describes it as low quality, while 1a.07 describes the strength of the recommendation as strong.*
- *Is the evidence directly applicable to the process of care being measured?*
- *Does the Standing Committee feel that the evidence supports the relationship between medical record documentation of life-sustaining preferences for seriously ill patients enrolled in hospice OR specialty palliative care in an acute hospital setting to desired patient outcomes and reduced utilization?*

Guidance from the Evidence Algorithm

Not a Health Outcome or PRO (Box 1) -> Process measure based on systematic review (Box 3) -> QQC presented (Box 4) -> Quantity: high; Quality: moderate; Consistency: moderate (Box 5) -> Moderate (Box 5b) -> Moderate

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

1b. Gap in Care/Opportunity for Improvement and 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 performance scores at the facility-level from the Hospice Quality Reporting Program (HQRP) for Fiscal Year (FY)15. The analyses encompassed 3,922 hospice organizations and approximately 1,218,786 patient stays.
  - The mean score was 98 percent with a range from 0 to 100 percent, the median was 100 percent, the interquartile range (IQR) was 1.5, and the standard deviation was 6.3.
  - The developer notes that 53.5 percent of hospices had perfect scores and 4.6 percent of hospices scored below 90 percent.
- The developer did not provide data specific to palliative care for clinician group/practices or facilities.
- The developer cited multiple studies and noted opportunities for improvement:
  - The Institute of Medicine (IOM) in 1997 identified poor communication regarding patient care preferences is a quality-of-care concern for palliative and end-of-life care services.

- A 1995 Study to Understand Prognosis and Preferences for Outcomes and Risks of Treatment (SUPPORT) found discrepancies between patient report of treatment preferences and provider awareness of or use of these preferences to guide treatment.
- A 2001 quantitative national survey showed that patients and families prioritize communication with providers and control over treatment choices when faced with serious or life-threatening illness.
- A 2004 study that assessed the effectiveness of different communication skills training courses for health professionals in cancer care found that physicians and other providers fail to open the door to these discussions at critical time points in illness progression.
- A 2007 systematic review found a consistent discrepancy between the quality and content of communication provided by the providers and experiences by the seriously ill patient and their families.

Disparities

- Racial disparity analysis:
  - The developer noted that patients with unknown racial and ethnic group information had the lowest rate of documentation of treatment preferences (94.5 percent), while patients identifying as White non-Hispanic had the highest rate (98.0 percent, $p<0.001$) of treatment preferences documented.
  - The developer also examined the differences across hospices with proportions of non-White patients that are greater than or equal to the national median proportion (11.9 percent), and hospices with fewer non-White patients than the national median.
    - The results showed that hospices with proportions of non-White patients that are greater than or equal to the national median proportion performed better at 98.4 compared to 97.6 for hospices with fewer non-White patients then the national median. The developer notes that while these results were statistically significant, they do not find them to be clinically substantial.

- Gender disparity analysis:
  - The developer compared documentation of treatment preferences between female and male patients and found a slightly lower rate of documentation of treatment preferences for female (97.6 percent) than for male patients (97.7 percent, $p=0.0029$).
  - The developer examined the differences in measure performance across hospices with proportions of female patients greater than or equal to the national median proportion (55.2 percent).
    - The results showed that hospices with proportions of female patients that are greater than or equal to the national median proportion did not perform better at 97.9 percent compared to 98.0 percent for hospices with fewer female patients then the national median.
- Socioeconomic disparity analysis:
  - The developer analyzed socioeconomic disparities using Medicaid status as a proxy measure of low socioeconomic status (SES).
  - The developer noted that both Medicaid and non-Medicaid patient groups show a comparable rate of documentation of treatment preferences at 97.4 percent.
  - The developer reports that hospices with a smaller proportion of Medicaid patients are less likely to provide treatment preferences to patients upon admission at 97.4 percent compared to hospices with a greater proportion of Medicaid patients at 98.5 percent ($p<0.001$).
- Rural-urban disparity analysis:

- The developer compared measure scores between rural and urban hospices and reported that measure performance was significantly different between rural and urban hospices (98.8 percent compared to 97.7 percent, p<0.001).

- The developer did not provide disparities data specific to palliative care for clinician group/practices or facilities.
- It is unclear if the developer reported clinician: group/practice disparity and performance gap data.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Does the Standing Committee have any concerns with the measure being topped out?*
- *Does the Standing Committee have any concerns with the level of analysis for the performance data provided (i.e., clinician/ group, facility, both)?*

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

Rationale:

- Low rating- The data presented by the developer indicates little opportunity for improvement, or overall, less-than-optimal performance across providers and/or population groups (topped-out).
- Insufficient rating-The developer did not report data for the clinician/group level.

Committee Pre-evaluation Comments:

1a. Evidence

- It is a little concerning that the although there is a strong recommendation for ACP, there is not consistent agreement regarding its benefit.
- There is some well known controversy in this space about the value of ACP. However, it is helpful that the developers have specified that, to qualify for the numerator, the presence of EOL documents in the chart is not enough. There must be documentation of a conversation with the patient/family. Clinical communication is key and known to result in desirable outcomes including satisfaction and cost. On the quality of the evidence, part of the issue is due to the vagaries of the two definitions involved. Low quality is defined as "Further research is very likely to have an important impact on our confidence in the estimate of effect". The only difference between Low and Moderate is in the latter "very" is dropped. In the Grade, there are only two choices, :"Strong" and "weak" Given that choice, it would seem "strong " is defensible. Staff rated the evidence low seemingly because the evidence is generally tangential. Again technically, that is likely the correct category. However, the evidence seems clearly sufficient. There is a new systematic review (Institute for Clinical Systems Improvement Palliative Care Guideline Workgroup. Palliative Care for Adults. Sixth edition. Updated January 2020. p.48). While this is technically "new" evidence, at best it presents a summary of already existing evidence that reaches the same conclusion that the committee has apparently reached twice, there seems no need to discuss or revote.
- Not aware of new evidence

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

- The data presented by the developer was fairly old and lacking in somme areas such as palliative care for clinician group/practice or facilities. I do not think a gap was demonstrated. Disparities were documented.
- Notably this data is almost 7 years old. At that time, the developers did an analysis of disparities data across gender, race, rural urban and socioeconomic status is. Compliance rates are all in the upper 90%s. No palliative care data provider. Hospice Quality Reporting Program 2015. It is not clear

anywhere in this report that there is any data from specialty palliative care patients in an acute hospital. They do report a small reliability study on this population initially but nothing since. The developer did not provide data specific to palliative care for clinician group/practices or facilities. However, the developers state that data collection in this area is in progress. It would be useful for the developers to tell the committee exactly where this effort is and when the committee can expect to see some data. I have no concerns about the level of analysis. I do have significant concerns about whether there is a gap of care that warrants a national measure and whether this measure has topped out. All of the data points reported are in the mid or upper 90s. The median is 98% and the 10th percentile is 95%. There are several gaps which are reported as statistically significant. However, with an N of over 1 million, the gap mostly turns out to be one or two percentage points with both scores being above 95%. These gaps do not seem to be clinically significant or actionable from a quality standpoint. With reimbursement to hospices lagging behind inflation and even more quality measures on the way (e.g. Hospice Visits in the Last Days of Life), it is hard to find a justification to keep this measure in the program.

