NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

Purple text represents the responses from measure developers.

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

Brief Measure Information

NQF #: 3235

Corresponding Measures:

De.2. Measure Title: Hospice and Palliative Care Composite Process Measure—Comprehensive Assessment at Admission

Co.1.1. Measure Steward: Centers for Medicare & Medicaid Services

De.3. Brief Description of Measure: The Hospice Comprehensive Assessment Measure assesses the percentage of hospice stays in which patients who received a comprehensive patient assessment at hospice admission. The measure focuses on hospice patients aged 18 years and older. A total of seven individual NQF endorsed component quality will provide the source data for this comprehensive assessment measure, including NQF #1634, NQF #1637, NQF #1639, NQF #1638, NQF #1617, NQF #1641, and NQF #1647. These seven measures are currently implemented in the CMS HQRP. These seven measures focus on care processes around hospice admission that are clinically recommended or required in the hospice Conditions of Participation, including patient preferences regarding life-sustaining treatments, care for spiritual and existential concerns, and management of pain, dyspnea, and bowels.

1b.1. Developer Rationale: The aim of this measure is to assess whether a comprehensive assessment is completed at hospice admission for each hospice patient based on seven QMs that assess high-priority care processes around admission as recognized by both leading hospice stakeholders and patients. Another key factor in creating a measure of comprehensive assessment at admission is to provide both consumers and providers with a single measure regarding the overall quality of the assessment of patient needs at hospice admission, which can then be easily used to compare quality across providers. Additionally, in the current HQRP QMs hospice performance scores are high with scores of 97% or higher; hospices perform lower on the Pain Assessment QM (95.3%). On average, 93.2% of patient stays in a hospice had documentation that all of these critical care processes were completed at admission. Thus, the comprehensive assessment measure sets a higher standard of care for hospices, and consequently reveals a larger performance gap. The performance gap identified by the comprehensive assessment measure creates opportunities for quality improvement and may motivate providers to conduct a greater number of high priority care processes for as many patients as possible upon admission.

S.4. Numerator Statement: The numerator of this measure is the number of patient stays in the denominator where the patient received all 7 care processes which are applicable to the patient at admission, as captured by the current HQRP quality measures. To be included in the comprehensive assessment measure numerator, a patient must meet the numerator criteria for each of the individual component quality measure (QM) that is

applicable to the patient. The numerator of this measure accounts for the three conditional measures in the current HQRP (NQF #1637 Pain Assessment, NQF #1638 Dyspnea Treatment, and NQF #1617 Bowel Regimen) as described below.

S.6. Denominator Statement: The denominator for the measure includes all hospice patient stays enrolled in hospice except those with exclusions.

S.8. Denominator Exclusions: Patient stays are excluded from the measure if they are under 18 years of age, or are a Type 2 (discharged stays missing the admission record) or Type 3 patient stay (active stays).

De.1. Measure Type: Composite

S.17. Data Source: Other

S.20. Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Jul 13, 2017 **Most Recent Endorsement Date:** Jul 13, 2017

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

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

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

Preliminary Analysis: Maintenance of Endorsement

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

Criteria 1: Importance to Measure and Report

- 1a. Evidence

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

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

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 2017

- The developer provided a rationale for the relationship between a comprehensive assessment of physical and psychosocial well-being and positive treatment outcomes, including improved quality of life, treatment consistent with preferences, improved management of symptoms including constipation, pain, and dyspnea, and meeting spiritual care needs.

- The developer cited the 2013 Institute for Clinical Systems Improvement (ICSI) Palliative Care for Adults guidelines to support the components in the composite. All of the recommendation statements from the ICSI guideline refer to inclusion of the measured components in the palliative care plan.
- The Committee concluded that the evidence presented is tangential to the foci of the measure, which assesses actual screening, assessment, discussions, or treatment not simply inclusion of these processes in the palliative care plan. The Committee recognized the evidence base linking dyspnea treatment, bowel regimens, and communication regarding treatment preferences to improved patient outcomes. However, members acknowledged that similar evidence for the other components of the measure (pain screening, pain assessment, dyspnea screening, and addressing spiritual and religious concerns) does not exist and likely would not be forthcoming. The Committee agreed that empirical evidence is not needed to hold providers accountable for those components of the measure, and agreed to invoke the exception to the evidence subcriterion.

Changes to evidence from last review

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

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

Questions for the Committee:

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

Guidance from the Evidence Algorithm

Process measure but graded systematic review can be considered tangential evidence (Box 3) →Other evidence not submitted (Box 7) →Other than the individual measures that correspond to the 7 components in the measure, other measures are not available (Box 10) →The ICSI guideline recommendations could be considered a systematic assessment of expert opinion relevant to this measure, as could other evidence presented for the individual NQF-endorsed measures (Box 11) →If Committee agrees it is okay or beneficial to hold providers accountable for performance in the absence of empirical evidence of benefits to patients, rate as INSUFFICIENT WITH EXCEPTION.

Preliminary rating for evidence: ☐ **High** ☐ **Moderate** ☐ **Low** ☒ **Insufficient (with Exception)**

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

Maintenance measures – increased emphasis on gap and variation

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

- The developer provided data from the Hospice Quality Reporting Program (HQRP) representing four calendar years from 2016-2019. Over time, the hospice-level mean score increased from 77.8% for patient stays admitted in 2016 to 89.6% in 2019, the median increased from 82.3% to 94.1%, the interquartile range (IQR) decreased from 22.9% to 12.0%, and the standard deviation (SD) decreased from 18.2% to 12.7%. For patient stays admitted between January 1, 2016 and December 31, 2016, only 2.2 percent of hospices (81 of 3,745 hospices) had perfect scores, and 34.6% of hospices scored lower than 75% on this QM. In 2019, 8.4% percent of hospices (344 of 4,080 hospices) had perfect scores, and 11.0% of hospices scored lower than 75%.

Disparities

- The developer analyzed data from 4,089 hospices, including 1,422,921 patient stays as reflected below:
    - Proportion of Non-White: the developer stated that rates of completion of the seven care processes within this composite across racial identities was found to be statistically significant. The lowest rate consisted of other non-Hispanic patients (90.9%) and the highest rate consisted of white non-Hispanic patients.
    - Race/Ethnicity: The median hospital performance among hospitals with more non-white patients was 88.1% while among hospitals with less non-white patients it was higher at 90.6%.
    - Gender: The developer stated that the analysis of gender shows similar rates among male and female patients.

Questions for the Committee:

- Is there a gap in care that warrants a national performance measure?
- Do disparities exist in this area of healthcare?

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

1c. Composite – Quality Construct and Rationale

Maintenance measures – same emphasis on quality construct and rationale as for new measures.

1c. Composite Quality Construct and Rationale. The quality construct and rationale should be explicitly articulated and logical; a description of how the aggregation and weighting of the components is consistent with the quality construct and rationale also should be explicitly articulated and logical.

- Composite Type – This is an all-or-nothing composite measure.
- Quality construct – The developer states that this measure is comprised of seven NQF-endorsed quality measures: pain screening (NQF #1634) and comprehensive pain assessment (NQF #1637), dyspnea screening (NQF #1639) and treatment (NQF #1638), patients treated with an opioid who are given a bowel regimen (NQF #1617), patient preferences for life-sustaining treatments (NQF #1641), and spiritual and existential concerns (NQF #1647). This measure reflects the overall quality of comprehensive assessment at hospice admission for each patient stay.
- Rationale – The developer states that the aggregation of the seven measure components will (1) incentivize hospices to conduct all critical care processes for each patient; (2) set a higher standard of care for hospices, which will reveal a larger performance gap and thus room for improvement (discussed below); and (3) provide consumers and providers with a single measure representing the overall quality and completeness of assessment of patient needs at hospice admission, which can be easily used to compare quality processes across providers.
- Aggregation method – the developer states that this measure calculates the percentage of patients who received all seven HQRP care processes at admission.
- All seven components are equally weighed. The score indicates the percentage of patients that received all seven care processes.

Questions for the Committee:

- Are the quality construct and a rationale for the composite explicitly stated and logical?
- Is the method for aggregation and weighting of the components explicitly stated and logical?

Preliminary rating for composite quality construct and rationale:
☒ High ☐ Moderate ☐ Low ☐ Insufficient

- **Committee Pre-evaluation Comments:**

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

1a. Evidence to Support Measure Focus: For all measures (structure, process, outcome, patient-reported structure/process), empirical data are required. How does the evidence relate to the specific structure, process, or outcome being measured? Does it apply directly or is it tangential? How does the structure, process, or outcome relate to desired outcomes? For maintenance measures –are you aware of any new studies/information that changes the evidence base for this measure that has not been cited in the submission? For measures derived from a patient report: Measures derived from a patient report must demonstrate that the target population values the measured outcome, process, or structure.

- wonderful body of evidence - no concerns
- This composite process measure is based on NQF individual measures, to evaluate comprehensive assessment. The measures include NQF #1634, NQF #1637, NQF #1639, NQF #1638, NQF #1617, NQF #1641, and NQF #1647.
- pass
- comfortable with the information provided -no new studies, target audience values the measured outcome
- I do not see a need for repeat discussion and vote on Evidence.
- Important to have a Comprehensive Assessment approach when admitted to hospice care
- The developer attests that the evidence has not changed since 201.
- The rationale is rationale for the relationship between a comprehensive assessment of physical and psychosocial well-being and positive treatment outcomes. 2013 Institute for Clinical Systems Improvement (ICSI) Palliative Care for Adults guidelines support the components in the composite. members acknowledged that similar evidence for some components of the measure (pain screening, pain assessment, dyspnea screening, and addressing spiritual and religious concerns) does not exist and likely would not be forthcoming. The Committee agreed that empirical evidence is not needed to hold providers accountable for those components of the measure, and agreed to invoke the exception to the evidence subcriterion. Underlying evidence for this measure has not changed since the last NQF endorsement review.
- Updated guidelines referenced for 2020 (ICSI)
- This measure is still not strong in demonstrating outcomes but I can support continuation
- I am concerned with the lack of evidence for the other components of the measure (pain screening, pain assessment, dyspnea screening, and addressing spiritual and religious concerns) does not exist and likely would not be forthcoming. It is important hold providers accountable for performance in the absence of empirical evidence of benefits to patients.
- NA, existing measure, no new evidence that I'm aware of so should streamline evaluation process

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

- though the performance gaps may seem small, they are real, and really important. and disparities data is compelling, considering there should be zero disparities expected
- Previously, the Committee recognized the evidence base linking dyspnea treatment, bowel regimens, and communication regarding treatment preferences to improved patient outcomes. However, members acknowledged that similar evidence for the other components of the measure (pain screening, pain assessment, dyspnea screening, and addressing spiritual and religious concerns) does

not exist and likely would not be forthcoming. The Committee agreed that empirical evidence is not needed to hold providers accountable for those components of the measure, and agreed to invoke the exception to the evidence subcriterion. The developer analyzed data from 4,089 hospices, including 1,422,921 patient stays, with disparities noted.

- no issue, gap is known
- needs to be tracked, performance gap persists
- Q1. I wonder if the developer looked at the intersection of race/ethnicity and gender? Q2. Based on research, disparities do exist with regard to use of hospice services and with regard to hospital performance.
- This measure demonstrates disparities in care
- The evidence demonstrates a performance gap. Data on disparities was provided.
- Performance gaps reported based on scores from 77.8%. 89.6%. 8.4% of hospices had perfect scores and 11% scored lower than 75% (2019). Subgroup data for non-white and gender noted
- Assessed for disparities; noted for race
- Criteria met to continue.
- Disparities do appear to be present, in that the median hospital performance among hospitals with more non-white patients was 88.1% while among hospitals with less non-white patients it was higher at 90.6%. . Over time, the hospice-level mean score increased.
- This measure is important as a way to gauge the quality of hospice care using important parameters.

1c. Composite Performance Measure - Quality Construct (if applicable): Are the following stated and logical: overall quality construct, component performance measures, and their relationships; rationale and distinctive and additive value; and aggregation and weighting rules?

- no concerns
- The developer states that this measure calculates the percentage of patients who received all seven HQRP care processes at admission. All seven components are equally weighed. The score indicates the percentage of patients that received all seven care processes.
- no issues
- yes
- I do think the quality construct and rational are explicitly stated and logical.
- This measure is comprised of seven NQF endorsed quality measures and provides rationale for the aggregation. Provides good argument for this being a composite quality construct
- The quality construct is logical.
- All-or-nothing composite measure. Developer says the aggregation of 7 measures incentivizes hospitals, sets a high standard and provides consumers with single measure of quality.
- Yes
- Meets criteria
- This measure reflects the overall quality of comprehensive assessment at hospice admission for each patient stay. All seven components are equally weighed. The score indicates the percentage of patients that received all seven care processes.
- NA, existing measure

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

2c. For composite measures: empirical analysis support composite approach

- Reliability

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

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

- Validity

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

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

Composite measures only:

2d. Empirical analysis to support composite construction. Empirical analysis should demonstrate that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct.

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

Evaluators: NQF Scientific Methods Panel

Methods Panel Review (Combined)

Methods Panel Evaluation Summary:

This measure was reviewed by the Scientific Methods Panel and discussed on the call. A summary of the measure and the Panel discussion is provided below.

- **Ratings for reliability:** H-5; M-3; L-0; 0-I → Measure passes with HIGH rating
  - Reliability testing conducted at the performance score level
  - The ICC coefficient for this measure was 0.86
  - The signal-to-noise ratio for this measure was 3.55
  - Stability analysis showed that approximately 70% of providers had a change in QM score of less than one standard deviation
  - The SMP members generally agreed that the reliability tests are appropriate, and the results show moderate-to-high reliability, although two members raised questions about the signal-to-noise ratio method and how to interpret the resulting score of 3.55
- **Ratings for validity:** H-2; M-5; L-1; I-0 → Measure passes with MODERATE rating
  - Validity testing conducted at the individual performance measures level by examining correlations between each component measure and the overall composite measure

- The p-values for all the Spearman correlation coefficients are significant (p-value < 0.01). There are significant positive correlations (ranging from 0.26 to 0.73) between the composite measure and each of the QMs
- Exclusion analysis showed the mean QM scores among hospices were within 0.02 percentage points with and without the age exclusion
- The missing rate for the majority of items were between 0.001 percent and 0.01 percent
- The SMP reviewers generally agreed that this measure passed validity criterion; although, a couple pointed out it was coarse in its ability to identify "meaningful" differences in performance as the distribution is fairly compressed near the top

Ratings for Composite: H-2; M-6; L-0; I-0 → Measure passes with MODERATE rating

- The SMP reviewers generally agreed that the construction of the composite measure from the seven individual NQF-endorsed measures was straightforward

Questions for the Committee regarding reliability:

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

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?
- The Scientific Methods Panel is satisfied with the validity analyses for the measure. Does the Committee think there is a need to discuss and/or vote on validity?