- Yes, there is a demographic difference between groups who discuss end of life wishes and/or document treatment preferences

Criteria 2: Scientific Acceptability of Measure Properties

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

Evaluators: Staff

2a. Reliability: Specifications and Testing

For 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.

For 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.

Specifications:

- The measure is specified at the Clinician: Group/Practice and Facility levels using the Hospice Item Set (HIS), which is a standardized, patient-level data collection instrument, part of the Hospice Quality Reporting Program (HQRP)
- After analysis of FY15 hospice data, the developer has changed the specifications to no longer exclude hospice stays of less than 7 days.
- The developer notes that the length of stay (LOS) change was reviewed during the previous maintenance submission and that removing the LOS exclusion had little impact on the distribution of hospices' QM scores.
- The Standing Committee requested clarification from the developers on what meets the criteria for patient care preference documentation. The developer responded that the measure requires evidence of a discussion with the patient (or with the surrogate decision-maker if the patient has lost decisional capacity) and that simply having the preferences included in the patient record (e.g., via a living will or a Do-Not-Resuscitate order) is not sufficient to meet the quality measure.

Reliability Testing:

- No new reliability testing was provided by the developer since the previous measure evaluation in 2016.
    - In 2016, the Standing Committee strongly recommended reliability testing for the clinician-level measure in the palliative care setting be updated and results provided to NQF when available.
- Reliability testing conducted at the Patient/Encounter Level:
    - The developer conducted inter-rater reliability testing in 2012 on a random subset of 20 seriously ill patients using two nurse abstractors and a structured chart abstraction tool.
        - The developer noted that numerator and denominator data were captured separately.
        - The developer reported that the abstractors achieved perfect inter-rater reliability for this measure with a kappa statistic of 1.0.
        - It is unclear from the results provided if the key data element results were reported.
- Reliability testing conducted at the Accountable Entity Level:
    - During the previous evaluation in 2016, the developer conducted split-half reliability analysis using HIS data from 10/2014 through 09/2015 on all facilities with 20 or more patients stays.
        - The developer reported an intraclass coefficient (ICC) of 0.91, indicating high internal reliability.
    - In 2016, the developer conducted a signal to noise ratio analysis to determine what proportion of total variance in the measure is attributable to differences among providers.
        - The signal-to-noise ratio was 0.98, which the developer indicates as strong reliability.
    - The developer conducted a stability analysis across four consecutive quarters in 2014-2015 to determine to what extent a provider's performance changes across time.
        - The developer indicated that slightly less than 95% of facilities had a change in quality measure scoring of <1 standard deviation across quarter ranging from 93.5 in Q4 2014-Q1 2015 to 93.9 in Q1-Q2 2015 to 93.6 in Q2-Q3 2015.

Questions for the Committee regarding reliability:

- *The developer attests the specifications have not changed and that additional reliability testing was not conducted. Does the Committee agree that the measure is still reliable and there is no need for repeat discussion and vote on Reliability?*
- *Do you have any concerns that the measure cannot be consistently implemented (i.e., are measure specifications adequate)?*
- *Does the Standing Committee have any concerns with the level of analysis for the performance data provided (i.e., clinician, group, facility, both)?*

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

- Moderate rating at the facility level.
- Insufficient rating at the Clinician/group-level level of analysis is not identified in the testing provided.

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

For maintenance measures – less emphasis if no new testing data provided.

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.

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

Validity Testing

- No new validity testing was provided by the developer since the previous measure evaluation in 2016.
    - The Standing Committee strongly recommended reliability testing for the clinician-level measure in the palliative care setting be updated and results provided to NQF when available during the 2016 evaluation.
- Validity testing conducted at the Accountable Entity Level:
    - Face validity was assessed in 2012 by a group of nursing and physician stakeholders who were asked to comment on the validity, accuracy, and actionability of the measure. The developer noted that the face validity results from the stakeholder group indicated broad endorsement of the face validity of the measure
        - It is unclear from the face validity testing provided if testing was accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.
        - The degree of consensus and any areas of disagreement was not provided or discussed.
    - The developer performed construct validity at the facility level in 2012 by comparing documentation of life sustaining preferences among seriously ill patients receiving specialty interdisciplinary palliative care with those patients who did not.
        - The results found that patients who received specialty palliative care were more likely to have documentation of their life-sustaining preferences compared to those who did not receive palliative care (91% vs 59%, $p<0.001$).
    - In 2016, the developers conducted non-parametric Spearman rank correlation analysis at the facility level with five (no longer NQF-endorsed) hospice quality measures:
        - #1634 *Hospice and Palliative Care -- Pain Screening*;
        - #1637 *Hospice and Palliative Care -- Pain Screening*,
        - #1638 *Hospice and Palliative Care - Dyspnea Treatment*;
        - #1639 *Hospice and Palliative Care -- Dyspnea Screening*; and
        - #1647 *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*.
        - The developer hypothesized that agencies should perform similarly on assessment processes at hospice admission and expected the resulting correlations to be high.
        - The developer noted that the measure had the strongest correlation with NQF #1647 Believes/Values Addressed (Spearman's rho= 0.64).

Exclusions

- The measure does not use exclusions.

Risk-Adjustment

- The measure is not risk adjusted or stratified.

Meaningful Differences

- The developer previously conducted confidence interval analysis at the facility level and found that 23.3 percent of hospices were more likely to report scores above the national mean of 98.0 percent compared to 9.9 percent of those hospices that were below the national mean.

- The developer also noted that 53.5 percent of hospitals had perfect measure scores (100 percent) and 4.6 hospices scored below 90 percent (range= 0-100 percent median= 100 percent, IQR= 1.5 percent, SD= 6.3 percent).

Missing Data

- The developer conducted analyses on two levels: eligible patient stays and hospice level. The developer examined three items (F2000B, F2100B, and F2200B) on the HIS that can include missing data and impact the validity of the measure. This analysis was conducted at the patient stay and hospice level.
    - It is unclear from the missing data provided what level of analysis is tested for patient-stay (i.e., clinician: group/practice, facility, both).
    - It is also unclear from the testing provided what the definitions are for the three HIS items used to determine the impact of missing data on validity.
    - The developer noted that the overall rate of missing data at the patient-level ranged from 0.01 to 0.02 percent.
    - At the hospice level, over 95 percent of hospices did not have any admissions with missing data related to these three items.