Questions for the Committee regarding composite construction:

- Do you have any concerns regarding the composite construction approach (e.g., do the component measures fit the quality construct and add value to the overall composite? Are the aggregation and weighting rules consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible?)?
- The Scientific Methods Panel is satisfied with the composite construction. Does the Committee think there is a need to discuss and/or vote on the composite construction approach?

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

- **Committee Pre-evaluation Comments:**

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

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

- no concerns
- Reliability tested at the individual measure level. The SMP members generally agreed that the reliability tests are appropriate, and the results show moderate-to-high reliability.

- no concerns
- no concerns
- I have no concerns.
- Reliability demonstrated
- No concerns
- ICC coefficient 0.86-good reliability
- No concerns
- It is limited and yet can support
- It appears that this measure can be consistently implemented.
- NA, existing measure

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

- no
- None noted.
- no concerns
- no concerns
- I have no concerns and see no reason to discuss and/or vote on reliability.
- Reliability demonstrated
- No
- No
- No
- limited concern
- no concerns discussion is not necessary
- No

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

- no
- None noted.
- no concerns
- no concerns
- I have no concerns and see no reason to discuss and/or vote on validity.
- Validity demonstrated
- No
- Measures by examining correlation between each component and the overall composite measure. p-values for spearman correlation coefficients are significant. Distribution fairly compressed near the top. No concerns
- No
- no
- no
- No

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

- no concerns
- This is a composite process measure.
- no concerns
- yes
- I do not see any issues with regard to exclusion or risk adjustment.
- No
- Risk adjustment was not used.
- Does not include patients < 18 years old.
- No risk adjustment used
- adequate
- Exclusions seem to be consistent, the analyses indicate acceptable results.
- NA, existing measure

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

- no concerns
- It does not. No concerns noted.
- no concerns
- no concerns
- I do not see any threats to validity.
- No
- No concerns
- coarse in its ability to identify “meaningful” differences in performance as the distribution is fairly compressed near the top. Low missing rate - 0.001% - 0.01%
- No
- outcome data would be helpful but difficult to obtain
- There does not appear to be any missing data, statistically significant differences, or multiple data sources.
- NA, existing measure

2c. Composite Performance Measure - Composite Analysis (if applicable): Do analyses demonstrate the component measures fit the quality construct and add value? Do analyses demonstrate the aggregation and weighting rules fit the quality construct and rationale?

- no concerns
- Construct of composite measure appropriate.
- no concerns, appropriate
- I have not concerns and see no need to discuss and or vote.
- Yes
- SMP member 9 suggests that we should discuss whether the validity of the underlying component measure extends logically to the composite.
- No concerns.
- Yes
- Fits requirements
- The construction of the composite measure from the seven individual NQF-endorsed measures was straightforward.
- NA, existing measure

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.

- This measure's data elements are generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)
- For this measure, all data elements are in defined fields in electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)

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?
- If an eCQM, does the eCQM Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?

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

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

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

- yes, very much should be routinely collected and thus feasible
- Elements are all routinely generated. For this measure, all data elements are in defined fields in electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS.

- no concern
- no concerns
- I see no issues with regard to feasibility. The data are collected at the time of the assessment and available to the team via an electronic format.
- The elements are routinely collected during care delivery and available in electronic form.
- All data is collected electronically.
- Requires that records are abstracted by someone besides the person who collects them. All data elements are in the defined field in the electronic data.
- No concerns
- no concerns
- For this measure, all data elements are in defined fields in electronic clinical data.
- None

Criterion 4: Usability and Use

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

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

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

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

Current uses of the measure

- This measure is used in CMS' Hospice Quality Reporting Program (HQRP).

Publicly reported? ☒ **Yes** ☐ **No**

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

OR

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

Accountability program details

- Public reporting: This measure is currently publicly reported by the Centers for Medicare & Medicaid Services (CMS) on the Hospice Compare website.
- Accountability Program: This measure is included in the Hospice Quality Reporting Program (HQRP).

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 receives feedback on this measure through the Hospice Quality HelpDesk, quarterly provider association calls, and ODFs and HQRP forum.
- The developer reports that CMS has received feedback from providers on the Hospice Comprehensive Assessment Measure following the addition of the measure to providers' CASPER QM reports. Some providers asked for clarification about the specifications, specifically about the "all or none" (rather than average) scoring methodology used and the inclusion of conditional measures in calculation.

***Questions for the Committee**:*

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

Preliminary rating for Use: ☒ Pass ☐ No Pass

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

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

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

Improvement results

- According to the developer, performance results on this measure presented in 1b under Performance Gap indicate that hospices have made significant improvements in completing a comprehensive assessment at hospice admission. The developer suggests that the results indicate this measure encourages hospices to conduct all critical care processes for each patient and also sets a higher standard of care for hospices.

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 does not note any unexpected findings from implementation of this measure.

Potential harms

- The developer does not note any potential harms from implementation of this measure.

***Questions for the Committee**:*

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

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

- **Committee Pre-evaluation Comments:**

Criteria 4: Usability and Use

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

- no concerns
- This measure is used in CMS' Hospice Quality Reporting Program (HQRP).
- no concern
- yes
- I see no issues with accountability or transparency, and it appears that the measures are being used both within agencies and by the public through the CMS Hospice Compare website.
- Yes publicly reported & used in an accountability program. Feedback also obtained.
- The measure is publicly reported and included in the HQRP. Feedback has been received from providers.
- used in CMS' Hospice Quality Reporting Program (HQRP). currently publicly reported by the Centers for Medicare & Medicaid Services (CMS) on the Hospice Compare website. Feedback - some providers asked for clarification about the specifications, specifically about the "all or none" (rather than average) scoring methodology used and the inclusion of conditional measures in calculation.
- Feedback provided to CMS by hospice organizations. Measure is publicly reported on hospice compare.
- These are addressed
- This measure is publicly reported and is included in the Hospice Quality Reporting Program.
- NA, existing measure

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

- no concerns
- No evidence of harms noted.
- no concern
- no concerns
- I think that the performance results can definitely be used to further improve the services provided by the hospice agency. I cannot see any actual intended consequences.
- Results will be utilized for quality improvement. No harms noted
- The benefits of improving care at the end of life outweighs any unintended negative consequence.
- Per developer - hospices have made significant improvements in completing a comprehensive assessment at hospice admission. The developer suggests that the results indicate this measure encourages hospices to conduct all critical care processes for each patient and also sets a higher standard of care for hospices.
- None
- none noted
- The developer suggests that the results indicate this measure encourages hospices to conduct all critical care processes for each patient and also sets a higher standard of care for hospices. No apparent harms.
- I am unaware of any harms

Criterion 5: Related and Competing Measures

Related or competing measures

- Hospice and Palliative Care – Pain Screening (NQF #1634),
- Hospice and Palliative Care – Pain Assessment (NQF #1637),
- Hospice and Palliative Care – Dyspnea Screening (NQF #1639),
- Hospice and Palliative Care – Dyspnea Treatment (NQF #1638),
- Patients Treated with an Opioid Who Are Given a Bowel Regimen (NQF #1617),
- Hospice and Palliative Care – Treatment Preferences (NQF #1641), and
- 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 (NQF #1647).

Harmonization

- Measures are harmonized to the extent possible. NQF will ask the Committee to discuss the utility of continued endorsement of the individual measures if the composite is endorsed.

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

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

- no competing measures
- •Hospice and Palliative Care – Pain Screening (NQF #1634), • Hospice and Palliative Care – Pain Assessment (NQF #1637), •Hospice and Palliative Care – Dyspnea Screening (NQF #1639), •Hospice and Palliative Care – Dyspnea Treatment (NQF #1638), •Patients Treated with an Opioid Who Are Given a Bowel Regimen (NQF #1617), • Hospice and Palliative Care – Treatment Preferences (NQF #1641), and • 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 (NQF #1647).
- harmonized, would favor composite endorsement instead of individual measures
- None that I am aware of.
- Yes there are individual measures that are included in this composite measure and would suggest re-examining if each of the individual measures still needed if included in this measure
- Do we need to continue to endorse the individual measures if the composite is endorsed?
- Hospice and Palliative Care – Pain Screening (NQF #1634), • Hospice and Palliative Care – Pain Assessment (NQF #1637), • Hospice and Palliative Care – Dyspnea Screening (NQF #1639), • Hospice and Palliative Care – Dyspnea Treatment (NQF #1638), • Patients Treated with an Opioid Who Are Given a Bowel Regimen (NQF #1617), • Hospice and Palliative Care – Treatment Preferences (NQF #1641), and • 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 (NQF #1647).
- This is a composite measure of 7.
- none of which I am aware
- There are 7 related measures.
- I am not aware of related or competing measures

Public and Member Comments

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

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

- Combined Methods Panel Scientific Acceptability Evaluation

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 3235

Measure Title: Hospice and Palliative Care Composite Process Measure—Comprehensive Assessment at Admission

Type of measure:

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

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

Data Source:

☐ **Claims** ☐ **Electronic Health Data** ☐ **Electronic Health Records** ☐ **Management Data** ☒ **Assessment Data** ☐ **Paper Medical Records** ☒ **Instrument-Based Data** ☐ **Registry Data** ☐ **Enrollment Data** ☒ **Other**

Level of Analysis:

☐ **Clinician: Group/Practice** ☐ **Clinician: Individual** ☒ **Facility** ☐ **Health Plan** ☐ **Population: Community, County or City** ☐ **Population: Regional and State** ☐ **Integrated Delivery System** ☐ **Other**

Measure is:

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

RELIABILITY: SPECIFICATIONS

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

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

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

2. **Briefly summarize any concerns about the measure specifications.**

   Panel Member 1: Specifications were documented well. Extensive.

   Panel Member 4: No concerns

   Panel Member 5: No concerns

   Panel Member 6: I have no concerns. The measure is a simple composite of 7 binary indicators.

Panel Member 7: I am somewhat concerned that the number of specifications associated with each of the measures included in the composite represents a substantial overall burden on data abstractors and increases the probability of errors.

Panel Member 8: The developer present the measure clearly and i have no concern.

RELIABILITY: TESTING

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

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

   ☐ **Yes** ☒ **No**

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

   Panel Member 3: Developers performed three types of reliability analyses. They refer to them as (1) split-half reliability, (2) signal-to-noise analysis, and (2) stability analysis.
   Method #1 used providers with at least 20 patient stays and estimated the intraclass correlation coefficient (ICC). The term ICC has multiple definitions so it would be helpful to clarify the interpretation of the quantity that their ICC was estimating. For example, is it simply the correlation coefficient between 2 estimates each calculated from half of the stays from each provider. If so, this would be expected to under-estimate the reliability that would be expected for composites calculated from all of a provider's stays (as opposed to only using half of the data). The reported ICC sounds like an acceptable value (0.86) so we may be able to conclude that reliability for the actual measure (using all of the data, not half) is quite high.
   For Method #2, the developers state that they "divided the average of composite scores at the hospice-level by the standard deviation of composite scores across all patient stays to determine the extent to which total variance in the measure is attributable to differences among providers." It's possible I am not understanding the above descriptions, but it's unclear to me how this measure sheds light on reliability or can be interpreted as a measure of signal-to-noise reliability. I would like to ask for clarification or suggest disregarding this analysis.
   Method #3 involved comparing performance scores from the same provider across 4 consecutive harvests to determine the extent to which scores remain stable. The developers report that scores changed by less than 1 standard deviation for approximately 70% of providers. This result is not simple to interpret because it describes reliability for a measure calculated using only a quarter of the data from each provider. (I am assuming that reporting of the measure is based on a 12-month measurement window.) If 70% of providers change by less than 1 SD, this result is mathematically consistent with a reliability value of ~0.55 for scores calculated on a quarter's worth of data. Using the Spearman-Brown prophecy formula, it would suggest reliability of ~0.83 for a full year's data.

   Panel Member 4: The methods use were: 1) split half reliability; 2) signal to nose analysis; and 3) stability analysis. These seem appropriate.

   Panel Member 5: Testing methods of reliability for this composite measure appears appropriate.

   Split-half reliability. Split-half reliability assesses the internal consistency

   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.

   Stability analysis. Stability analysis describes the extent to which providers' performance assessed by a QM changes across time' [p6-7]

Panel Member 6: Weinhandl: Split-half reliability, signal-to-noise analysis, and stability analysis were all conducted.

Panel Member 7: Kaplan: Split-half reliability within facility, signal-to-noise and "stability analysis" were within NQF current guidance as acceptable assessments of reliability.

Panel Member 8: Lin: The developer included three types of score level reliability analyses: Split-half reliability, signal-to-noise analysis, and stability analysis. As described, only split-half reliability analysis is appropriate while stability analysis is useful and informative. Signal-to-noise analysis as described is not what one would expect, the resulting score of 3.55 confirms that.

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

Panel Member 3: See above

Panel Member 4: The results were: 1) split half reliability =0.86; 2) signal to nose analysis = 3.55; and 3) stability analysis =70% of providers had a change of less than 1 SD. These results indicate moderate to high reliability.

Panel Member 5: Testing results of reliability for this composite measure are high in general across the tests performed.

'Split-half reliability. The ICC coefficient for this measure is 0.86.

Signal-to-noise analysis. The signal-to-noise ratio for this measure is 3.55.

Stability analysis. ***Figure 1*** below illustrates the observed change in facility scores between the four consecutive quarters analyzed (Q4 2018 and Q1 2019, Q1 and Q2 2019, and Q2 and Q3 2019), where the changes in facility scores are reported in standard deviations.

Approximately 70% of providers had a change in QM score of less than one standard deviation. …. Less than 1 percent of providers had a change in QM score between one and two standard deviations and very few providers had a change in QM within two or more standard deviations.' [p7]

Panel Member 6: The estimated intraclass correlation coefficient in split-half reliability analysis was 0.86—very high. The signal-to-noise ratio was 3.56—likewise, very high. Stability analysis showed that nearly three-fourths of all hospices had measure scores differences of less than one standard deviation in a series of four consecutive quarters.

Panel Member 7: Results provided indicate adequate reliability for split-half and signal-to-noise, along with stability over the three time periods assessed.

Panel Member 8: Split half reliability of 0.86 indicates excellent reliability. It is not clear how to interpret the reliability of 3.55 based on "signal-to-noise" reliability.

Panel Member 9: Results indicate an acceptable level of reliability at the measure score level, assuming an adequate sample size at each hospice program being evaluated**.**

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

Submission document: Testing attachment, section 2a2.2

☒ **Yes**

☐ **No**

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

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

Submission document: Testing attachment, section 2a2.2

☒ **Yes**

☐ **No**

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

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

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

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

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

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

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

Panel Member 1: Testing results.
Panel Member 3: I had several questions about the methods but I think it's a safe bet that reliability is likely to be acceptable.