Comparability

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

Questions for the Committee regarding validity:

- *The developer attests the specifications have not changed and that additional validity testing was not conducted. Does the Standing Committee agree that the measure is still valid and there is no need for repeat discussion and vote on Validity*?
- *Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?*
- *Does the Standing Committee have any concerns with the level of analysis for the performance data provided (i.e., clinician, group, facility, both)?*

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

- Moderate rating at the facility level.
- Insufficient rating: Clinician/group-level testing is not identified in the testing provided.

Committee Pre-evaluation Comments:

2a1. Reliability-Specifications

- I still have concerns about the criteria for patient care preference documentation as I did the first time this was reviewed.
- The data elements are clearly specified. Originally patients with a stay < 7 days were excluded . At the 2016 review the committee agreed to remove this exclusion. The developers demonstrated that this exclusion did not impact the reliability. During the last review, the developers also clarified for the committee what is considered adequate documentation of patient care preferences. The staff and committee appear to have no other concerns with the specifications or definitions. I agree with that evaluation. The measure is specified at the Clinician: Group/Practice and Facility levels using the Hospice Item Set (HIS), which is a standardized, patient-level data collection instrument, part of the Hospice Quality Reporting Program (HQRP). It is not clear how the hospital data is being collected and what data is being collected. It would be helpful for the developers to be clearer about where and how

hospital palliative care patients are included and where they are not. If not, what plans are under way to collect and include this data.

- While this measure can clearly show performance among those receiving hospice care, not sure how it does that for those receiving palliative care

2a2. Reliability-Testing

- The Standing Committee recommended additional testing in 2016 and I am disappointed that the developer chose not to do so.
- In 2016, the developers did interrater reliability testing at the Patient/Encounter level, reliability testing at the Accountable Entity level, sign Signal to noise ratio analysis and stability analysis. All the results were within acceptable ranges. There is no apparent reason to believe that the reliability testing needs to be redone at least in the hospice population. It is unclear what testing has been done on the palliative care population. In 2016, the Standing Committee strongly recommended reliability testing for the clinician-level measure in the palliative care setting be updated and results provided to NQF when available. It is not clear that this testing was ever done or data provided to NQF as requested. It would be helpful for the developer to address this issue. The developers state that they have two studies in progress which apparently will update this data. They expect the data to be available next year. The issues with the hospital data should be discussed and reliability revoted.
- There is mixed information about reliability for this measure but I don't know how to address or resolve that.

2b. Validity

- The only concern I have is in regard to no clinician/group level testing.
- Validity was tested for both face validity and construct validity. I have no concerns.
- The validity may be low although I lack the expertise to make that determination

2b2-3. Other Threats to Validity

- I have no concerns about exclusions or risks.
- There are no exclusions and the measure is not risk adjusted.
- This is not risk ajusted but perhaps should be due to demographic differences

2b4-7. Other Threats to Validity

- I have no concerns.
- There are no apparent threats to validity and no missing data.
- There are meaningful difference by demographic group, not sure how that affects the measure's overall performance

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 a combination of electronic sources.
- Data elements are abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)
- The developer reports that since no unique identifiers are required in record abstraction, confidentiality is preserved.

Questions for the Committee:

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

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

Committee Pre-evaluation Comments:

3. Feasibility

- No concerns about feasibility.
- All data elements are available electronically. The measure has been embedded in EMRs. The hospice date again comes from the quality reporting program so can easily be extracted. There are no difficulties of note with the hospice data. Again, it is unclear how the hospital palliative care data is being provided if at all. There are no unique identifiers required in the extraction of the data.
- No concerns

Criterion 4: Use and Usability

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

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

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

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

Current uses of the measure

Publicly reported? ☐ Yes ☒ No

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

Planned use in an accountability program? ☐ Yes ☒ No ☐ NA

Accountability program details

- The developer highlights that this measure is one of ten recommended quality measures in the Hospice and Palliative Nurses Association and the American Academy of Hospice and Palliative Medicine (AAHPM/HPNA) Measuring What Matters (MWM) project.
- The developer noted that MWM measures are part of the Palliative Care Quality Collaborative (PCQC) national data registry.
- The developer further notes that the measure is used as a quality metric in the Public Hospital Redesign and Incentives in MEdi-Cal (PRIME) pay-for-performance program to enhance care delivery and quality across California health care systems.

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

Feedback on the measure by those being measured or others

The developer notes the following:

- PCQC provides on-demand structured reports of quality-of-care data to all participating palliative care organizations along with on demand explanation of metrics and opportunities to participate in quality improvement collaboratives.
- PRIME provides structured reports of quality metrics according to California state law and regulations.
- It is unclear from the developer's submission how feedback was obtained from those being measures, the type of feedback, and how the feedback has been considered when developing or revising the measure specifications.
- It is unclear from the submission if the data is reported publicly.

***Questions for the Committee**:*

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

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

It is unclear from the developer's submission if the performance results are publicly reported or if the data or performance results are publicly available. It is also unclear if there is a feedback loop in place for those being measured.

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

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

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

Improvement results

- The developer provided performance scores from the Hospice Quality Reporting Program (HQRP) for FY15 during the last measure evaluation in 2016.
    - The mean score was 98.0 percent with a range from 0 to 100 percent, the median was 100 percent, the interquartile range (IQR) was 1.5, and the standard deviation was 6.3.
    - The developer notes that 53.5 percent of hospices had perfect scores and 4.6 percent of hospices scored below 90 percent.
    - The developer did not provide data specific to palliative care for clinician group/practice in the hospital setting.

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

Unexpected findings (positive or negative) during implementation

- The developer did not report any unexpected findings.

Potential harms

- The developer did not report any unintended consequences.

***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: Usability is rated low for the hospice setting as there is little opportunity for improvement. It is unclear from the improvement results (previous data is from 2016) whether progress toward achieving the goal of high-quality, efficient care for individuals and populations is demonstrated. The results provided also do not specify which level of analysis.

Committee Pre-evaluation Comments:

4a. Use

- The measure does not appear to be publicly reported and there is a lack of clarity regarding feedback by those being measured or others. These are two important concerns for the Committee to consider.
- Data is collected by both the Palliative Care Quality Collaborative and the PRIME which is a funded California program aimed at improving care in a select group of hospitals over the next few years. In both of these case, I believe the data is available only to a limited number of providers. The developers provide no indication that the data is being used by either of these programs. Also, given the very high percentage of compliance with this measure, it is not clear how data from this measure would be actionable in a quality improvement program. It is not clear that any feedback has been solicited or provided
- This seems mostly a hospice measure and I wonder if it shouldn't be renamed accordingly. The name along is not clear as to what this refers to.

4a. Usability

- Opportunity for improvement with regard to the hospice setting appears to be small.
- Developer reports that there are now multiple years of data available to CMS. There is no discussion of or data presented on improvement over time or whether such analysis has been attempted. I concur with the staff opinion that the measure does not pass on usability
- No harms identified

Criterion 5: Related and Competing Measures

Related measures

- NQF #0326 Advance Care Plan
- NQF #1626 Patients Admitted to ICU who Have Care Preferences Documented

Harmonization

- The developer states that the specifications are harmonized to the extent possible.