Panel Member 4: No concerns.

Panel Member 5: As noted in Q7: Testing results of reliability for this composite measure are high in general across the tests performed.

Panel Member 6: Weinhandl: The measure has very high reliability.

Panel Member 7: Kaplan: The data provided indicate that reliability data provided are within the current NQF guidance for moderate reliability (ICC for split-half=.86, signal-to-noise ratio=3.55 and >70% of providers fell within one standard deviation of their scores for the prior observation period.

Panel Member 8: Lin: All seven components of the composite score are based on NQF endorsed measures. The score level split-half reliability is pretty high.

Panel Member 9: Nerenz: Three separate tests of reliability at the measure score level were conducted, and all showed reasonably high levels of reliability. The combination of ICC approach and a test-retest stability approach was unusual and commendable.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

Panel Member 3: : None

Panel Member 4: No concerns.

Panel Member 5: Concerns follow:

[1] Regarding exclusion of under age 18:

- Table 2 has errors in the reporting. In the columns titled "Percentage of pediatric patients in numerator" & "Percentage of non-pediatric patients in numerator" contain figures of 96%+. Of course, that is impossible. [p10]
- Per Table 3: While the mean rate of cases under age 18 is only 1.2%, we see that in general there variably of such excluded cases is fairly substantial: 1st percentile: 0.05%, 99th percentile: 9.09%
- There's no rationale provided for excluding cases under ae 18.

[2] The MIF states the following exclusions that are not addressed in the testing form:

- Patients with Type 2 (discharged stays missing the admission record)
- Type 3 patient stays (active stays) [p9]

Panel Member 6: I have no concerns. Pediatric patients are excluded.

Panel Member 7: None
Panel Member 8: No concern

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

Panel Member 3: Results in 2b5.2 indicate that the measure was able to classify a substantial proportion of providers (65%) as having performance above or below the overall mean. The interquartile ranges across participants was 85% to 97.05%. This strikes me as a meaningful difference. It's unclear to be how much of this variation is due to true signal variation but I assume it's substantial.

Panel Member 4: No concerns.

Panel Member 5: No concerns. Measure expresses a high degree of variation across providers when testing at the 95% CI. '…65.32% 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 (37.54% vs 27.78%, respectively).' [p13]

Panel Member 6: Weinhandl: I have no concerns. A high percentage of hospices exhibit significantly higher or lower performance in the composite measure, relative to the national mean.

Panel Member 7: Kaplan: The data appear to be skewed in the direction of favorable results (median QM score=93.64%) although the interquartile range was ~12% and ~8% of hospices had perfect scores. The meaning of small differences between facilities is not clear.

Panel Member 8: Lin: No concern

Panel Member 9: Nerenz: The analyses on this point showed that the measure is capable of dividing hospices into three groups – those significantly better than average, those significantly worse than average, and all others who are no different from average. The distribution is fairly compressed near the top, and it isn't quite clear how clinically significant it is to be significantly higher or lower than average when almost all hospices do pretty well. The measure is as good as many other endorsed measures on this criterion, but the ability to just distinguish three groups vs. some more fine-grained distinction means that the measure is very coarse in its ability to identify "meaningful" differences in performance.

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**

Submission document: Testing attachment, section 2b5.

Panel Member 4: No concerns.

Panel Member 5: NA – 1 data source was used.

Panel Member 6: This is not applicable.

15. **Please describe any concerns you have regarding missing data.**

Submission document: Testing attachment, section 2b7.

Panel Member 4: No concerns.

Panel Member 5: No concerns. The level of missingness across the measures was very low. The code with the highest frequency of missing occurred at a rate of 0.016%. [p15]

Panel Member 6: I have no concerns. Missing component measure values are exceedingly rare.

Panel Member 7: None. Proportion of missing data appears small.

Panel Member 8: No concern.

16. **Risk Adjustment**

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

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

☐ Yes ☒ No ☒ Not applicable

16c. **Social risk adjustment:**

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

16c.2 Conceptual rationale for social risk factors included? ☒ Yes ☒

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

16d. **Risk adjustment summary:**

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

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

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

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

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

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

Panel Member 1: Farquhar: Information located in the Evidence and Performance Gap section.

Panel Member 4: Warholak: The developer indicated in 1b4 in their racial and ethnic disparity analysis at the patient stay level and the hospice level that the "Differences in the rate of completion of all seven care processes by racial identification were found to be statistically significant (p-value < 0.01)" and "The results showed that the QM score was significantly different between the two groups of hospices (88.1% for hospices with a proportion of nonwhite patients greater than the national median compared with 90.6% for hospices with a proportion of nonwhite patients less than or equal to the national median, p-value < 0.01)." This would indicate to me the need to risk adjust but no risk adjustment model is presented for the measure.

Panel Member 5: NA – No risk adjustment. Disappointing no rationale presented as to the lack of risk adjustment

Panel Member 6: Weinhandl: he correlations of the composite measure and each of its components were estimated.

Panel Member 7: Kaplan: Section 2b4 was not completed in the application I received**.**

Panel Member 9: This is a composite process measure and it is reasonable to not adjust for either clinical or social factors, as any effects in either arena are handled at the exclusion stage, and the process steps should be performed for all patients regardless of presence/absence of potential risk variables. The concept of "risk" doesn't really apply here in the same way it would for an outcome measure.

For cost/resource use measures ONLY:

17. **Are the specifications in alignment with the stated measure intent?**

    ☐ **Yes** ☐ **Somewhat** ☐ **No (If "Somewhat" or "No", please explain)**

18. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**

VALIDITY: TESTING

19. **Validity testing level:** ☒ **Measure score** ☒ **Data element** ☒ **Both**

20. **Method of establishing validity of the measure score:**

    ☐ **Face validity**

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

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

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

Panel Member 1: Appropriate
Panel Member 3: Obrien: Developers examine correlations between each component measure and the overall composite and they assess the impact of removing each individual component from the composite

one at a time. For the latter analysis, they report the percent of outliers that were identified by the original composite that were also identified when a given component was removed. I would have preferred to report correlation or rank correlation coefficients between the original composite and each modified composite. For looking at agreement with respect to statistical classifications, I would have preferred to see a set of 3x3 contingency tables comparing classifications based on the original composite (above, no different, below average) to classifications based on the modified composite.
The developers also report on the frequency of missing data.

Panel Member 4: Warholak: The developer used correlations to show convergent validity of the individual component measures to the composite measure score. While all of the correlations were statistically significant at the $p<0.001$ level, the correlations range from low (0.26 for treatment preferences and 0.33 for dypsnea) too high for pain assessment (0.73). While this seems rational, it seems that additional validity evidence would be nice to have.

Panel Member 5: Regarding measure score testing:

The correlation test occurred using measures within the composite. Would have been preferable to test the correlation of the composite with other gold standard measures in this space.

Regarding data element testing:

The test of missing data is not helpful. Further the measure steward refers to missing data test results in 2b7, which simply notes the degree of missingness of data elements.

Correlations. Providers should perform similarly on QMs that reflect the quality of similar care processes. Thus, a common strategy used to evaluate validity is to examine the correlation among measures that capture related clinical care processes: convergent validity.

Missing data. We also conducted analysis of missing data to support the validity of this measure. [p9]

Panel Member 6: Weinhandl: he correlations of the composite measure and each of its components were estimated.
Panel Member 7: Kaplan: The developer correlated each of the elements of the composite with the aggregate composite measure, the equivalent of 'item-to-total' correlation, which would be expected if all measures reflect the underlying construct. However, in a reflective measurement model, this assessment (as used to improve or estimate the contribution of each measure to reliability coefficients in Cronbach's alpha) is an estimate of internal consistency, not validity. The use of quality measures not part of the composite would have been more helpful in the assessment of construct validity, or the prediction of some other aspect of hospice performance in another time period could have been used to provide evidence of predictive validity.

Panel Member 8: The developer assessed the correlation between the composite score and all seven HQRP QMs, given that the composite score was based on those seven components, there are inherent relationships between components and composite. It would be preferable if other external quality markers were used.

Panel Member 9: The method for establishing validity is not acceptable. The developers show correlations between the composite score and each of the individual elements of the composite. The correlations that are seen are essentially guaranteed by definition or by construction – an entity with a higher composite score must at some level have higher scores on the composite measures. This is not an independent test of validity that involves comparison of the composite against some separate measure of quality of care.

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

Submission document: Testing attachment, section 2b2.3

Panel Member 1: Reasonable.
Panel Member : Individual components are correlated with the overall composite as expected. No single item is correlated with the composite to the point that it appears to dominate the composite or make the other items irrelevant. Items with average scores above 99% do not appear to contribute substantial

statistical information but there is a conceptual basis for retaining all 7 items in order to incentive continued high perform on all of them. Missing data are very rare and are therefore unlikely to have a large impact on score results.

Panel Member 4: While this seems rational, it seems that additional validity evidence would be nice to have. I'm not sure that checking a box really means that the assessment was done… but it is a process measure… and I guess that is the nature of process measures.

Panel Member 5: Regarding measure score testing:

The correlation test results in Table 1 [p9] are modest and acceptable.

Regarding data element testing:

The measure steward refers to missing data test results in 2b7, which simply notes the degree of missingness of data elements. The test result is moot as the type of test is not helpful in ascertaining validity.

Panel Member 6: Weinhandl: All of the estimated correlations were positive and statistically significant ($p < 0.01$).

Panel Member 7: Kaplan: See above.

Panel Member 8: The moderate correlation between components and the composite is expected given their inherent relationships.

Panel Member 9: Nerenz: The results seem to show some kind of validity, but the premise underlying the analyses is not acceptable**.**

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

    Submission document: Testing attachment, section 2b1.

    ☒ **Yes**

    ☒ **No**

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

24. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?**

    NOTE that data element validation from the literature is acceptable.

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

    ☒ **Yes**

    ☒ **No**

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

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

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

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

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

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

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

Panel Member 1: NQF endorsement of individual measures in the composite and the correlation values provided.

Panel Member 3:: Individual components are correlated with the overall composite as expected. No single item is correlated with the composite to the point that it appears to dominate the composite or make the other items irrelevant. Items with average scores above 99% do not appear to contribute substantial statistical information but there is a conceptual basis for retaining all 7 items in order to incentive continued high perform on all of them. Missing data are very rare and are therefore unlikely to have a large impact on score results.

Panel Member 4: No other concerns.

Panel Member 5:

Q12: Concerns follow:

[1] Regarding exclusion of under age 18:

- Table 2 has errors in the reporting. In the columns titled "Percentage of pediatric patients in numerator" & "Percentage of non-pediatric patients in numerator" contain figures of 96%+. Of course, that is impossible. [p10]
- Per Table 3: While the mean rate of cases under age 18 is only 1.2%, we see that in general there variably of such excluded cases is fairly substantial: 1st percentile: 0.05%, 99th percentile: 9.09%
- There's no rationale provided for excluding cases under ae 18.

[2] The MIF states the following exclusions that are not addressed in the testing form:

- Patients with Type 2 (discharged stays missing the admission record)
- Type 3 patient stays (active stays) [p9]

Q16e: No risk adjustment. Disappointing no rationale presented as to the lack of risk adjustment

Q22: Regarding measure score testing:

The correlation test results in Table 1 [p9] are modest and acceptable.

Regarding data element testing:

The measure steward refers to missing data test results in 2b7, which simply notes the degree of missingness of data elements. The test result is moot as the type of test is not helpful in ascertaining validity.

Panel Member 6: All of the components of the composite measure are NQF-endorsed measures, and the composite measure itself is positively correlated with each of its components.

Panel Member 7 It is difficult to address the issue of validity given the concerns about the relationship of each of the measures included in the composite to the overall composite, as that is usually an assessment of internal consistency reliability, not validity for reflective measurement models. If this were a formative measurement model, the relationship of each of the measures in the composite as well as the overall composite to some external validity variable would have been more appropriate.

Panel Member 8: It would be much preferred had the developer included additional validity testing. All seven components were based on NQF endorsed measures.

Panel Member 9: As noted above, simply showing the correlations between a score on a composite measure and scores on the elements of the composite is not sufficient to show validity of the composite measure. Positive correlations are essentially guaranteed by definition and by the construction of the composite.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

27. **What is the level of certainty or confidence that the empirical analysis demonstrates that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct?**

    ☒ **High**

    ☒ **Moderate**

    ☐ **Low**

☐ **Insufficient**

28. **Briefly explain rationale for rating of EMPIRICAL ANALYSES TO SUPPORT COMPOSITE CONSTRUCTION**

Panel Member 1: Demonstrated by positive, significant correlations of the measure to each individual quality measure. Weak correlations between the quality measures which indicates that a minimum of overlapping quality information among the measures in the composite.
Panel Member 3: Each component measure is endorsed by NQF and measures processes that are considered to be important components of high quality hospice care. The proportion of patients who receive all 7 processes is a straightforward concept and it is measured directly by the calculated composite.

Panel Member 4:: No major concerns.

Panel Member 5: Unimpressed with the face validity testing, such as small sample of caregivers, unstated number of others and no explanation as to how these candidates were sampled to ensure representativeness and avoid conflicts of interest.

The Spearman rank correlation analysis was reasonable regarding measures that comprise the composite.

'A mixed methods approach was used. Qualitative information gathering activities—including an environmental scan, focus groups with 6 hospice patient caregivers (representing consumers), and a Technical expert panels involving clinical stakeholders —were conducted to support this measure concept and inform the measure's specifications. Quantitative analyses conducted include a nonparametric Spearman rank correlation analysis between the composite measure and all seven HQRP QMs.' [p16]

Panel Member 6: Weinhandl: Moderate, positive correlations of the composite measure with each of its components support a value-added concept of the composite measure.
Panel Member 7: Kaplan: The developer provided data suggesting that each of the individual measures included in the composite provided unique information regarding the identification of outliers. However, the contribution of each of the measures to the overall score did not appear to be substantial, especially #1641, treatment preferences.
Panel Member 9: Nerenz: No particular concerns – the construction of the composite from the seven underlying measures is straightforward**.**

ADDITIONAL RECOMMENDATIONS

29. **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.**

Panel Member 3: No concerns
Panel Member 5: No concerns

Panel Member 9: The analyses of validity are the key problem here. If each of the component measures has been shown to be valid in earlier rounds of NQF review and endorsement, then the composite measure may be judged valid just by logical extension. However, the empirical evidence presented here doesn't add any new information on the validity of the composite, and in fact does not really speak to the validity of the composite. The Standing Committee could judge the measure to be valid if it believes that the validity of the underlying component measures extends logically and naturally to the composite. That is an issue best handled by the Standing Committee; the formal statistical evidence presented here does not speak to validity in a way that is acceptable from the SMP perspective.