Committee Pre-evaluation Comments:

5: Related and Competing Measures

- My answer to all these questions is 'none that I am aware of.'
- NQF #326 Advance Care Plan and NQF # 1626 Patients Admitted to ICU with Care Preferences Documented. Developers report that these measures are harmonized as far as possible.
- There are related advance care planning measures but not for this population.

Public and NQF Member Comments (Submitted as of June 17, 2022)

Member Expression of Support

- One NQF member submitted an expression of "do not support" for the measure.

Comments

Comment 1 by: Submitted by Koryn Rubin, American Medical Association

The American Medical Association (AMA) requests clarification on whether this measure is truly intended to be used at the group/practice level. On review of the measure testing and its current and planned future uses, we believe that the measure has only been tested and implemented for use at the facility level. We ask that the Committee consider this inconsistency and ensure that the measure is only endorsed at the appropriate level(s) of analysis.

Scientific Acceptability Evaluation

RELIABILITY: SPECIFICATIONS

1. **Have measure specifications changed since the last review?** ☐ **Yes** ☒ **No**
2. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?** ☒ **Yes** ☐ **No**
3. **Briefly summarize any changes to the measure specifications and/or concerns about the measure specifications.**
   - Measure specifications have not changed since last review.

RELIABILITY: TESTING

4. **Did the developer conduct new reliability testing?** ☐ **Yes** ☒ **No**

   4a. If no, summarize the Standing Committee's previous feedback:
   - The developer provided inter-rater reliability analysis based on data from 20 patients in one hospital (kappa=1.0) during the 2012 endorsement evaluation.
   - The developer provided updated reliability testing during the 2016 evaluation, developers updated by providing score-level testing results for the hospice setting from two different analyses (a split-half analysis with an intra-class correlation coefficient of 0.91, and a signal-to-noise analysis with a signal-to-noise ratio of 0.98).
   - The Standing Committee voiced no concerns regarding reliability for the hospice setting, but strongly recommended that reliability testing for the clinician-level measure in the palliative care setting be updated and the results provided to NQF when available.

   4b. If yes, describe any differences between the new and old testing and summarize any relevant Standing Committee's feedback from the previous review:
5. **Reliability testing level:** ☒ **Accountable-Entity Level** ☒ **Patient/Encounter Level** ☐ **Neither**

6. **Reliability testing was conducted with the data source and level of analysis indicated for this measure:**

   ☐ **Yes** ☒ **No**

7. If accountable-entity level and/or patient/encounter level reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of patient-level data conducted?

   ☐ **Yes** ☐ **No**

8. **Assess the method(s) used for reliability testing:**
   - The developer conducted inter-rater reliability testing using two nurse abstractors and a structured chart abstraction tool (n=20 seriously ill patients, kappa 1.0).
   - The developer conducted split-half reliability analysis on all facilities with 20 or more patients stays.
   - An analysis of variance (ANOVA) was conducted to determine what proportion of total variance in the measure is attributable to differences among providers.
   - A stability analysis was conducted to determine, to what extent a provider's performance changes across time.

9. **Assess the results of reliability testing**
   - Split-half reliability analysis: The intraclass coefficient (ICC) is 0.91, indicating high internal reliability.
   - Analysis of variance (ANOVA): The signal-to-noise ratio was 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 developer indicated that slightly less than 95% of facilities had a change in quality measure scoring of <1 standard deviation across quarter ranging from 93.5 to 93.9 to 93.6.

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

    ☒ **Yes** ☐ **No** ☐ **Not applicable**

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

    ☒ **Yes** ☐ **No** ☐ **Not applicable** (patient/encounter level testing was not performed)

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

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

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

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

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

13. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**
    - Facility level: Precise specifications (Box 1) -> Empirical reliability testing with measure as specified (Box 2) -> Score-level testing (Box 4) -> Appropriate method (Box 5) -> Moderate certainty that measure results are reliable (Box 6b) -> Moderate
    - Clinician: Group/Practice level: Precise specifications (Box 1) -> Empirical reliability testing with measure as specified (Box 2) -> Empirical reliability testing for EACH level of analysis in the tested population is not provided (Box 4)-> Empirical reliability testing conducted on all data elements (Box 8) -> Inter-abstractor (rater) agreement (kappa for overall agreement only assessed) (Box 9)-> Insufficient rating

VALIDITY: TESTING

14. **Did the developer conduct new validity testing?** ☐ **Yes** ☒ **No**

14a. If no, summarize the Standing Committee's previous feedback:

- Validity testing at the time of the 2012 endorsement evaluation included both a face validity assessment and a score-level construct validity analysis for the palliative care setting. Results from this analysis indicated documenting treatment preferences was statistically significantly different for seriously ill patients seen in specialty interdisciplinary palliative care consultations (91%) compared to those who did not receive these services (59%).
- For the 2016 evaluation, developers updated their validity testing by conducting a nonparametric Spearman rank correlation analysis between this measure and 5 other hospice quality measures. Although the magnitude of the correlations from this analysis was lower than expected, they were positive and statistically significant, confirming the developers' hypothesis that hospice agencies perform similarly on various assessment activities conducted at the time of hospice admission.
- The Standing Committee members did not express concern regarding the validity of the measure.

14b. If yes, describe any differences between the new and old testing and summarize any relevant Standing Committee's feedback from the previous review:

15. **Validity testing level (check all that apply):**

☒ **Accountable-Entity Level** ☐ **Patient or Encounter-Level** ☐ **Both**

NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

16. **If patient/encounter level validity testing was provided, was the method described and appropriate for assessing the accuracy of ALL critical data elements? NOTE:** Data element validation from the literature is acceptable.

☐ **Yes**

☐ **No**

☒ **Not applicable** (patient/encounter level testing was not performed)

17. **Method of establishing validity at the accountable-entity level:**

☒ **Face validity**

☒ **Empirical validity testing at the accountable-entity level**

☐ **N/A (accountable-entity level testing not conducted)**

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

☒ **Yes**

☐ **No**

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

19. **Assess the method(s) for establishing validity**

- 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.
- The developer performed construct validity testing by comparing documentation of life sustaining preferences among seriously ill patients receiving specialty interdisciplinary palliative care and patients who did not in a facility setting.
- The developers conducted non-parametric Spearman rank correlation analysis using FY15 HQRS data (n=3,992 hospice organizations; n= 1,218,786 patient stays) between this measure and five additional hospice quality measures.

20. **Assess the results(s) for establishing validity**
    - The developer does not provide results of a systemic review of face validity for 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).
    - The developer indicates that facility-level correlations were positive and statistically significant but does not provide results from statistical testing to support this statement.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

21. **Please describe any concerns you have with measure exclusions.**
    - The measure does not use exclusions.
22. **Risk Adjustment**

    22a. **Risk-adjustment method**

    ☒ None (only answer Question 20b and 20e) ☐ Statistical model ☐ Stratification

    ☐ Other method assessing risk factors (please specify)

    22b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

    ☐ Yes ☒ No ☐ Not applicable

    22c. **Social risk adjustment:**

    22c.1 Are social risk factors included in risk model? ☐ Yes ☐ No ☒ Not applicable

    22c.2 Conceptual rationale for social risk factors included? ☐ Yes ☐ No

    22c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☐ Yes ☐ No

    22d.**Risk adjustment summary:**

    22d.1 All of the risk-adjustment variables present at the start of care? ☐ Yes ☐ No
    22d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐ Yes ☐ No
    22d.3 Is the risk adjustment approach appropriately developed and assessed? ☐ Yes ☐ No
    22d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☐ Yes ☐ No
    22d.5.Appropriate risk-adjustment strategy included in the measure? ☐ Yes ☐ No

    22e**. Assess the risk-adjustment approach**
    - The measure is not risk adjusted or stratified.
23. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.**
    - The developer found that 23.3 percent of hospices were more likely to report scores above the national mean of 98.0 percent compared to 9.9 percent of those hospices that were below the national mean.
    - The developer also noted that 53.5 percent of hospitals had perfect measure scores (100 percent) and 4.6 hospices scored below 90 percent.
24. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
    - The measure only uses one set of specifications for this measure.
25. **Please describe any concerns you have regarding missing data.**

- The developer conducted analyses on two levels: eligible patient stays and hospice level. The developer examined three items on the HIS that can include missing data and impact the validity of the measure. This analysis was conducted at the patient stay and hospice level.
    - The developer noted that the overall rate of missing data at the patient-level ranged from 0.01 to 0.02 percent.
    - At the hospice level, over 95 percent of hospices did not have any admissions with missing data related to these three items.

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

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

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

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

    ☒ **Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the accountable-entity level and the patient/encounter level is required; if not conducted, should rate as INSUFFICIENT.)

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

    - Facility level: Potential threats to validity assessed (Box 1) -> Empirical validity testing conducted using the measure as specified and appropriate statistical testing (Box 2) -> Validity testing conducted with computed performance measure scores of each measured entity (Box 5) -> Method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships (Box 6) -> Moderate certainty or confidence (Box 7b) -> Moderate rating
    - Clinician: Group/Practice level: Potential threats to validity assessed (Box 1) -> Empirical validity testing conducted using the measure as specified and appropriate statistical testing (Box 2) -> Validity testing not conducted for EACH level of analysis in the tested population (Box 5)-> Empirical validity not conducted on all critical data elements (Box 9) -> Method not appropriate for assessing the accuracy of all critical patient/encounter data elements (Box 10) -> Insufficient rating

ADDITIONAL RECOMMENDATIONS

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

Developer Submission

Criteria 1: Importance to Measure and Report

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

1a. Evidence

1ma.01. Indicate whether there is new evidence about the measure since the most recent maintenance evaluation. If yes, please briefly summarize the new evidence, and ensure you have updated entries in the Evidence section as needed.

[Response Begins]
Yes
[Yes, Please Explain]
NQF 1641 was used as a quality metric in the Public hospital Redesign and Incentives in MEdi-Cal (PRIME) program.
Summary data from 52 California hospitals indicates achievement of this quality metric at the organizational level:
mean 82.9%
median 89.4%
range (0-100%)

[Response Ends]

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Measure and Report: Evidence section. For example:
2021 Submission:
Updated evidence information here.
2018 Submission:
Evidence from the previous submission here.

1a.01. Provide a logic model.

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

[Response Begins]
2016 Submission:
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. (7-13)

[Response Ends]

1a.02. Select the type of source for the systematic review of the body of evidence that supports the performance measure.

A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data.
[Response Begins]
Clinical Practice Guideline recommendation (with evidence review)
[Response Ends]

If the evidence is not based on a systematic review, skip to the end of the section and do not complete the repeatable question group below. If you wish to include more than one systematic review, add additional tables by clicking "Add" after the final question in the group.

Evidence - Systematic Reviews Table (Repeatable)
Group 1 - Evidence - Systematic Reviews Table

1a.03. Provide the title, author, date, citation (including page number) and URL for the systematic review.

[Response Begins]
Institute for Clinical Systems Improvement Palliative Care Guideline Workgroup. Palliative Care for Adults. Sixth edition. Updated January 2020. p.48
https://www.icsi.org/wp-content/uploads/2021/11/PalliativeCare_6th-Ed_2020_v2.pdf

[Response Ends]

1a.04. Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the systematic review.

[Response Begins]
Aim 4: Improve the accessibility and ongoing utilization of the patient's palliative care plan, health care directive, patient values and preferences, and involvement of the health care proxy across the continuum of care (inpatient, ED, outpatient, home care, senior residence, etc.).
Recommendation: Facilitation of advance care planning conversations is appropriate for all adult patients. Regular review of goals and wishes should occur as the patient's condition or life circumstances change.

[Response Ends]

1a.05. Provide the grade assigned to the evidence associated with the recommendation, and include the definition of the grade.

[Response Begins]
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
[Response Ends]

1a.06. Provide all other grades and definitions from the evidence grading system.

[Response Begins]
High quality: Further research is very unlikely to change our confidence in the estimate of effect
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
Very low quality: Any estimate of effect is very uncertain

[Response Ends]

1a.07. Provide the grade assigned to the recommendation, with definition of the grade.

[Response Begins]
Strength of recommendation: Strong
The GRADE system offers two grades of recommendations: “strong” and “weak”. When the desirable effects of an intervention clearly outweigh the undesirable effects, or clearly do not, guideline panels offer strong recommendations.

[Response Ends]

1a.08. Provide all other grades and definitions from the recommendation grading system.

[Response Begins]
ICSI utilizes the Grading of Recommendations Assessment, Development and Evaluation (GRADE) methodology system. GRADE involves systematically evaluating the quality of evidence (high, moderate, low, very low) and developing a strength of recommendation (strong, weak). For more detailed information on GRADE, please visit http://www.gradeworkinggroup.org/. In addition, when GRADE methodology could not be applied, the work group developed consensus recommendations.
[Response Ends]

1a.09. Detail the quantity (how many studies) and quality (the type of studies) of the evidence.

[Response Begins]
ICSI identified 63 references supporting their formal recommendations. Quality of evidence was rated low.
[Response Ends]

1a.10. Provide the estimates of benefit, and consistency across studies.

[Response Begins]
Regular review of ACP ensures patient wishes for treatment are accurately documented and family understands the benefits and burdens of available treatment options.
[Response Ends]

1a.11. Indicate what, if any, harms were identified in the study.

[Response Begins]
Opportunity costs and limited available resources may be a barrier. Systems may have difficulty capturing, storing and accessing ACP documents when needed. Although it is difficult to have this detailed discussion with all adult patients, it is extremely helpful to patients, family and clinicians in the event that the patient cannot express his/her desires.
[Response Ends]

1a.12. Identify any new studies conducted since the systematic review, and indicate whether the new studies change the conclusions from the systematic review.