Developer Submission

Additional evaluations and submission materials attachments…

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

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

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

NQF_Measure_Evidence_Form_3235_final.DOCX

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

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

No

- 1a. Evidence (subcriterion 1a)

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

Outcome

☐ Outcome:

☐ Patient-reported outcome (PRO):

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

☐ Intermediate clinical outcome (*e.g., lab value*):

☒ Process: Screening/management of pain, dyspnea, and bowels around the time of hospice admission, as well as discussions of patient preferences regarding life sustaining treatments and spiritual/existential concerns.

☐ Appropriate use measure:

☐ Structure:

☒ Composite: The overall quality of comprehensive assessment at hospice admission for each patient stay based on care processes for the screening/management of pain, dyspnea, and bowels around the time of hospice admission, as well as discussions of patient preferences regarding life sustaining treatments and spiritual/existential concerns.

1a.12 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

This composite QM for comprehensive assessment at admission addresses high priority aspects of quality hospice care as identified by both leading hospice stakeholders and patients. The Medicare Hospice Conditions of Participation (CoPs) require that hospice comprehensive assessments identify patients' physical, psychosocial, emotional, and spiritual needs, and address them to promote the hospice patient's comfort

throughout the end of life process[1]. Furthermore, the person-centered approach to care, as well as family and caregiver perspectives on end-of-life care, align with the domains identified by the CoPs, as patients and their families/caregiver also place value on physical symptom management and spiritual/psychosocial care as important factors at the end of life[2,3]. A composite measure serves to ensure all hospice patients receive a comprehensive assessment for both physical and psychosocial needs at admission. Below is further evidence of the importance of each measure that constitutes the composite.

NQF #1634: Pain Screening & NQF #1637 Pain Assessment: Pain is under-recognized and undertreated, resulting in excess suffering for patients with serious illness. Pain screening improves the provider's awareness of the presence of pain and it is the essential first step for quality pain management and treatment. Without initial screening to identify patients in pain, and clinical assessment to determine the severity, etiology, and effect on function, effective treatment cannot be administered. Pain screening and assessment are necessary processes in order to improve the patient-centered outcome of pain management.

NQF #1639 Dyspnea Screening & NQF #1638 Dyspnea Treatment: Dyspnea is a prevalent yet undertreated condition for many seriously ill patients. Dyspnea screening and assessment are necessary to detect the presence of dyspnea and to understand its severity. Screening will form the basis for treatment decision-making and will facilitate opportunities for effective treatment and symptom alleviation. Effective treatment for dyspnea should be made available to alleviate symptom distress for patients with dyspnea. Timely screening will facilitate opportunities for effective treatment and symptom alleviation.

NQF #1617 Bowel Regimen: This quality measure serves to ensure that patients on opioids are offered or prescribed a bowel regimen. Most patients prescribed opioids to manage pain or other symptoms develop some degree of constipation after opioid initiation or dose increases. Older adults who are immobile and dehydrated are at increased risk for constipation, though adults of any age, report constipation as a side effect from opioid use. Reducing opioid-induced constipation can reduce patient discomfort and improve quality of life.

NQF #1647 Beliefs/Values Addressed: This quality measure helps agencies improve processes for addressing spiritual/religious concerns of patients and families receiving hospice care and promotes a truly interdisciplinary approach by ensuring that the spiritual needs of the patient are incorporated into the provision of care in conjunction with his/her physical and psychological needs. A discussion of beliefs and values is the core of a rigorous assessment of spiritual care needs and is essential to ensuring these needs are met.

NQF #1641 Treatment Preferences: This quality measure facilitates opportunities for patients to express life-sustaining treatment preferences, which enhances patient autonomy over treatments, facilitates patient-centered decision-making, improves patient and family satisfaction outcomes, improves transitions to hospice and palliative care, and communicates patient preferences via documentation to other treating providers. Patients given this opportunity are more likely to receive care consistent with their values, improving patient and family outcomes, including greater satisfaction with care.

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

N/A

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

---

[1] Medicare and Medicaid Programs: Hospice Conditions of Participation, Part 418 subpart 54. Centers for Medicare and Medicaid Services, June 5, 2008.

[2] Singer PA, Martin DK, Kelner M. Quality End-of-Life Care: Patients' Perspectives. JAMA. 1999;281(2):163-168. doi:10.1001/jama.281.2.163.

[3] Steinhauser KE, Christakis NA, Clipp EC, McNeilly M, McIntyre L, Tulsky JA. Factors Considered Important at the End of Life by Patients, Family, Physicians, and Other Care Providers. JAMA. 2000;284(19):2476-2482. doi:10.1001/jama.284.19.2476.

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

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

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

☑ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Systematic Review

Evidence

Source of Systematic Review:

- Title
- Author
- Date
- Citation, including page number
- URL

The Health Care Guideline on Palliative Care for Adults from the Institute for Clinical Systems Improvement (ICSI) is evidenced based and uses GRADE methodology to evaluate the literature.

Citation: McCusker M, Benson J, Dvorkin J, Hadzic S, Hansen A, Jolkvosky M, Rosielle D, Ruff R, Schmidt M. Institute for Clinical Systems Improvement. Palliative Care for Adults. Updated January 2020.

URL for Guideline: https://www.icsi.org/guideline/palliative-care/

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

Guideline 2. Assess Patient's Palliative Care Needs Based on the Following Domains of Palliative Care: Clinicians should use validated assessment tools, where available, to assess palliative care needs for each of these domains: physical, cultural, psychological, social, spiritual, and ethical/legal. (p. 18)

Guideline 3. Begin Advance Care Planning Process: 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. (p. 23)

Guideline 4. Physical Aspects of Care: The control of physical symptoms is an important part of the palliative care plan. (p. 27)

Guideline 5. Cultural Aspects of Care: The cultural assessment promotes patient/family-centered decision-making and offers the opportunity to identify care practices. Cultural decisions affecting palliative care also include attention to gender, age, generation, education level, diet/food, and ritual. Clinicians should ask the patient/family about these considerations. (p. 28)

Guideline 8. Spirituality is recognized as an integral part of the palliative care plan. Clinicians should screen for spiritual beliefs and practices and respond respectfully. (p. 33)

Grade assigned to the **evidence** associated with the recommendation with the definition of the grade

All evidence-based recommendations for palliative care (Guideline 2: Assess Domains of Palliative Care; Guideline 3: Advance Care Planning) were rated as having a **Low** quality of evidence. In the GRADE approach, each outcome is rated as one of four categories: Very Low, Low, Moderate, or High (Guyatt et al., 2011 (Intro to GRADE)). The quality of evidence rating depends on a number of factors: study design (randomized trial, observational study), risk of bias, inconsistency, indirectness, imprecision, publication bias, effect size, dose response, and consideration of plausible confounding variables (Guyatt et al., 2011). A low quality of evidence rating means that "The true effect might be markedly different from the estimated effect." (https://bestpractice.bmj.com/info/us/toolkit/learn-ebm/what-is-grade/)

Provide all other grades and definitions from the evidence grading system

Very low: The true effect is probably markedly different from the estimated effect.

Moderate: The authors believe that the true effect is probably close to the estimated effect.

High: The authors have a lot of confidence that the true effect is similar to the estimated effect.

(https://bestpractice.bmj.com/info/us/toolkit/learn-ebm/what-is-grade/)

Grade assigned to the **recommendation** with definition of the grade

All evidence-based recommendations for palliative care (Guideline 2: Assess Domains of Palliative Care; Guideline 3: Advance Care Planning) were rated as **Strong** recommendations. The direction and strength of the recommendation depends on the quality of the evidence as well as the balance between desirable and undesirable outcomes and the application of patients' values and preferences (Guyatt et al., 2011). In GRADE, recommendations can be strong or weak, in favor or against intervention. Strong recommendations suggest that all or almost all persons would choose that intervention.
(https://bestpractice.bmj.com/info/us/toolkit/learn-ebm/what-is-grade/)

Provide all other grades and definitions from the recommendation grading system

Weak recommendations imply that there is likely to be an important variation in the decision that informed persons are likely to make.

(https://bestpractice.bmj.com/info/us/toolkit/learn-ebm/what-is-grade/)

Body of evidence:

- Quantity – how many studies?
- Quality – what type of studies?

Studies included in the assessment of evidence were: systematic reviews, meta-analyses, randomized controlled trials, implementation studies, observational studies, technical briefs, case reports, case studies, and summary reports.

Guideline 2. Clinicians should use validated assessment tools, where available, to assess palliative care needs for each of these domains: physical, cultural, psychological, social, spiritual, and ethnical/legal.

- 18 articles were included in this evidence review.
- This included 1 technical brief, 1 randomized control trial, 1 report, 6 case reports, 4 observational studies, and 5 summary reports.

Guideline 3. 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.

- 7 articles were included in this evidence review.
- This included 1 report, 1 case report, 1 case study, and 4 summary reports.

Guideline 4. The control of physical symptoms is an important part of the palliative care plan.

- 9 references were included in this recommendation.

Guideline 5. The cultural assessment promotes patient/family-centered decision-making and offers the opportunity to identify care practices. Cultural decisions affecting palliative care also include attention to gender, age, generation, education level, diet/food, and ritual. Clinicians should ask the patient/family about these considerations.

- 6 references were included in this recommendation.

Guideline 8. Spirituality is recognized as an integral part of the palliative care plan. Clinicians should screen for spiritual beliefs and practices and respond respectfully.

- 14 references were included in this recommendation

Estimates of benefit and consistency across studies

Guideline 2. Clinicians should use validated assessment tools, where available, to assess palliative care needs for each of these domains: physical, cultural, psychological, social, spiritual, and ethnical/legal.

- Thorough assessment is less likely to miss symptoms in need of management. The consistent use of tools creates reliability in assessment over time and potentially with different providers.

Guideline 3: 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.

- Regular review of advance care planning ensures patient wishes for treatment are accurately documented and family understands the benefits and burdens of available treatment option.

Guideline 4: The control of physical symptoms is an important part of the palliative care plan.

- Patients receiving palliative care interventions (including for dyspnea and pain) experience improved quality of life.

Guideline 5. The cultural assessment promotes patient/family-centered decision-making and offers the opportunity to identify care practices. Cultural decisions affecting palliative care also include attention to gender, age, generation, education level, diet/food, and ritual. Clinicians should ask the patient/family about these considerations.

- The cultural assessment promotes patient/family-centered decision-making and offers the opportunity to identify care practices.

Guideline 8. Spirituality is recognized as an integral part of the palliative care plan. Clinicians should screen for spiritual beliefs and practices and respond respectfully.

- Support of patient's spiritual needs at end of life is associated with better quality of life. Attending to a patient's spirituality can deepen the relationship between patient and clinical and build trust.

What harms were identified?

Guideline 2. Clinicians should use validated assessment tools, where available, to assess palliative care needs for each of these domains: physical, cultural, psychological, social, spiritual, and ethnical/legal.

- Non-standardized assessment may lead to specific needs going unaddressed.

Guideline 3: 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.

- Opportunity costs and limited available resources may be a barrier. Systems may have difficulty capturing, storing and accessing advance care planning documents when needed.

Guideline 4: The control of physical symptoms is an important part of the palliative care plan.

- Patients may fear that accepting palliative care interventions will shorten their lives.

Guideline 5. The cultural assessment promotes patient/family-centered decision-making and offers the opportunity to identify care practices. Cultural decisions affecting palliative care also include attention to gender, age, generation, education level, diet/food, and ritual. Clinicians should ask the patient/family about these considerations.

- Clinicians may underestimate the striking differences between the culture of medicine and the distinct beliefs and traditions that patients may value.

Guideline 8. Spirituality is recognized as an integral part of the palliative care plan. Clinicians should screen for spiritual beliefs and practices and respond respectfully.

- Spiritual and religious concerns can at times create distress and increase the burden of illness.

Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?

☑ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Systematic Review

Evidence

Source of Systematic Review:

- Title
- Author
- Date
- Citation, including page number
- URL

The National Consensus Project Clinical Practice Guidelines for Quality Palliative from the National Coalition for Hospice and Palliative Care formalize and delineate available evidence-based processes and practices as well as consensus recommendations for the provision of safe and reliable high-quality palliative care for adults, children, and families with serious illness in all care settings. A systematic review of the evidence for the Guidelines was conducted, and the complete findings were published separately (see below).

Citation: National Consensus Project for Quality Palliative Care. Clinical Practice Guidelines for Quality Palliative Care, 4th edition. Richmond, VA: National Coalition for Hospice and Palliative Care; 2018.

URL: https://www.nationalcoalitionhpc.org/ncp

The supporting systematic review focuses on specific questions and synthesizes evidence for palliative care interventions to inform the National Coalition for Hospice and Palliative Care clinical practice guidelines.

Citation: Ahluwalia S, Chen C, Raaen L, Motala A, Walling A, Chamberlin M, O'Hanlon C, Larkin J, Lorenz K, Akinniranye O, Hempel S. (2018). A systematic review in support of the National Consensus Project Clinical Practice Guidelines for Quality Palliative Care, Fourth Edition. *Journal of Pain and Symptom Management, 56* (6), 831-870.

URL: https://doi.org/10.1016/j.jpainsymman.2018.09.008

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

Domain 2. Physical Aspects of Care

- Guideline 2.1. The palliative care interdisciplinary team (IDT) endeavors to relieve suffering and improve quality of life, as defined by the patient and family, through the safe and timely reduction of the physical symptoms and functional impairment associated with serious illness.
- Guideline 2.2. The IDT assesses physical symptoms and their impact on well-being, quality of life, and functional status.
- Guideline 2.3. Interdisciplinary care plans to address physical symptoms, maximize functional status, and enhance quality of life are developed in the context of the patient's goals of care, disease, prognosis, functional limitations, culture, and care setting. An essential component of palliative care is ongoing management of physical symptoms, anticipating changes in health status, and monitoring of potential risk factors associated with the disease and side effects due to treatment regimens.

Domain 5. Spiritual, Religious, and Existential Aspects of Care

- Guideline 5.1. Patient and family spiritual beliefs and practices are assessed and respected.
- Guideline 5.2. The spiritual assessment process has three distinct components – spiritual screening, spiritual history, and a full spiritual assessment. The spiritual screening is conducted with every patient and family to identify spiritual needs and/or distress. The history and assessment identify the spiritual background, preferences, and related beliefs, values, rituals, and practices of the patient and family. Symptoms, such as spiritual distress and spiritual strengths and resources, are identified and documented.

Domain 8. Ethical and Legal Aspects of Care

- Guideline 8.1. The core ethical principles of autonomy substituted judgment, beneficence, justice, and nonmaleficence underpin the provision of palliative care.
- Guideline 8.3. The patient's preferences and goals for medical care are elicited using core ethical principles and documented.