[Response Begins]
N/A
[Response Ends]

1a.13. If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, describe the evidence on which you are basing the performance measure.

[Response Begins]
[Response Ends]

1a.14. Briefly synthesize the evidence that supports the measure.

[Response Begins]
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 address hospice patients, whose broader treatment preferences must be addressed in advance of hospice care in order to elect that option.
[Response Ends]

1a.15. Detail the process used to identify the evidence.

[Response Begins]
[Response Ends]

1a.16. Provide the citation(s) for the evidence.

[Response Begins]
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.
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.

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.
14. 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.
15. Institute for Clinical Systems Improvement Palliative Care Guideline Workgroup. Palliative Care for Adults. Sixth edition. Updated January 2020. p.48

[Response Ends]

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

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

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

[Response Begins]
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.
[Response Ends]

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

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

[Response Begins]
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.
[Response Ends]

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

[Response Begins]
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.
[Response Ends]

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

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

[Response Begins]
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.
[Response Ends]

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

[Response Begins]
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.
[Response Ends]

Criteria 2: Scientific Acceptability of Measure Properties

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

spma.01. Indicate whether there are changes to the specifications since the last updates/submission. If yes, update the specifications in the Measure Specifications section of the Measure Submission Form, and explain your reasoning for the changes below.

[Response Begins]
No
[Response Ends]

spma.02. Briefly describe any important changes to the measure specifications since the last measure update and provide a rationale.

For annual updates, please explain how the change in specifications affects the measure results. If a material change in specification is identified, data from re-testing of the measure with the new specifications is required for early maintenance review.

For example, specifications may have been updated based on suggestions from a previous NQF CDP review.

[Response Begins]
We removed 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.

[Response Ends]

sp.01. Provide the measure title.

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

[Response Begins]
Hospice and Palliative Care – Treatment Preferences
[Response Ends]

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

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

[Response Begins]
Percentage of patients with chart documentation of preferences for life sustaining treatments.
[Response Ends]

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

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.
Please do not select:

- *Surgery: General*

[Response Begins]
Palliative Care and End-of-Life Care
[Response Ends]

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

[Response Begins]
Person-and Family-Centered Care: Person-and Family-Centered Care
[Response Ends]

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

Select only those target populations which can be stratified in the reporting of the measure's result.
Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.
Please do not select:

- *Populations at Risk: Populations at Risk*

[Response Begins]
Elderly (Age >= 65)
[Response Ends]

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

Check ONLY the levels of analysis for which the measure is SPECIFIED and TESTED.
Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.
Please do not select:

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

[Response Begins]
Clinician: Group/Practice
Facility
[Response Ends]

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

Check ONLY the settings for which the measure is SPECIFIED and TESTED.
[Response Begins]
Home Care
Inpatient/Hospital
[Response Ends]

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

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

[Response Begins]
PEACE Hospice and Palliative Care Quality Measures: http://www.med.unc.edu/pcare/resources/PEACE-Quality-Measures
[Response Ends]

sp.11. Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). Excel formats (.xlsx or .csv) are preferred.

Attach an excel or csv file; if this poses an issue, contact staff. Provide descriptors for any codes. Use one file with multiple worksheets, if needed.
[Response Begins]
No data dictionary/code table – all information provided in the submission form
[Response Ends]

sp.12. State the numerator.

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

[Response Begins]
Patients whose medical record includes documentation of life sustaining preferences
[Response Ends]

sp.13. Provide details needed to calculate the numerator.

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

[Response Begins]
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.
[Response Ends]

sp.14. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]
Seriously ill patients enrolled in hospice OR receiving specialty palliative care in an acute hospital setting.
[Response Ends]

sp.15. Provide details needed to calculate the denominator.

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

[Response Begins]
The 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.
[Response Ends]

sp.16. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]
There are no denominator exclusions for this measure.
[Response Ends]

sp.17. Provide details needed to calculate the denominator exclusions.

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

[Response Begins]
N/A
[Response Ends]

sp.18. Provide all information required to stratify the measure results, if necessary.

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

[Response Begins]
N/A
[Response Ends]

sp.19. Select the risk adjustment type.

Select type. Provide specifications for risk stratification and/or risk models in the Scientific Acceptability section.
[Response Begins]
No risk adjustment or risk stratification
[Response Ends]

sp.20. Select the most relevant type of score.

Attachment: If available, please provide a sample report.
[Response Begins]
Rate/proportion
[Response Ends]

sp.21. Select the appropriate interpretation of the measure score.

Classifies interpretation of score according to whether better quality or resource use is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score
[Response Begins]
Better quality = Higher score
[Response Ends]

sp.22. Diagram or describe the calculation of the measure score as an ordered sequence of steps.

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

[Response Begins]
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
[Response Ends]

sp.25. If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.

[Response Begins]
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.
[Response Ends]

sp.28. Select only the data sources for which the measure is specified.

[Response Begins]
Assessment Data
Electronic Health Records
[Response Ends]

sp.29. Identify the specific data source or data collection instrument.

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

[Response Begins]
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
[Response Ends]

sp.30. Provide the data collection instrument.

[Response Begins]
No data collection instrument provided
[Response Ends]

2ma.01. Indicate whether additional empirical reliability testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Reliability - Testing. Include information on all testing conducted (prior testing as well as any new testing).

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:
Current Submission:
Updated testing information here.
Previous Submission:
Testing from the previous submission here.

[Response Begins]
No
[Response Ends]

2ma.02. Indicate whether additional empirical validity testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Validity - Testing. Include information on all testing conducted (prior testing as well as any new testing).

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:
Current Submission:
Updated testing information here.
Previous Submission:
Testing from the previous submission here.

[Response Begins]
No
[Response Ends]

2ma.03. For outcome, patient-reported outcome, resource use, cost, and some process measures, risk adjustment/stratification may be conducted. Did you perform a risk adjustment or stratification analysis?

[Response Begins]
No
[Response Ends]

2ma.04. For maintenance measures in which risk adjustment/stratification has been performed, indicate whether additional risk adjustment testing has been conducted since the most recent maintenance evaluation. This may include updates to the risk adjustment analysis with additional clinical, demographic, and social risk factors.

Please update the Scientific Acceptability: Validity - Other Threats to Validity section.

Note: This section must be updated even if social risk factors are not included in the risk adjustment strategy.

[Response Begins]
No additional risk adjustment analysis included
[Response Ends]

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

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

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

2a. Reliability testing demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.
2b1. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.
2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure;
AND
If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).
2b3. For outcome measures and other measures when indicated (e.g., resource use):
• an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; 14,15 and has demonstrated adequate discrimination and calibration
OR
• rationale/data support no risk adjustment/ stratification.
2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful 16 differences in performance;
OR
there is evidence of overall less-than-optimal performance.
2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.
2b6. Analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.
2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:
2c1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and
2c2. the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.
(if not conducted or results not adequate, justification must be submitted and accepted)

Definitions
Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).
Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.
Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.
Risk factors that influence outcomes should not be specified as exclusions.

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

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Scientific Acceptability sections. For example:
2021 Submission:
Updated testing information here.
2018 Submission:
Testing from the previous submission here.

2a. Reliability

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

[Response Begins]
Assessment Data
Electronic Health Records
[Response Ends]

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

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

[Response Begins]
2021 Submission:
No updated testing information
2016 Submission:
Hospice: Hospice analysis uses the Hospice Item Set (HIS) as the data source to calculate the quality measure.

[Response Ends]

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

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

[Response Begins]
Hospice: 10-01-2014 - 09-30-2015
[Response Ends]

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

Testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan.
Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.
Please do not select:

- *Clinician: Clinician*

- *Population: Population*

[Response Begins]
Clinician: Group/Practice
Facility
[Response Ends]

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

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

[Response Begins]
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.
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%
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.

[Response Ends]

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

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

[Response Begins]
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%).

[Response Ends]

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

[Response Begins]
N/A
[Response Ends]

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

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

[Response Begins]
Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.07 check patient or encounter-level data; in 2a.08 enter "see validity testing section of data elements"; and enter "N/A" for 2a.09 and 2a.10.
[Response Ends]

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.07 check patient or encounter-level data; in 2a.08 enter "see validity testing section of data elements"; and enter "N/A" for 2a.09 and 2a.10.

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

Choose one or both levels.
[Response Begins]
Patient or Encounter-Level (e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements)
Accountable Entity Level (e.g., signal-to-noise analysis)
[Response Ends]

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

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

[Response Begins]

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 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.

[Response Ends]

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

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

[Response Begins]
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.

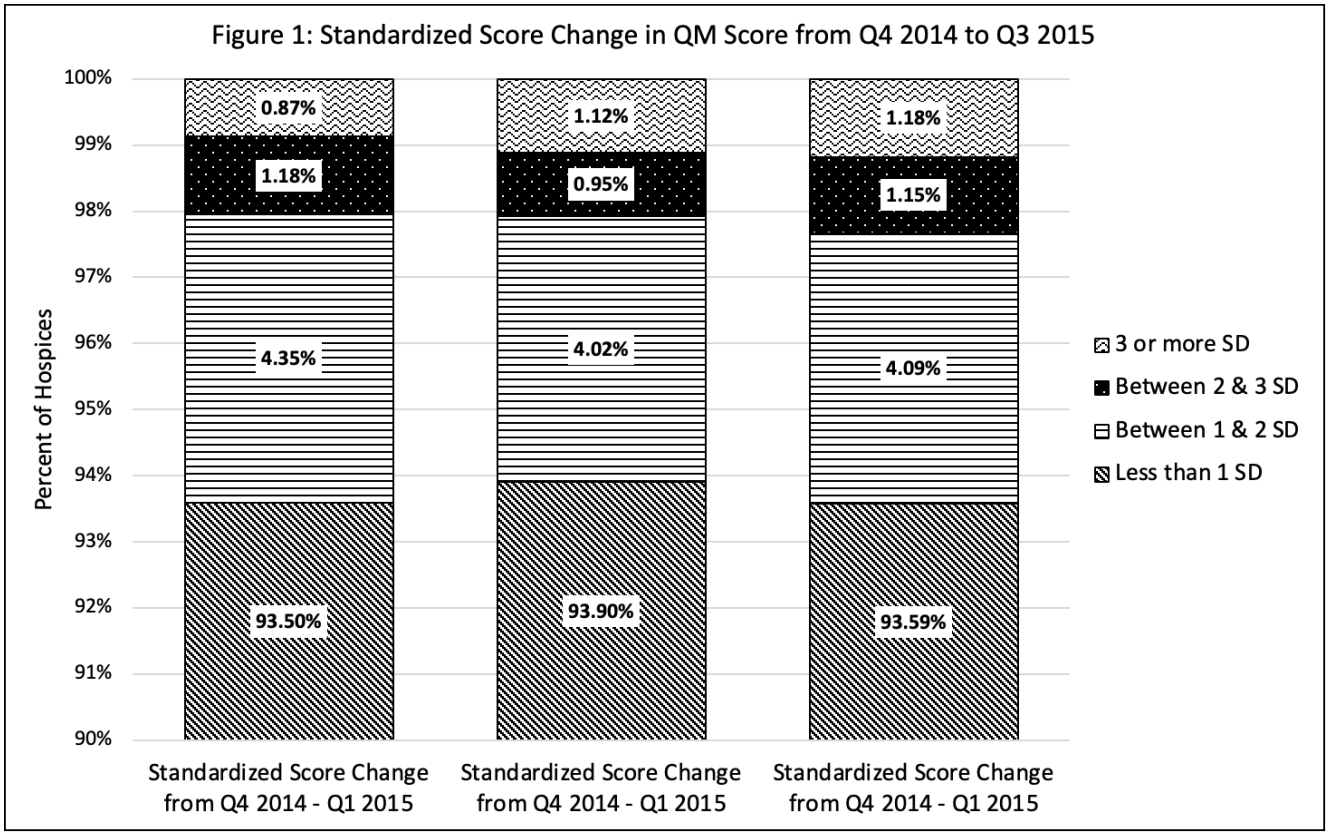

Figure 1 is a stacked bar chart illustrating the percent of standardized score change in facility scores between Q4 of 2014 to Q1 2015, Q1 to Q2 2015, and Q2 to Q3 2015. Slightly less than 95% of facilities had a change in QM score of less than one standard deviation, indicating high stability of the QM.

Palliative Care: Kappa scores range from 0 to 1 with higher scores indicating better agreement. The 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.

[Response Ends]

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

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

[Response Begins]
Hospice: The ICC coefficient for this measure is 0.91, indicating high internal reliability. 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 indicated that facility scores were very stable and that the measure is generally quite stable.
[Response Ends]

2b. Validity

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

[Response Begins]
Accountable Entity Level (e.g. hospitals, clinicians)
Empirical validity testing
[Response Ends]

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

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

[Response Begins]
In the prior NQF endorsement review, the measure was tested at the Clinician Group/Practice level for the palliative care hospital setting using face and construct validity. 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. Construct validity was tested by comparing measure results for seriously ill patients seen in specialty interdisciplinary palliative care consultations in one hospital to those who did not receive these services.
Additionally, at the facility level in the hospice setting, using FY15 data from the Hospice Quality Reporting System (n=3,992 hospice organizations; n=1,218,786 patient stays), we conducted non-parametric Spearman rank correlation analysis between this measure and 5 other hospice quality measures - NQF #1634, 1637, 1638, 1639, and 1647. We hypothesized that agencies should perform similarly on assessment processes at hospice admission and expected the resulting correlations to be high.

[Response Ends]

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

Examples may include correlations or t-test results.