Grade assigned to the **evidence** associated with the recommendation with the definition of the grade

The evidence was reviewed using the GRADE approach, in which each outcome is rated as one of four categories: Very Low, Low, Moderate, or High (Guyatt et al., 2011 (Intro to GRADE)). Definition of assigned grades below:

Very low: The true effect is probably markedly different from the estimated effect.

Low: The true effect might be markedly different from the estimated effect.

Moderate: The authors believe that the true effect is probably close to the estimated effect.

Much of the evidence for palliative care remains **low** quality, due to inconsistency in study findings, the lack of precise effect estimates to support the effectiveness of interventions, and large variation in study designs, with few RCTs that allow strong evidence statements contributing to the evidence base.

Domain 2. Physical Aspects of Care

- The systematic review addressed the following key question: KQ2) What is the impact of palliative care interventions on physical symptom screening, assessment, and management of patients?
- Much of the evidence in this domain is **low** quality largely due to inconsistent findings regarding the impact of interventions on symptoms.

Domain 5. Spiritual, Religious, and Existential Aspects of Care

- The systematic review addressed the following key question: KQ5) What is the effect of a spiritual assessment and/or interventions on patient and family/caregiver spiritual and emotional wellbeing?
- There is **moderate** quality evidence for the positive impact of life review/dignity therapy on spiritual well-being.
- There is **very low** quality evidence for the positive impact of spiritual/religious interventions on spiritual well-being.
- There is **very low** quality evidence for a positive impact of other meaning-centered interventions.

Domain 8. Ethical and Legal Aspects of Care

- The systematic review addressed the following key question: KQ8) What is the impact of advance care planning on substituted decision-making regarding life-sustaining treatments?
- There is **moderate** quality evidence that advance directive interventions lead to preference-concordant care and to increased preference documentation.
- There is **moderate** quality evidence for the positive impact of care planning discussions on preference concordant care, concordance between patient and family wishes, and documentation of advance care planning processes and documents.

Provide all other grades and definitions from the evidence grading system

High: The authors have a lot of confidence that the true effect is similar to the estimated effect.

(https://bestpractice.bmj.com/info/us/toolkit/learn-ebm/what-is-grade/)

Grade assigned to the **recommendation** with definition of the grade

The NCP Guidelines do not use GRADE to describe the strength of recommendations; rather, the Guidelines formalize and delineate available evidence-based processes and practices as well as consensus recommendations. Practitioners are encouraged to use the NCP Guidelines to strengthen knowledge and skills to better meet the needs of people living with serious illness.

The systematic review focuses on the quality of evidence. It does not make any recommendations or assign a grade to the recommendations included in the Guidelines.

Provide all other grades and definitions from the recommendation grading system

Body of evidence:

- Quantity – how many studies?
- Quality – what type of studies?

Domain 2. Physical Aspects of Care

- 48 systematic reviews were identified pertaining to KQ2.

Domain 5. Spiritual, Religious, and Existential Aspects of Care

- 11 systematic reviews were identified pertaining to KQ5.

Domain 8. Ethical and Legal Aspects of Care

- 36 systematic reviews were identified pertaining to KQ8.

Estimates of benefit and consistency across studies

Domain 2. Physical Aspects of Care

- There was evidence demonstrating the positive impact of music/art therapy on pain management outcomes.
- There was evidence for the impact of a comprehensive palliative care team for adults on symptom burden.
- The evidence for pharmacological interventions showed inconsistent findings across studies and a lack of pooled effect estimates.

Domain 5. Spiritual, Religious, and Existential Aspects of Care

- There was evidence for the positive impact of life review/dignity therapy on spiritual well-being.
- There was very low quality evidence for the positive impact of spiritual/religious interventions on spiritual well-being.
- There was very low quality evidence for a positive impact of other meaning-centered interventions.

Domain 8. Ethical and Legal Aspects of Care

- There was evidence that advance directive interventions led to preference-concordant care and to increased preference documentation.

- There was evidence for the positive impact of care planning discussions on preference concordant care, concordance between patient and family wishes, and documentation of advance care planning processes and documents.
- There was inconsistency in findings related to the impact of decision aids as well as a lack of pooled effect estimates.

What harms were identified?

Domain 2. Physical Aspects of Care

- Physical concerns, including ongoing access to medications, can be exacerbated as patients transfer across settings of care.

Domain 5. Spiritual, Religious, and Existential Aspects of Care

Domain 8. Ethical and Legal Aspects of Care

- Familiarity with local and state laws is needed relating to advance care planning, decisions regarding life-sustaining treatments, and evolving treatments with legal ramifications (e.g., medical marijuana), especially when caring for vulnerable populations, such as minors, prisoners, or those with developmental disability or psychiatric illness.

Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?

1a.4 OTHER SOURCE OF EVIDENCE

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

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

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

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

- 1b. Performance Gap

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

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

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

The aim of this measure is to assess whether a comprehensive assessment is completed at hospice admission for each hospice patient based on seven QMs that assess high-priority care processes around admission as recognized by both leading hospice stakeholders and patients. Another key factor in creating a measure of comprehensive assessment at admission is to provide both consumers and providers with a single measure regarding the overall quality of the assessment of patient needs at hospice admission, which can then be easily used to compare quality across providers. Additionally, in the current HQRP QMs hospice performance scores are high with scores of 97% or higher; hospices perform lower on the Pain Assessment QM (95.3%). On average, 93.2% of patient stays in a hospice had documentation that all of these critical care processes were completed at admission. Thus, the comprehensive assessment measure sets a higher standard of care for hospices, and consequently reveals a larger performance gap. The performance gap identified by the comprehensive assessment measure creates opportunities for quality improvement and may motivate providers to conduct a greater number of high priority care processes for as many patients as possible upon admission.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile*

range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.

2 Based on 1,418,348 patient stays discharged from 4,089 hospices between Q4 2018 and Q3 2019, on average, 89.3% of patient stays in a hospice had documentation that all of seven HQRP desirable care process were completed at admission. Below we present the hospice-level score distribution for the comprehensive assessment measure based on the year of patient stay admissions, representing 4 four calendar years of data. This includes data from January 1, 2016 to December 31, 2019. Additionally, we assessed the QM's variability by focusing on the percentage of hospices with perfect scores (i.e., critical care processes performed for 100 percent of patients), the percentage of lower-performing hospices, and the interquartile range of the QM scores. A meaningful and useful QM should be able to distinguish between high- and low-quality hospices within a given reporting period and have sufficient variability across providers. This analysis produced the following results based on hospices meeting the minimum reporting threshold:

- Over time, the hospice-level mean score for this QM increased from 77.8% for patient stays admitted in 2016 to 89.6% in 2019, the median increased from 82.3% to 94.1%, the interquartile range (IQR) decreased from 22.9% to 12.0%, and the standard deviation (SD) decreased from 18.2% to 12.7%.
- For patient stays admitted between January 1, 2016 and December 31, 2016, only 2.2 percent of hospices (81 of 3,745 hospices) had perfect scores, and 34.6% of hospices scored lower than 75% on this QM. In 2019, 8.4% percent of hospices (344 of 4,080 hospices) had perfect scores, and 11.0% of hospices scored lower than 75% on this QM.

Results from these trend analysis suggest that hospices' performance on this QM has been improving overtime. However, a performance gap still remains, supporting the importance of this measure.

			Percentile					
CY of Admission	N	Mean	Std. Dev	10th	25th	50th	75th	90th
2016	3,745	77.8%	18.2%	51.7%	68.9%	82.3%	91.8%	96.8%
2017	3,860	83.9%	14.8%	65.0%	77.5%	88.0%	94.9%	98.3%
2018	4,041	88.1%	13.3%	70.6%	83.6%	92.6%	97.4%	99.3%
2019	4,080	89.6%	12.7%	73.3%	86.0%	94.1%	98.0%	99.7%

CY of Admission	N with Perfect Score	% with Perfect Score
2016	81	2.2%
2017	124	3.2%
2018	229	5.7%
2019	344	8.4%

CY of Admission	N with Average Less than 75% Score	% Less than 75%
2016	1,296	34.6%
2017	807	20.9%
2018	533	13.2%
2019	448	11.0%

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

Data are available and described in 1b.2.

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

We conducted a series of disparity analyses at both the patient stay and the hospice levels to assess how the measures are affected by the sociodemographic characteristics of hospice patients. At the patient stay level, we compared the QM scores across sociodemographic groups. At the hospice level, we compared the distribution of QM scores between hospices with varying population sociodemographic characteristics. Analyses were conducted using HIS data from Q4 2018 through Q3 2019. Patient-level analyses encompassed 4,566 hospice organizations and 1,422,921 patient stays across the United States. Hospice-level analyses reflect 4,089 hospices that met the minimum reporting threshold. The results of these analyses are presented below.

Racial and ethnic disparity analysis. We conducted racial and ethnic disparity analyses at both the patient stay and hospice levels. At the patient stay level, we compared the percentage of patients who received all seven care process 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, as Asian, as Native Hawaiian or other Pacific Islander, or as more than one race. A Chi-square test was performed to determine if there were any statistically significant differences in completion of all seven care processes between groups. The lowest rate of completion of the comprehensive assessment was found for other non-Hispanic patients (90.9%), and the highest rate was among patients identifying as White non-Hispanic (93.5%). Differences in the rate of completion of all seven care processes by racial identification were found to be statistically significant (p-value < 0.01).

Analyses at the hospice level examined the differences in this measure across two groups: hospices with proportions of nonwhite patients that are less than or equal to the national median (18.2%), and hospices with proportions of nonwhite patients that are greater 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 nonwhite. 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 (88.1% for hospices with a proportion of nonwhite patients greater than the national median compared with 90.6% for hospices with a proportion of nonwhite patients less than or equal to the national median, p-value < 0.01).

Gender disparity analysis. We performed both the patient stay and hospice-level analyses on gender disparity using the same methods described in the racial and ethnic disparity analyses above. Females received all 7 care processes at a similar rate to male patients (93.17% and 93.22% respectively). Differences in the rate of completion of all 7 care processes by gender were not statistically significant (p-value = 0.314). At the hospice level, we examined the differences in this measure across two groups: hospices with proportions of female patients that are less than or equal to the national median (55.1%), and hospices with proportions of female patients that are greater than the national median. The results showed that the QM score was not statistically significantly different between the two groups of hospices (89.2% for hospices with a proportion of female patients above the national median compared with 89.5% for hospices with a proportion of female patients at or below the national median, p-value = 0.375).

Socioeconomic disparity analysis. We performed socioeconomic disparity analyses at both the patient stay and hospice levels using the same methods described in the racial and ethnic disparity analyses above. Medicaid status was used as a proxy measure of low socioeconomic status. There were similar rates of comprehensive assessment completion for non-Medicaid patients (93.0%) compared with Medicaid patients (93.2%) and the

difference was not statistically significant at p-val < 0.01. At the hospice level, the results showed that the QM score was significantly different between hospices with proportions of Medicaid patients less than or equal to the national median (14.1%) and hospices with proportions of Medicaid patients greater than the national median (90.6% for hospices above the median compared with 88.2% for hospices below the median, p-value < 0.01). The statistically insignificant results at the patient stay level do not indicate that quality of hospice care for Medicaid patients, measured by the completion of the comprehensive assessment, is better than for non-Medicaid patients. The significant findings at the hospice level indicate that hospices with a greater proportion of Medicaid patients are more likely to complete all 7 care process for every patient at admission.

Rural-urban disparity analysis. We compared the average QM score between rural and urban hospices using a Wilcoxon-Mann-Whitney test. The results showed that the average QM score was not significantly different between rural and urban hospices (89.1% for urban hospices compared with 89.4% for rural hospices, p-value = 0.377). This indicates that rural hospices and urban hospices perform similarly on the comprehensive assessment measure.

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

Data are available and described in 1b.4.

- 1c. Composite Quality Construct and Rationale

1c.1. A composite performance measure is a combination of two or more component measures, each of which individually reflects quality of care, into a single performance measure with a single score.

For purposes of NQF measure submission, evaluation, and endorsement, the following will be considered composites:

- Measures with two or more individual performance measure scores combined into one score for an accountable entity.
- Measures with two or more individual component measures assessed separately for each patient and then aggregated into one score for an accountable entity:
    - all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient);

1c.1. Please identify the composite measure construction: all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient)

1c.2. Describe the quality construct, including:

- the overall area of quality
- included component measures and
- the relationship of the component measures to the overall composite and to each other.

The Hospice and Palliative Care Composite Process Measure—Comprehensive Assessment at Admission (herein after referred to as the 'Hospice Comprehensive Assessment Measure') is designed to reflect the overall quality of comprehensive assessment at hospice admission for each patient stay. This measure captures whether a comprehensive assessment is completed at hospice admission by assessing the number of individual care processes completed upon admission for each hospice patient stay. These individual care processes will be based on the seven NQF-endorsed quality measures (QMs) currently implemented in the Centers for Medicare & Medicaid Services (CMS) Hospice Quality Reporting Program (HQRP). These seven measures capture care processes for five different domain areas: pain screening (NQF #1634) and comprehensive pain assessment (NQF #1637), dyspnea screening (NQF #1639) and treatment (NQF #1638), patients treated with an opioid who are given a bowel regimen (NQF #1617), patient preferences for life-sustaining treatments (NQF #1641), and spiritual and existential concerns (NQF #1647). All seven measures address high-priority aspects of quality hospice care as identified by the National Consensus Project, are

required by the Medicare Hospice Conditions of Participation, and are supported by hospice stakeholders. The Comprehensive Assessment Measure, therefore, reflects the overall quality of comprehensive assessment at hospice admission for each patient stay.

1c.3. Describe the rationale for constructing a composite measure, including how the composite provides a distinctive or additive value over the component measures individually.

The Hospice Comprehensive Assessment Measure provides the overall quality of assessment of patient needs at hospice admission, quality information that cannot be determined from any single measure alone. By assessing whether a comprehensive assessment is performed at hospice admission in a single measure, this measure will (1) incentivize hospices to conduct all critical care processes for each patient; (2) set a higher standard of care for hospices, which will reveal a larger performance gap and thus room for improvement (discussed below); and (3) provide consumers and providers with a single measure representing the overall quality and completeness of assessment of patient needs at hospice admission, which can be easily used to compare quality processes across providers.

For all seven existing component measures, hospice performance scores are 95% or higher; most patients are receiving individual care processes recommended as part of high quality care at admission. For example, on average, 97.3% of patient stays in a hospice had documentation showing that the patient received a pain screening at admission. On the other hand, hospices on average score 89.3% on the Comprehensive Assessment Measure revealing still room for improvement and thus creating incentives for hospices to improve quality.

1c.4. Describe how the aggregation and weighting of the component measures are consistent with the stated quality construct and rationale.