[Response Begins]
In the prior NQF endorsement review:
Facility-level/Hospice: Overall, this measure had 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 (Spearman's rho = 0.64). We expected the strong correlation between these two QMs because they both address the competency of the hospice to solicit the patient's preferences.
Palliative Care: Face validity results from the initial stakeholder group of nursing and physician leaders from MICU, SICU, geriatrics, oncology, and palliative care indicated broad endorsement of the face validity of the measure. 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.
Face validity was further affirmed by selection of this measure for endorsement by the American Academy of Hospice and Palliative Medicine and Hospice and Palliative Nursing Association through the expert panel guiding the Measuring What Matters project. (http://aahpm.org/quality/measuring-what-matters)
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%, $p<0.001$).

[Response Ends]

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

[Response Begins]
Facility-level validity for hospice is established due to strong correlation coefficient (Spearman's rho = 0.64) between quality measures requiring elicitation of patient values for treatment preferences and for spiritual preferences.
Face validity findings and construct validity demonstrate this quality measure has broad stakeholder support including at the national level and correlates with access to specialty palliative care.

[Response Ends]

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

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

[Response Begins]
Prior analysis used 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.

[Response Ends]

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

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

[Response Begins]
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).

[Response Ends]

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

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

[Response Begins]
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).
[Response Ends]

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

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

[Response Begins]

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.
[Response Ends]

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

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

[Response Begins]
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.

[Response Ends]

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

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

[Response Begins]
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.
[Response Ends]

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

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

[Response Begins]
No, there is only one set of specifications for this measure
[Response Ends]

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

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

[Response Begins]
[Response Ends]

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

Examples may include correlation, and/or rank order.

[Response Begins]
[Response Ends]

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

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

[Response Begins]
[Response Ends]

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

[Response Begins]
N/A or no exclusions
[Response Ends]

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

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

[Response Begins]
n/a
[Response Ends]

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

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

[Response Begins]
n/a
[Response Ends]

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

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

[Response Begins]
n/a
[Response Ends]

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

[Response Begins]
No risk adjustment or stratification
[Response Ends]

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

[Response Begins]
[Response Ends]

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

[Response Begins]
n/a
[Response Ends]

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

[Response Begins]
[Response Ends]

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

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

[Response Begins]
[Response Ends]

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

[Response Begins]
[Response Ends]

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

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

[Response Begins]
[Response Ends]

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

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

[Response Begins]
[Response Ends]

2b.27. Provide risk model discrimination statistics.

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

[Response Begins]
[Response Ends]

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

[Response Begins]
n/a
[Response Ends]

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

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

[Response Begins]
[Response Ends]

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

[Response Begins]
[Response Ends]

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

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

[Response Begins]
[Response Ends]

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

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

[Response Begins]
[Response Ends]

Criteria 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.

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

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

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

In other words, indicate whether data elements that are needed to compute the performance measure score are in defined, computer-readable fields.
[Response Begins]
ALL data elements are in defined fields in a combination of electronic sources
[Response Ends]

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

[Response Begins]
N/A
[Response Ends]

3.04. Describe any efforts to develop an eCQM.

[Response Begins]
Hospice EMR vendors have embedded an electronic version of NQF 1641 to facilitate data collection as part of the CMS Hospice Item Set.
[Response Ends]

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

[Response Begins]
2016 Submission: 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.

[Response Ends]

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

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

Attach the fee schedule here, if applicable.

[Response Begins]
N/A
[Response Ends]

Criteria 4: Use and Usability

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.

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making.
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 demonstrating performance improvement.

4a. Use

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

Name of program and sponsor

URL

Purpose

Geographic area and number and percentage of accountable entities and patients included

Level of measurement and setting

[Response Begins]
Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
[Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Please Explain]
NQF 1641 is one of the recommended quality measures in AAHPM/HPNA Measuring What Matters (MWM) project. It is one of 10 quality measures recommended for nationwide use in hospice and palliative care in the United States. MWM quality measures are embedded in Palliative Care Quality Collaborative (PCQC) national registry data.
In addition, it was chosen as a quality metric for the Public hospital Redesign and Incentives in MEdi-Cal (PRIME) program to enhance care delivery and quality across these California organizations.

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]
Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
[Response Ends]

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

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

[Response Begins]
[Response Ends]

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

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

[Response Begins]
[Response Ends]

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

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

[Response Begins]
PCQC provides on-demand structured reports of quality of care data to all participating palliative care organizations.
PRIME provides structured reports of quality metrics to participating public hospitals in California.

[Response Ends]

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

[Response Begins]
PCQC provides on-demand structured reports of quality of care data to all participating palliative care organizations along with on demand explanation of metrics and opportunities to participate in quality improvement collaboratives.
PRIME provides structured reports of quality metrics according to California state law and regulations.

[Response Ends]

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

[Response Begins]
PCQC and PRIME control feedback and implementation processes.
[Response Ends]

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

[Response Begins]
PCQC and PRIME control feedback and implementation processes.
[Response Ends]

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

[Response Begins]
PCQC and PRIME control feedback and implementation processes.
[Response Ends]

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

[Response Begins]
PCQC and PRIME control feedback and implementation processes. The measure has not been modified.
[Response Ends]

4b. Usability

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

[Response Begins]
In the prior submission we had only one year of data to report, which is not enough to show trends over time.
Due to the use of this quality measure in national hospice data (Hospice Item Set) multiple years of data are available to CMS.

[Response Ends]

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

[Response Begins]
None
[Response Ends]

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

[Response Begins]
None
[Response Ends]

Criteria 5: Related and 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.

If you are updating a maintenance measure submission for the first time in MIMS, please note that the previous related and competing data appearing in question 5.03 may need to be entered in to 5.01 and 5.02, if the measures are NQF endorsed. Please review and update questions 5.01, 5.02, and 5.03 accordingly.

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

(Can search and select measures.)
[Response Begins]
0326: Advance Care Plan
[Response Ends]

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

(Can search and select measures.)
[Response Begins]
[Response Ends]

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

[Response Begins]
N/A
[Response Ends]

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

[Response Begins]
Yes
[Response Ends]

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

[Response Begins]
N/A
[Response Ends]

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

Provide analyses when possible.

[Response Begins]
This measure is superior to 0326 because it includes all age groups receiving hospice or palliative care and is not limited to age 65 and older. In addition, this measure is inclusive of advance care plans noted in 0326 and also includes discussion of current goals of care and current treatment preferences.

[Response Ends]

Appendix

Supplemental materials may be provided in an appendix.:

No appendix

Contact Information

Measure Steward (Intellectual Property Owner): University of North Carolina-Chapel Hill
Measure Steward Point of Contact: Hanson, Laura, lhanson@med.unc.edu
Wessell, Kathryn, kwessell@ad.unc.edu

Measure Developer if different from Measure Steward: University of North Carolina-Chapel Hill
Measure Developer Point(s) of Contact: Hanson, Laura, lhanson@med.unc.edu