The Hospice Comprehensive Assessment uses an all-or-none scoring approach. This scoring approach calculates the percentage of patients who received all seven HQRP care processes at admission. In other words, the seven component measures are equally weighted to create the composite measure. These seven QM components are a standard practice by any hospice provider for every patient, where each component should be conducted as part of the initial comprehensive assessment. Therefore, the all-or-none scoring approach sets a clear quality expectation that all care processes captured by the seven QM components are expected to be performed, and missing any one of these measures could be recognized as an indicator of suboptimal care. Our environmental scan and Technical Expert Panel (TEP) also supported the all-or-non scoring approach. The results from our analyses presented above demonstrate a performance gap and highlight the potential for quality improvement across hospice providers. The all-or-none scoring approach is most effective in ensuring that a multidimensional assessment has been completed for all patients on admission.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

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

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

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

https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Hospice-Quality-Reporting/Current-Measures.html

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

This is not an eMeasure **Attachment:**

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

No data dictionary **Attachment:**

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

Yes, this is an instrument-based measure **Attachment:** HQRP-HIS-v2000-Admission-637320621175353646.pdf

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

Clinician

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

No

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

There are no significant changes to the measure specifications.

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

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

The numerator of this measure is the number of patient stays in the denominator where the patient received all 7 care processes which are applicable to the patient at admission, as captured by the current HQRP quality measures. To be included in the comprehensive assessment measure numerator, a patient must meet the numerator criteria for each of the individual component quality measure (QM) that is applicable to the patient. The numerator of this measure accounts for the three conditional measures in the current HQRP (NQF #1637 Pain Assessment, NQF #1638 Dyspnea Treatment, and NQF #1617 Bowel Regimen) as described below.

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

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

The numerator of this measure is the number of patient stays in the denominator where the patient received all the 7 care processes which are applicable to the patient at admission, as captured in the current HQRP quality measures. This includes patients who received all 7 care process which are applicable to them at

admission, as well as patients for whom the three individual conditional component QMs do not apply. The numerator criteria for the individual measures are:

1. NQF 1634: Patient stays that include a screening for the presence or absence of pain (and if present, rating of its severity) using a standardized quantitative tool during the admission evaluation for hospice / initial encounter for palliative care.
2. NQF 1637: Patient stays who received a comprehensive clinical assessment to determine the severity, etiology and impact of their pain within 24 hours of screening positive for pain.
3. NQF 1639: Patient stays that include a screening for the presence or absence of dyspnea and its severity during the hospice admission evaluation / initial encounter for palliative care.
4. NQF 1638: Patient stays that include a positive screening for dyspnea who received treatment within 24 hours of screening.
5. NQF 1617: Patient stays that are given a bowel regimen when appropriate or there is documentation as to why this was not needed
6. NQF 1641: Patient stays with a medical record that includes documentation of life sustaining preferences
7. NQF 1647: Patient stays with a medical record that includes documentation that the patient and/or caregiver was asked about spiritual/existential concerns within 5 days of the admission date.

Therefore, the numerator for this measure includes all patient stays from the denominator in which the patient meets the numerator criteria for all of the individual component QMs. Patient stays are included in the numerator if they meet the following criteria:

1. The patient/responsible party was asked about preference regarding the use of cardiopulmonary resuscitation (F2000A = [1,2]) OR preferences regarding life-sustaining treatments other than CPR (F2100A = [1,2]) OR preference regarding hospitalization (F2200A = [1,2]) no more than 7 days prior to admission or within 5 days of the admission date (-7 = F2000B – A0220 = 5 and F2000B ? [-,^])

AND

2. The patient and/or caregiver was asked about spiritual/existential concerns (F3000A = [1,2]) no more than 7 days prior to admission or within 5 days of the admission date (-7 = F3000B – A0220 = 5 and F3000B ? [-,^])

AND

3. The patient was screened for pain within 2 days of the admission date (J0900B - A0220 = 2 and J0900B ? [-,^]) and reported that they had no pain (J0900C = [0]) OR The patient was screened for pain within 2 days of the admission date (J0900B - A0220 = 2 and J0900B ? [-,^]), the patient's pain severity was rated mild, moderate, or severe (J0900C = [1,2,3]), and a standardized pain tool was used (J0900D = [1,2,3,4]))

AND*

4. A comprehensive pain assessment was completed within 1 day of the initial nursing assessment during which the patient screened positive for pain (J0910B – J0900B = 1 and J0910B and J0900B ? [-,^]) and included at least 5 of the following characteristics: location, severity, character, duration, frequency, what relieves or worsens the pain, and the effect on function or quality of life (5 or more items in J0910C1 – J0910C7 checked and not all J0910C boxes = [-,^])

AND

5. The patient was screened for shortness of breath within 2 days of the admission date (J2030B - A0220 = 2 and J2030B ? [-,^])

AND*

6. The patient declined treatment (J2040A = [1]) OR Treatment for shortness of breath was initiated prior to the initial nursing assessment or within 1 day of the initial nursing assessment during which the patient screened positive for shortness of breath (J2040B – J2030B = 1 and J2040B and J2030B ? [-,^])

AND*

7. There is documentation of why a bowel regimen was not initiated or continued (N0520 = [1]) OR A bowel regimen was initiated or continued within 1 day of a scheduled opioid being initiated or continued (N0520B – N0500B = [1] and N0520B and N0500B ? [-,^])

NOTE: *denotes paired measures. For some patient stays, the second component of the paired measure may not be applicable. In this instance, in the calculation of the comprehensive assessment measure, the patient will be included in the numerator for the composite measures as long as the patient meets the numerator criteria for the first measure in the pair as if hospices completed both care processes for the patients. For example, if a patient screened negative for pain, the comprehensive pain assessment measure will not be applicable, however, in the comprehensive assessment measure, the hospice would be 'given credit' for completing the comprehensive pain assessment. This logic also applies to NQF #1617 Bowel Regimen. While NQF #1617 is not a paired measure, the patient must have a scheduled opioid initiated or continued in order to complete item N0520, which assess whether a bowel regimen was initiated or continued.

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

The denominator for the measure includes all hospice patient stays enrolled in hospice except those with exclusions.

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

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

The denominator for the measure includes all hospice patient stays except for those with exclusions as identified in S.8 and S.9 below.

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

Patient stays are excluded from the measure if they are under 18 years of age, or are a Type 2 (discharged stays missing the admission record) or Type 3 patient stay (active stays).

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

The exclusion criteria are:

1. Patients under 18 years of age as indicated by the birth date (A0900) and admission date (A0220)
2. Patients with Type 2 (discharged stays missing the admission record) and Type 3 patient stays (active stays)

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

N/A

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

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Higher score

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

Step one: Calculate the total number of Type 1 stays that do not meet the exclusion criteria.

Step two: Calculate the number of patient stays where the patient meets the numerator criteria for all the individual component QMs, that is, the number of patient stays where each patient received all care processes at admission for which the patient is eligible. This includes patients who are eligible for and received all 7 care process at admission, as well as patients who may not be included in the individual paired component QMs.

Step three: Divide the hospice's numerator count by its denominator count to obtain the hospice's observed score; that is, divide the result of step (2) by the result of step (1). The quality measure score is converted to a percent value by multiplying by 100.

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

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

N/A

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

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

N/A

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

If other, please describe in S.18.

Other

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

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

Hospice Item Set (HIS). The HIS is a standardized, patient-level data collection instrument part of the HQRP as finalized 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.

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

Available in attached appendix at A.1

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

Facility

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

Other

If other: Hospice

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

All component measures included in the comprehensive assessment measure are weighted equally and have all received NQF endorsement.

2. Validity – See attached Measure Testing Submission Form

NQF_Measure_Testing_Form_final_final.docx

2.1 For maintenance of endorsement

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

Yes

2.2 For maintenance of endorsement

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

Yes

2.3 For maintenance of endorsement

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

No - This measure is not risk-adjusted

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

Measure Number (*if previously endorsed*)**:** 3235
Composite Measure Title: Hospice and Palliative Care Composite Process Measure—Comprehensive Assessment at Admission
Date of Submission: 8/3/2020

Composite Construction:

☐Two or more individual performance measure scores combined into one score

☒ All-or-none measures (e.g., all essential care processes received or outcomes experienced by each patient)

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

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

The Hospice Item Set (HIS), a standardized, patient-level data collection instrument, was implemented by the Centers for Medicare and Medicaid Services (CMS) 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 V1.00 collects data to calculate seven quality measures which are endorsed by the National Quality Forum (NQF). These measures (NQF #1634, NQF #1637, NQF #1639, NQF #1638, NQF #1617, NQF #1647, NQF #1641) focus on patient preferences regarding life-sustaining treatments; care for spiritual and existential concerns; and management of pain, dyspnea, and bowels. These measures are the components of the composite measure. Three of the components are "paired" measures (NQF #1637, NQF #1638, NQF #1617), such that they are only relevant for a sub-set of patient stays. For these measures, patients for whom the component is not relevant are automatically included in the component's numerator for the purposes of calculting the composite measure. For example, if a patient screened negative for pain, they are not eligible for the component pain assessment measure, however, in the composite measure, they would be included in the numerator for the comprehensive pain assessment.

The analyses for this measure were conducted on HIS-Admission and HIS-Discharge records for patient stays admitted between October 1, 2018 and September 30, 2019. The stay-level analysis included 1,422,921 patient stays in 4,566 hospices. The hospice-level analysis includes 4,089 hospices which met the minimum reporting threshold of 20 stays within the period.

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

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒ other: Hospice Item Set (HIS)	☒ other: Hospice Item Set (HIS)

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

The dataset used for testing is the Hospice Item Set (HIS).

1.3. What are the dates of the data used in testing? Patient stays admitted between October 1, 2018 and September 30, 2019.

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

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.26*)	Measure Tested at Level of:
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other:	☐ other:

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

All United States accountable entities and patients were included. Analyses encompassed 3,922 Medicare-certified hospice organizations across the United States. Characteristics of these providers are as follows:

Facility Type:

- Affiliated with Hospital: 8.07%
- Affiliated with Skilled Nursing Facility: 0.20%
- Affiliated with Home Health: 9.44%
- Freestanding: 81.19%

Profit status

- Nonprofit: 23.26%
- For-profit: 62.92%
- Government and other: 12.72%
- Missing Data: 0.84%

Regions:

- South: 36.95%
- West: 29.22%
- Midwest: 21.62%
- Northeast: 10.03%
- Territories: 1.08%
- Unknown/Missing Data: 1.10%

Urban/Rural

- Urban: 78.97%
- Rural: 19.93%

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

Analyses encompassed 1,422,912 hospice patient stays across the United States for individuals aged 18 and older, representing all patients stays served by the 4,566 Medicare-certified hospices during the analysis period. Average patient characteristics are as follows:

Race\ethnicity:

- White non-Hispanic: 77.42%
- Black non-Hispanic: 8.69%
- Other non-Hispanic: 3.26%
- Hispanic: 7.96%
- Missing: 2.63%

Sex:

- Male: 45.14%
- Female: 54.86%

Primary diagnosis:

- Cancer: 31.92%
- Dementia/Alzheimer's: 14.77%
- None of the above: 53.31%

Mean Age: 79.56

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

All hospice-level analyses include hospice providers with a minimum denominator size of 20 patient stays to yield statistically meaningful QM scores. All stay-level analyses include the entire data set and do not apply a minimum denominator size.

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

We performed socioeconomic disparity analyses at both the patient stay and hospice levels. Medicaid status was used as a proxy measure of low socioeconomic status. At the patient level, we examined the rate of receiving the composite measure (i.e. receiving all seven care processes) between Medicaid and non-Medicaid patients and found statistically significant differences at the $p < 0.01$ level. However, the magnitude of the difference was minimal. The average composite score was 93.05 percent among stays for Medicaid patients and 93.23 percent among stays for non-Medicaid patients. At the hospice level, we examined the differences in this measure between hospices with proportions of Medicaid patients less than or equal to the national median and hospices with proportions of Medicaid patients greater than the national median. In the HIS data for October 1, 2018 to September 30, 2019, the national median of the proportion of Medicaid patients in a hospice is 14.13 percent. We found a statistically significant difference at the $p <0.01$ level in average composite scores for hospices above or below the national median. Again, the magnitude of the difference in

average scores was minimal; the average scores were 90.60 percent and 88.02 percent for hospices above and below the median, respectively.

2a2. RELIABILITY TESTING

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

Note*: Current guidance for composite measure evaluation states that reliability must be demonstrated for the composite performance measure score.*

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. 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*)

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 providers with a minimum denominator size of 20 patient stays, and used the Intraclass 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 divided the average of composite scores at the hospice-level by the standard deviation of composite scores across all patient stays to determine the extent to which total variance in the measure is attributable to differences among providers. Higher ratios indicate better reliability.

Stability analysis. Stability analysis describes the extent to which providers' performance assessed by a QM changes across time. Dramatic and unsystematic changes in the QM score across time may indicate measure instability rather than true changes in quality. Therefore, QMs with higher stability are considered to be more reliable. We examine changes in hospice-level QM scores between four consecutive quarters (Q4 2014 and Q1 2015, Q1 and Q2 2015, and Q2 and Q3 2015) and reported changes in facility scores in standard deviations.

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

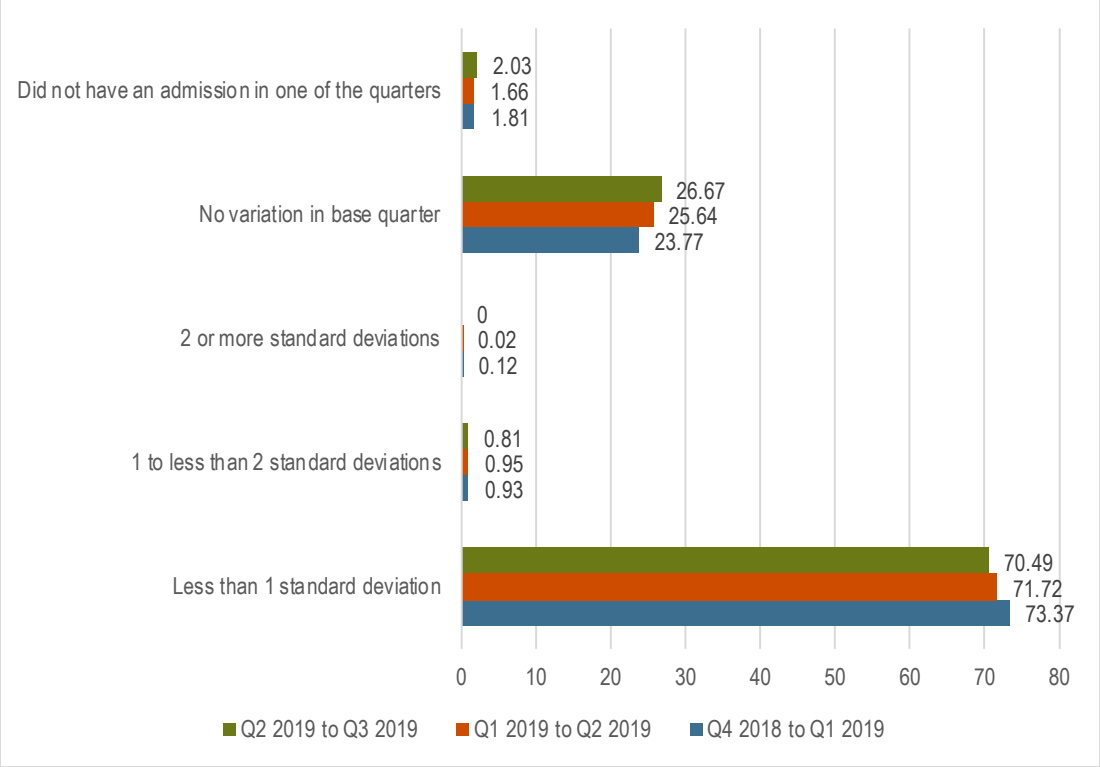
Split-half reliability. The ICC coefficient for this measure is 0.86.

Signal-to-noise analysis. The signal-to-noise ratio for this measure is 3.55.

Stability analysis. ***Figure 1*** below illustrates the observed change in facility scores between the four consecutive quarters analyzed (Q4 2014 and Q1 2015, Q1 and Q2 2015, and Q2 and Q3 2015), where the changes in facility scores are reported in standard deviations.

Approximately 70% of providers had a change in QM score of less than one standard deviation. The number of providers with a change in QM score of less than one standard deviation increased from 70.44% to 73.39% across quarters. Less than 1 percent of providers had a change in QM score between one and two standard deviations and very few providers had a change in QM within two or more standard deviations. For almost a quarter of hospices, there was no variation in the base quarters. Hospices with no variation in the composite score had the same composite score for all patient stays. About 2% of hospices did not have an admission in one or both quarters.

Figure 1: Standardized Score Change in QM Score from Q4 2014 to Q3 2015



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

This comprehensive assessment measure has high reliability, demonstrated by high stability, internal consistency, and signal-to-noise ratio.

Split-half reliability. An ICC coefficient of 0.86, as observed, indicates high internal reliability for the comprehensive assessment measure.

Signal-to-noise analysis. The analysis results found the signal-to-noise ratio to be 3.56, indicating a high reliability for this measure.

Stability analysis. The results of this analysis indicate that facility scores were very stable. The high rate of providers that had a change in QM score of less than one standard deviation (from 70.49 percent to 73.37 percent), indicates high stability of the comprehensive assessment QM and suggests the measure is generally stable.

2b2. VALIDITY TESTING

***Note**: Current guidance for composite measure evaluation states that validity should be demonstrated for the composite performance measure score. If not feasible for initial endorsement, acceptable alternatives include assessment of content or face validity of the composite OR demonstration of validity for each component. Empirical validity testing of the composite measure score is expected by the time of endorsement maintenance.*

2b2.1. What level of validity testing was conducted?

☒ **Composite performance measure score**

☐ **Empirical validity testing**

☐ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

☐ **Systematic assessment of content validity**

☒ **Validity testing for component measures** *(check all that apply)*

***Note**: applies to ALL component measures, unless already endorsed or are being submitted for individual endorsement.*

☒ **Endorsed (or submitted) as individual performance measures**

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

☐ **Empirical validity testing of the component measure score(s)**

☐ **Systematic assessment of face validity of component measure score(s) as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

Correlations. Providers should perform similarly on QMs that reflect the quality of similar care processes. Thus, a common strategy used to evaluate validity is to examine the correlation among measures that capture related clinical care processes: convergent validity. We conducted nonparametric Spearman rank correlation analysis between the composite measure and all seven HQRP QMs. In this analysis, we did not adjust the paired QMs' numerators and denominators for align with the composite measure. We used nonparametric methods because the data are heavily skewed (most providers perform well, while a small proportion have lower scores). The magnitude of the Spearman correlation coefficient becomes larger as the correlation between the two QM scores become stronger, with 0 indicating no correlation and –1 and 1 indicating a perfect monotonic relationship. The absolute value of the Spearman correlation coefficient indicates the strength of the correlation. In general, a value less than 0.4 indicates weak correlation; a value between 0.4 and 0.6 indicates moderate correlation; a value of 0.6 and above indicates strong correlation.

Missing data. We also conducted analysis of missing data to support the validity of this measure. The method description and results are presented in section 2b7. Below.

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

Correlations. The p-values for all the Spearman correlation coefficients are significant (p-value < 0.01). There are significant positive correlations between the composite measure and each of the QMs. ***Table 1*** below presents the Spearman correlation coefficients for all 1,422,921 patient stays.

Table 1: Correlation of Composite Measure and Component QMs for Patient Stays

Quality measure	NQF #1641 Treatment Preferences	NQF #1647 Beliefs & Values	NQF #1634 Pain Screening	NQF #1637 Pain Assessment	NQF #1639 Dyspnea Screening	NQF #1638 Dyspnea Treatment	NQF #1617 Bowel Regimen
Comprehensive Assessment QM	0.26*	0.47*	0.50*	0.73*	0.33*	0.47*	0.46*

* indicates significant correlation at p-value < 0.01

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

This composite measure has moderate to high validity, demonstrated by positive, significant correlations of the measure to each of the individual QMs. The significant and positive correlations observed between the composite measure and each of the QMs indicates that high-performing hospices in the component measures are more likely to be high-preforming on the composite measure. Overall, the correlations between the QMs are weak or moderate; if the correlations were very high (closer to a value of 1), then this would indicate that two measures share overlapping quality information and would have diminished the value of developing this measure. The results of the missing analysis, presented in section 2b7. below also support the validity of this measure.

_____

2b3. EXCLUSIONS ANALYSIS

Note: *Applies to the composite performance measure, as well all component measures unless they are already endorsed or are being submitted for individual endorsement.*

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

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

The composite measure excludes patients less than 18 years from the measure denominator. This exclusion criterion is consistent with the one used for the component measures. We examined the effect of excluding patients less than 18 years of age from the measure denominator on QM performance for the component measures.

To test the impact of this exclusion on QM performance, we conducted three sets of analyses.

1. At the patient stay level, we examined the proportion of patient stays that are excluded from the denominator.
2. At the hospice level, we examined the distribution of hospice-level exclusion rates by each exclusion criterion.
3. At the QM level, we examined the distribution of hospice QM scores with and without applying the exclusion criteria.

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Table 2 below describes the prevalence of the age exclusion at the patient stay level and compares the percentage of patients with and without this exclusion that meet the criteria for the measure numerator (for example, patients were screened for pain within 2 days of hospice admission).There were 3,112 pediatric patients that comprised 0.22% of 1,426,033 total patient stays. From the results in this table, we see that non-pediatric patients are only slightly more likely to receive the desired care process compared with pediatric patients (for example, 96.88 percent received a pain screening within 2 days of admission, compared with 98.24 percent among patients who were at least 18 years old). This suggests that non-pediatric patients are

almost as likely to receive the desired care process compared with pediatric patients.

Table 2: Percentage of Patients Counted in the Component QM Numerator by Age

< 18 years of age exclusion	Number of pediatric patients in the numerator	Percentage of pediatric patients in numerator	Percentage of non-pediatric patients in numerator
NQF #1641	3,058	98.26%	99.50%
NQF #1647	3,032	97.43%	98.42%
NQF #1634	3,015	96.88%	98.24%
NQF #1637	2,964	95.24%	97.18%
NQF #1639	3,067	98.55%	99.23%
NQF #1638	3,048	97.94%	98.51%
NQF #1617	3,092	99.36%	99.49%

Table 3 below describes the percent of patient stays that meet the age exclusion criteria in each hospice to determine if excluded patients are evenly distributed across facilities or if they disproportionately affect some hospices' denominator size. "N" represents the number of hospices with pediatric patients in the numerator. Out of 4,089 hospices, the number of hospices with pediatric patients included in the QM numerator ranges from 768 to 788, or approximately 19 percent of hospices. Across hospices with at least one pediatric patient stays included in the QM numerator, the mean percent of stays for patients < 18 years of age was at most 1.22 percent, the median was 0.55 percent or 0.56 percent, and in more than 99 percent of hospices, less than 10 percent of patients were pediatric patients.

< 18 years of age exclusion	N	Mean	SD	Percentile: 1	Percentile: 10	Percentile: 25	Percentile: 50	Percentile: 75	Percentile: 90	Percentile: 99	Min	Max
NQF #1641	781	1.22%	3.52%	0.05%	0.14%	0.27%	0.56%	1.17%	2.47%	9.09%	0.03 %	76.19%
NQF #1647	776	1.22%	3.53%	0.05%	0.13%	0.27%	0.56%	1.16%	2.50%	9.09%	0.03 %	76.19%
NQF #1634	775	1.22%	3.54%	0.05%	0.14%	0.27%	0.56%	1.16%	2.47%	9.09%	0.03 %	76.19%
NQF #1637	768	1.22%	3.55%	0.05%	0.13%	0.26%	0.55%	1.16%	2.50%	9.09%	0.03 %	76.19%
NQF #1639	780	1.22%	3.53%	0.05%	0.14%	0.27%	0.55%	1.16%	2.48%	9.09%	0.03 %	76.19%
NQF #1638	777	1.22%	3.53%	0.05%	0.13%	0.27%	0.55%	1.17%	2.50%	9.09%	0.03 %	76.19%
NQF #1617	788	1.21%	3.51%	0.05%	0.13%	0.26%	0.55%	1.16%	2.47%	9.09%	0.03 %	76.19%

Table 3: Distribution of Hospice-level Percent of Excluded Stays, by Component QM

Table 4 below examines the impact of excluding patients < 18 years of age on the mean QM score for each of the component measures. From the results below, we see that the mean QM scores among hospices are within 0.02 percentage points with and without the age exclusion.

Table 4: Hospice-level Mean QM Score: Impact of Excluding Young Patients by Component QM

Measure	N of hospices	N of Patient Stays	NQF #1641	NQF #1647	NQF #1634	NQF #1637	NQF #1639	NQF #1638	NQF #1617
Include < 18 years of age	4,090	1,421,450	99.21%	97.76%	97.25%	95.30%	98.79%	97.48%	98.98%
Exclude < 18 years of age	4,089	1,418,348	99.22%	97.76%	97.26%	95.32%	98.79%	97.48%	98.98%

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

The findings from these analyses indicate that the age exclusion affects only a small proportion of hospices nationwide, and that for most hospices that serve pediatric patients, the impact of the exclusion on denominator size is minimal. Additionally, the impact of the age exclusion on the mean QM score distribution is very small.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
Note: *Applies to all outcome or resource use component measures, unless already endorsed or are being submitted for individual endorsement.*
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used? *(check all that apply)*

☐ **Endorsed (or submitted) as individual performance measures**

☐ **No risk adjustment or stratification**

☐ **Statistical risk model**

☐ **Stratification by risk categories**

☐ **Other**

2b4.1.1 If using statistical risk models, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

2b4.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

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

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

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

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

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

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

2b4.9. Results of Risk Stratification Analysis:

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

_____

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

Note: *Applies to the composite performance measure.*

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

Confidence interval analysis. We examined proportions of hospices with the composite QM scores that are significantly different from the national hospice-level mean. If a high proportion of hospices have a composite measure score significantly different from the mean, the QM can identify providers 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 providers should have scores that are significantly above average, and low-performing providers should be significantly below average.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

The mean score for this QM was 89.31% with a range from 1.59% to 100%, the median was 93.64%, the interquartile range was 85.29% to 97.95%, and the standard deviation was 12.70%. For this QM, 313 hospices (7.65%) had perfect scores.

Across all hospices, 65.32% 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 (37.54% vs 27.78%, respectively).

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

The QM is able to identify those hospices that are performing well (higher than the national mean) and those that are performing lower than the national mean.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

Note: *Applies to all component measures, unless already endorsed or are being submitted for individual endorsement.*

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

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

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted?*)

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

Note: *Applies to the overall composite measure.*

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

The comprehensive assessment measure includes 7 HQRP measures, where each measure includes two or three items that can be missing. If missing, these items are coded as dashes and analyzed as though the process did not happen. In turn, they are not included in the comprehensive assessment measure numerator. For example, the pain screening measure includes three items that can include missing data – J0900B (pain screening), J0900C (pain severity), and J0900D (type of standardized pain tool used).

In order to assess how these missing data impact the comprehensive assessment measure, we conducted patient stay- and hospice-level analyses. For the patient stay-level analysis, we calculated the number and percentage of eligible patient stays for which any item from the seven individual component QM's on the HIS-Admission records was completed with a dash. For the hospice-level analysis we calculated each hospice's percent of eligible admissions for which any item from the seven individual component QM's was completed with a dash.

Below is the list of items that can be coded as missing on each of the individual component measures.

1. NQF #1634: For the pain screening measure, there are three items on the HIS that can include missing data – J0900B, J0900C, and J0900D
2. NQF #1637: For the pain assessment measure, there are one item on the HIS that can include missing data – J0910B.
3. NQF #1639: For the dyspnea screening measure, there are two items on the HIS that can include missing data – J2030B and J2030C.
4. NQF #1638: For the dyspnea treatment measure, there are two items on the HIS that can include missing data – J2040B and J2040C.
5. NQF #1617: For the bowel regimen measure, there are three items on the HIS that can include missing data – N0500B, N0510B, and N0520B
6. NQF #1641: For the treatment preferences measure, there are three items on the HIS that can include missing data – F2000B, F2100B, and F2200B.
7. NQF #1647: For the beliefs/values measure, there is only one item on the HIS that can include missing data – F3000B.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

Table 5 below presents the missing rate for each item used in the calculation of the composite QM. At the patient stay level, the overall rate of missing data ranged from 0.001 percent to 0.017 percent for any given item between October 2018 and September 2019. The missing rate for the majority of items were between 0.001 percent and 0.01 percent. Item J02030C, "Did the screening indicate the patient had shortness of breath?" had the highest missing rate of 0.013 percent.

Table 5: Stay Level Analysis of the Missing Item Rate

Item Number	Number of Missing (N)	Percentage of Missing (%)
J0900B	43	0.003%
J0900C	112	0.008%
J0900D	235	0.016%
J0910B	46	0.003%
J2030B	22	0.002%
J2030C	191	0.013%
J2040B	106	0.007%
J2040C1	22	0.002%
N0500B	61	0.004%
N0520B	85	0.006%
F3000B	55	0.004%
F2000B	21	0.001%
F2100B	30	0.002%
F2200B	29	0.002%

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

These results indicate that the missing rate is low and thus should have a negligible effect on the comprehensive assessment measure.

2c. EMPIRICAL ANALYSIS TO SUPPORT COMPOSITE CONSTRUCTION APPROACH

Note: *If empirical analyses do not provide adequate results—or are not conducted—justification must be provided and accepted in order to meet the must-pass criterion of Scientific Acceptability of Measure Properties. Each of the following questions has instructions if there is no empirical analysis.*

2d1. Empirical analysis demonstrating that the component measures fit the quality construct, add value to the overall composite, and achieve the object of parsimony to the extent possible.

2d1.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

A mixed methods approach was used. Qualitative information gathering activities—including an environmental scan, focus groups with 6 hospice patient caregivers (representing consumers), and a Technical expert panels involving clinical stakeholders —were conducted to support this measure concept and inform

the measure's specifications. Quantitative analyses conducted include a nonparametric Spearman rank correlation analysis between the composite measure and all seven HQRP QMs.

2d1.2. What were the statistical results obtained from the analysis of the components? (e.g., *correlations, contribution of each component to the composite score, etc.*; *if no empirical analysis, identify the components that were considered and the pros and cons of each*)

Correlations. ***Table 1*** above presents the Spearman correlation coefficients. The p-values for all the Spearman correlation coefficients are significant (p-value < 0.01). There are significant positive correlations between the comprehensive assessment measure and each of the QMs.

Focus Groups and Qualitative Information Gathering: Caregivers reported difficulties disentangling information from individual measures. Caregivers believed that a composite process measure assessing whether patients received a comprehensive assessment (covering physical symptoms, treatment preferences, and spiritual and existential concerns) upon hospice admission would help alleviate some of this confusion. Additionally, from the provider perspective, this measure is more actionable and identifies room for provider improvement.

2d1.3. What is your interpretation of the results in terms of demonstrating that the components included in the composite are consistent with the described quality construct and add value to the overall composite? (i.*e., what do the results mean in terms of supporting inclusion of the components; if no empirical analysis, provide rationale for the components that were selected)*

The significant and positive correlations observed between the comprehensive assessment measure and each of the QMs suggests that the composite measure is moderately and significantly related to the individual measures, sharing a common underlying focus. Overall, the correlations between the QMs are weak or moderate; if the correlations were very high (closer to a value of 1), then this would indicate that two measures share overlapping quality information and would have diminished the value of developing this measure. QMs with weak correlations are those that have a high percent of patient stays in the numerator and contribute little quality information to the overall composite measure.

2d2. Empirical analysis demonstrating that the aggregations and weighting rules are consistent with the quality construct and achieve the objective of simplicity to the extent possible

2d2.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

Impact of Each Component QM on Comprehensive Assessment QM Score: We assessed the contribution and impact of each individual component measure on the comprehensive assessment QM score by constructing seven 'alternative' composite measures. These alternative measures were built using only 6 out of the 7 component QMs, each time changing the QM that was omitted. In this analysis, we calculate seven different versions of the numerator and QM score, and compared these QM scores to the score of the complete comprehensive assessment measure.

Impact of Each Component QM on Hospice Score and Ranking: We examined the impact of each individual component measure on each hospice by creating 7 scores for each hospice based on the differences between the original composite mean score and one of the alternative mean scores created in in the analyses described above (impact of each component QM on Composite QM score). We also examined whether each of the alternative constructions were able to identify the same poor-quality outliers as the complete comprehensive assessment measure, based on the analysis of 200 hospices with the lowest score identified through each approach.

2d2.2. What were the statistical results obtained from the analysis of the aggregation and weighting rules? (e.g., *results of sensitivity analysis of effect of different aggregations and/or weighting rules; if no empirical analysis, identify the aggregation and weighting rules that were considered and the pros and cons of each*)

Impact of Each Component QM on Composite QM Score: The mean QM score calculated differed amongst all the alternative composite measures where a different component measure was removed. The mean QM score when all individual components were included was 93.19%. The mean QM scores ranged from a score of 93.22% when QM 1639 Dyspnea Screening was removed to a score of 94.91% when the QM 1637 Pain Assessment was removed. The mean QM score produced by each of the alternative construction approaches was higher than the mean score of the complete comprehensive assessment measure (that is, the difference in mean score between the complete comprehensive assessment QM and each of the alternative scores created was negative). Scores differed by 0.03% when QM 1639 Dyspnea Screening to 1.85% when QM 1637 Pain Assessment was removed.

Impact of Each Component QM on Hospice Ranking: ***Table 6*** below presents the overlapping outliers identified by each alternative approach. Each of the alternative measures identified a different but overlapping group of poor quality outliers, compared to those identified by the original composite measure. A smaller overlap between the outlier groups indicates more quality information contributed by the component measure that is missing from the alternative measure. Only 74.0% of the same outliers were identified when QM 1637 Pain Assessment was removed. The greatest overlap was seen when QM 1639 Dyspnea Screening was removed, capturing all of the same outliers. The majority of alternative approaches identified 90% of the outliers identified by the complete composite measure.

Table 6: Overlapping Outliers Identified in the Alternative QM Construction and Comprehensive Assessment QM

QM Removed	N	%
NQF #1641	195	97.5%
NQF #1647	157	78.5%
NQF #1634	179	89.5%
NQF #1637	148	74.0%
NQF #1639	200	100.0%
NQF #1638	185	92.5%
NQF #1617	192	96.0%

2d2.3. What is your interpretation of the results in terms of demonstrating the aggregation and weighting rules are consistent with the described quality construct? (i.*e., what do the results mean in terms of supporting the selected rules for aggregation and weighting; if no empirical analysis, provide rationale for the selected rules for aggregation and weighting*)

These results suggest that each QM does not meaningfully contribute to the composite assessment measure. In particular, the QMs with a high percent of overlapping outliers are those with average scores above 99 percent, such that there are few hospices where a significant proportion of patient stays do not meet the QM. These hospices are also those with low average composite scores.

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

If other:

3b. Electronic Sources

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

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

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

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

N/A

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

Attachment:

3c. Data Collection Strategy

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

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

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

Hospices administer the Hospice Item Set (HIS) to collect the information necessary to calculate the measure. The Hospice Quality Reporting Program employees a pay-for-reporting approach; since FY2020 and for all subsequent years, at least 90% of all required HIS records must be submitted and accepted within a 30-day submission deadline to avoid a 2 percentage-point payment rate increase reduction. Given the payment incentive, the vast majority of hospices submit information necessary to calculate the measure. Hospices are given 30 days after the patient´s admission data to submit data (and each patient only has one HIS admission assessment). For several months after data submission, hospices are encouraged to review their submission records for data accuracies prior to measure calculation. The HIS admission record was implemented with the

objective of calculating the seven component process measures from which this composite measure is calculated. As such, it is a shorter, lesser burdensome collection tool than those used in other care settings.

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

There are no fees, licensing, or other requirements to use any aspect of the measure as specified. The measure is part of CMS´s Hospice Quality Reporting Program, implemented with the goal of providing free and useful information to the public to empower patients and caregivers to make informed health care decisions. The data collection tool and measure specifications used to calculate the measure are publicly-available through the CMS website. CMS´s contractors calculate measure scores using data submitted by hospices. These scores are then publicly-reported with free access.

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
Quality Improvement (Internal to the specific organization)	Public Reporting https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Hospice-Quality-Reporting/Current-Measures Hospice Quality Reporting Program

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

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

The Hospice and Palliative Care Composite Process Measure—Comprehensive Assessment at Admission (NQF #3235) is currently publicly reported by the Centers for Medicare & Medicaid Services (CMS) on the Hospice Compare website. Hospice-level measure scores are reported for all nationwide hospices certified for Medicare, and publicly reported if there are at least 20 individuals in the denominator (non-suppressed scores are provided to all hospices). The analyses in this testing form closely approximate the number of providers and individuals in 4 quarters of data that are publicly reported. Our testing included 1.4 million patient stays in over 4,500 hospices.

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

The comprehensive assessment measure is currently publicly reported.

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

The comprehensive assessment measure is currently publicly reported.

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

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

The Hospice Comprehensive Assessment Measure was introduced to hospice providers at a National HQRP Provider Training on January 18, 2017. This training included information about the individual component measures and the measure specifications. Providers were given the opportunity to see a sample calculation of the Comprehensive Assessment Measure and ask questions about the measure. The materials and video recordings from this training session are accessible to providers on the HQRP website here: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Hospice-Quality-Reporting/Hospice-Quality-Reporting-Training-Training-and-Education-Library.html.

Providers have been able to view their performance results on the Comprehensive Assessment Measure in the Certification and Survey Enhanced Reporting (CASPER) system since February 2018. All Medicare-certified providers should have access to these CASPER QM reports. These CASPER QM reports provide hospice providers with feedback on their quality measure scores, both at the hospice and patient-stay level, helping them to improve the quality of care delivered. A factsheet is available to help hospice providers interpret these reports. The factsheet is available here: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Hospice-Quality-Reporting/Downloads/Fact-Sheet_CASPER-QM-Reports.pdf.

In advance of reporting on Hospice Compare, providers are given the opportunity to preview their HIS QM results during a 30-day preview period using a HIS Provider Preview Report. The purpose of these reports is to give providers the opportunity to preview their quality measure results on each quality measure prior to public display on Hospice Compare. The Comprehensive Assessment Measure was reported on HIS Provider Preview Reports in September 2018 and is regularly updated on Hospice Compare since 2018.

Finally, CMS provides a Help Desk to answer provider questions about all HQRP quality measures and reporting requirements. As part of the Help Desk functionality, CMS posts a HQRP Quarterly Update document at the end of each calendar year quarter that includes frequently asked Help Desk questions from the previous quarter and the corresponding responses on the CMS HQRP Requirements and Best Practices page here: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Hospice-Quality-Reporting/HQRP-Requirements-and-Best-Practices.html.

Additionally, in November 2018, CMS posted a factsheet on the CMS HQRP website that explains how the Hospice Comprehensive Assessment Measure is calculated and how providers can use their CASPER QM reports to understand their hospice's performance on the Hospice Comprehensive Assessment measure. Also, in December 2018, CMS hosted a live webinar training to explain the background of the measure, how the measure is calculated, and how providers can use their CASPER QM reports to understand their hospice's performance on the Hospice Comprehensive Assessment measure as well as provide the opportunity for providers to ask questions about the measure.

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

The CASPER QM reports allow providers to not only view national average scores on the Comprehensive Assessment Measure, but also specify a reporting period and view their own quality data at both the patient-

stay level and hospice level. These reports are on-demand and thus enable providers to view and compare their performance on this measure to a national comparison group at any time and for any reporting period of their choice.

CMS has offered several education and support opportunities. First, the National HQRP Provider Training served as a large-scale venue to introduce and explain the Comprehensive Assessment Measure to providers. This training including education about the component measures, the measure numerator and denominator, exclusion criteria, and overall measure calculation. Providers were also able to ask questions about the measure and its reporting requirements. Second, the factsheet that accompanies the two QM reports, the Hospice Comprehensive Assessment Measure QM Background and Methodology factsheet, and the HQRP Help Desk will help providers interpret their results and have their questions about the Comprehensive Assessment Measure and its component QMs answered. A "Stay on Target" factsheet was posted online in November 2019 and can be accessed here: https://www.cms.gov/files/document/hospice-comprehensive-assessment-measureone-pager.pdf.

Finally, CMS offered a live webinar training in December 2018. The December live webinar training served as a large-scale opportunity to train providers on the Hospice Comprehensive Assessment Measure. The training included education about the background of the measure, the measure numerator and denominator, exclusion criteria, overall measure calculation, and how providers can use their CASPER QM reports to understand their performance on this measure. Providers were given the opportunity to ask questions about the measure during the training.

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

Describe how feedback was obtained.

CMS is committed to receiving feedback on measures implemented as part of the HQRP. CMS takes into consideration feedback and input on measure performance and implementation through the appropriate sub-regulatory communication channels, including but not limited to: NQF public comment periods held as part of endorsement processes, feedback from providers on the Hospice Quality HelpDesk, feedback from provider associations through quarterly provider association calls, and feedback from the provider community on ODFs and HQRP forum. CMS will continue to gather and review feedback on this measure.

CMS issued the FY 2020 Hospice Payment Rate Update Final Rule to include a discussion on the Hospice and Palliative Care Composite Process Measure.

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

CMS has received feedback from providers on the Hospice Comprehensive Assessment Measure following the addition of the measure to providers' CASPER QM reports. Some providers asked for clarification about the specifications, specifically about the "all or none" (rather than average) scoring methodology used and the inclusion of conditional measures in calculation.

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

CMS monitors feedback from other, non-provider, users through the Help Desk, HQRP forum, and through rulemaking.

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

CMS takes all feedback into account when considering future measure refinement. The questions asked to-date are largely related to providers seeking clarification on the specifications of this new measure, and do not point to any issues with the measure specifications. Provider questions about the measure appear to decrease when CMS clarifies the measure specifications through education and outreach activities.

Improvement

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

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

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

Performance results on this measure presented in 1b indicate that hospices have made significant improvements in completing a comprehensive assessment at hospice admission. These results suggest that this measure encourages hospices to conduct all critical care processes for each patient and also sets a higher standard of care for hospices.

4b2. Unintended Consequences

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

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

N/A

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

N/A

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

No

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

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

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):
Are the measure specifications harmonized to the extent possible?

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

N/A

5b. Competing Measures

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

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

Appendix

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

No appendix Attachment:

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): Centers for Medicare & Medicaid Services

Co.2 Point of Contact: Helen, Dollar-Maples, Helen.Dollar-Maples@cms.hhs.gov, 410-786-7214-

Co.3 Measure Developer if different from Measure Steward: Abt Associates

Co.4 Point of Contact: T.J., Christian, Thomas_Christian@abtassoc.com, 617-520-2637-

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

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

CMS convened a 13-member HQRP Composite Measure Technical Expert Panel (TEP) of nationally recognized experts with extensive experience in the following areas: medical or nursing expertise in hospice and palliative care, methods and instrumentation, and quality improvement. Using criteria provided by the NQF, TEP members rated this measure on four criteria: importance, scientific soundness, feasibility and usability. Please see the attached TEP Charter document for TEP members' organizational affiliation and areas of expertise.

Thomas Caprio, MD, MPH, CMD, HMDC, FACP

Cordt Kassner, PhD

Hazel Crews, MHA, MHS, CPHQ

Kama Ferryman, RN, CPHQ

Stacie Hansen, RN, MSN

San Keller, PhD

Ellen Martin, CPHQ, CHPCA

Dana Mukamel, PhD

Sharon-Lise T. Normand, PhD, MSc, FACC, FAHA, Fellow of American Statistical Association

Sally A. Norton, PhD, RN

Joan Piteo, RN, MSN, COS-C

Helena Temkin-Greener, PhD, MPH

Susan Wallace, MSW

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2017

Ad.3 Month and Year of most recent revision: 11, 2020

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

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

Ad.6 Copyright statement:

Ad.7 Disclaimers:

Ad.8 Additional Information/Comments: