NATIONAL QUALITY FORUM
Driving measurable health improvements together

Measure Worksheet

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.
To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 3645
Corresponding Measures:
Measure Title: Hospice Visits in the Last Days of Life
Measure Steward: CMS - DCPAC
sp.02. Brief Description of Measure: The proportion of hospice patients who have received visits from a Registered Nurse or Medical Social Worker (non-telephonically) on at least two out of the final three days of the patient's life.
1b.01. Developer Rationale: This measure addresses a high-priority area by assessing hospice staff visits to patients and caregivers during the final days of life when patients and caregivers typically experience higher symptom and caregiving burdens, and therefore a higher need for care. Collecting information about hospice staff visits for measuring quality of care will encourage hospices to visit patients and caregivers and provide services that will address their care needs and improve quality of life during the patients' last days of life.
This measure focuses on the provision of consistent professional staff visits during the last three days of life during this time of increased symptom burden. The disciplines included in this measure – registered nurse and medical social worker service visits – were selected not only because they are the focus of CMS's payment incentive policy for end-of-life visits (the "Service Intensity Add-On") but also empirical testing found these visits had the highest correlation with CAHPS Hospice scores. Additionally, CAHPS hospice ratings were found to be greatest when visits were provided on two of the last three days of life. Two days would seemingly be preferable to one day for providing more supportive services; as to why it would be preferable to three days, it may be that family prefers at least some privacy during an end-of-life vigil (and not requiring a visit every day of the last three days of life also grants reprieve to hospice providers to still achieve success at this measure).

---

sp.12. Numerator Statement: The numerator of this measure is the number of patient stays in the denominator in which the patient and/or caregiver received visits from registered nurses or medical social workers on at least two of the final three days of the patient's life, as captured by hospice claims records.
sp.14. Denominator Statement: The denominator for the measure includes all hospice patient enrollments in hospice with the patient discharged to death except those meeting exclusion criteria outlined below.
sp.16. Denominator Exclusions: Patient stays are excluded from the measure if the patient (1) received any continuous home care, respite care, or general inpatient care in the final three days of life or (2) if the patient was enrolled in hospice fewer than three calendar days.

Measure Type: Process
sp.28. Data Source: Claims
sp.07. Level of Analysis: Facility

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

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

IF this measure is paired/grouped, NQF#/title:
sp.03. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?:

Preliminary Analysis: New Measure

Criteria 1: Importance to Measure and Report

1a. Evidence

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**

Evidence Summary

- This is a process measure using claims data at the facility level to determines the proportion of hospice patients who have received visits from a registered nurse or medical social worker (non-telephonically) on at least two out of the final three days of the patient's life.
- The logic model depicts clinician visits to patients at the end of life are associated with decreased risk of hospitalization and emergency room visits in the last 2 weeks of the patients' life, decreased likelihood of a hospital-related disenrollment, as well as decreased odds of dying in the hospital.
- The developer provides evidence from a review of the literature from a variety of sources (i.e., academic journals and other publicly available reports), including clinical organizations and panels, as well as individual studies supporting clinician visits at the end of life are associated with improved outcomes for both the patients and their caregivers.
- The developer also notes that several organizations and panels have identified care of the imminently dying patient as an important domain of palliative and hospice care and established guidelines and recommendations related to this high priority aspect of healthcare that affects many people including the National Consensus Project (NCP) Clinical Practice Guidelines for Quality Palliative Care and the American College of Physicians Clinical Practice Guidelines.

Questions for the Committee:

- How strong is the evidence for this relationship?
- Is the evidence directly applicable to the process of care being measured?
- Was the provided evidence comprehensive as an assessment without a systemic review of QQC and grading?

Guidance from the Evidence Algorithm
Not a health outcome (Box 1)→ Graded systematic review was not provided for process measure (Box 3)→ Empirical evidence provided (Box 7)→ Studies summarized (Box 8) → Benefits outweigh undesirable effects (Box 9) → Moderate

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

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

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

- The developer provided data using nationwide Medicare Part A hospice claims representing fee-for-service claims data from 1,561,465 hospice patients receiving service from 3,997 hospice entities.
- 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%.
- The performance scores show a nationwide average of 63.2% (standard deviation is 22.5%; interquartile range 50.2% - 80.5%).

Disparities

- Medicare Enrollment Database beneficiary data was compared using an analysis of variance model:
  - Race/ethnicity: white patients are more likely to achieve success in the measure (68.5%) vs. other groups (black 61.0%, Asian 57.2%, Hispanic 57.5%, Other/Unknown 63.7%)
  - Medicare/Medicaid dual eligibility status (to proxy for socioeconomic status): Medicare-only patients are slightly more likely to achieve success (69.1% vs. 63.2%)
  - Urban/rural: Rural patients are more likely to achieve success (rural 71.0% vs. urban 67.0%)

Questions for the Committee:

- Is there a gap in care that warrants a national performance measure?

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

Committee Pre-evaluation Comments:

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

1a. 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.

- Moderate appropriate, process measure, no evidence for necessity, guideline, will help with social determinant of health comparison
- Process measures - correlations with HCAPS
- The evidence relates to the measure although it is not as robust as desired.
- Evidence applies directly to the desired outcome
- Moderate - No systematic review. Based on input from different articles-experts and recommendations from organizations
- The developer provides research that documents the intensity of the last 2-3 days of life and how the presence of hospice professionals can diminish burden, uphold wishes and prevent unwanted hospitalization.

- The evidence provided and description of why the measure is important mentions the need for an Integrated Care Team and Clinician Visits that would include End of life visits by not only an RN and Social Worker, but also physician and midlevel practitioners involved in the care (NPs/PAs) I am not sure why the measure is only including visits by RNs & SWs rather than other licensed clinicians who are part of the integrated care team.
- The evidence to support this measure I believe falls in the moderate category. The developer did not supply a systematic review specific to the measure but did supply individual studies that provide evidence of an association of clinician visits at the end of life with improved patient and caregiver outcomes, and several panels and organizations noting these visits as a priority. One study noted the visits that addressed palliation and emotional support as important, which provides some evidence for focusing on nurses and social workers. The evidence does apply directly in terms of the focus on the end of life (two weeks, etc.) and the type of clinicians (one study that I can see). I am unaware of any new studies.
- 1. How do we protect against families being "encouraged" to have an in-home visit so that the hospice can do well on this measure even when they really don't want one? 2. While it may very well be that most families will choose a nurse or social worker as their preferred visitor, by limiting the staff who qualify to visit, specifically leaving out chaplains, there is a clear danger that a nurse or social worker will be sent to visit a family and patient whose need and desire is to address spiritual issues with a chaplain. This may be particularly the case with families of color who are more likely to rely on spiritual/religious support at this time. 3. The text specifies that the visits need to be done by "staff who can assess symptoms and make changes to the plans of care as well as work with the patient and the primary caregiver to provide the appropriate palliation and emotional support (nurses, social workers, and physicians) …". While this statement is accurate as far as it goes (1) it again leaves out spiritual support which is so important at EOL and (2) seems to ignore the fact that fully trained professional chaplains have the same capabilities to assess symptoms and recommend changes in care plans as social workers. 4. Should "telephonic" be interpreted to mean only visits by telephone or all virtual visits? If the latter, that restriction does not seem to be consistent with the emerging practice in health care generally and certainly in hospice. Why not virtual visits? Does "telephonic" mean by telephone or all virtual visits?
- The evidence of need is strong, they have provided evidence from literature reviews. Several organizations and panels have identified care of the imminently dying patient as an important domain.

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?

- Yes, gap exists that warrants
- Yes, gap is present
- There is a performance gap that this measure addresses as well as an opportunity to reduce disparities in care
- Performance data was supplied and demonstrates a gap in care in terms of variability in performance that warrants a national performance measure. Some data on the measure by population subgroups was provided and demonstrates disparities in care as evidenced by poorer performance for black, Hispanic and poorer socioeconomic groups
- Data provided. Gap noted (nationwide average of 63.2%). Data about subgroups provided for Race/ethnicity, Medicare/Medicaid, urban/rural) – High

- There appear to be meaningful gaps by race, dual eligibility, and rural/urban location.
- Measure developer provided gaps in care based on race/ethnicity along with disparities in care data
- The developer did supply data from Medicare Part A hospice claims. I was surprised that rural success with regard to the measure was significantly better than urban, but I was not surprised by while beneficiaries doing better than all other beneficiaries of color (data was provided by population subgroups). I think there is moderate evidence of the gap.
- Performance gap and disparities data demonstrating racial disparities.
- There are opportunities for improvement and there are disparities present based on race/ethnicity, SES, and urban/rural locations.

Criteria 2: Scientific Acceptability of Measure Properties

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

Reliability

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

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.

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.

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

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

Evaluators: NQF Staff

NQF Staff Evaluation Summary:

Reliability

- Reliability testing at the Accountable Entity Level:
    - The developer uses 100% Medicare hospice data set of 4,811 Medicare-certified hospices for reliability testing detailing the hospice population with both beneficiary and facility descriptive statistics in percentages.
    - The final testing set included 3,997 hospices with 814 excluded for not meeting the 20-beneficiary reporting minimum. Hospices that were excluded from testing were freestanding (94.8 percent), for-profit (62.6%), and rural (86.2%). For rural hospices, the average number of beneficiary stays was 7.5. Excluded beneficiaries more likely to be Black, male, have a primary diagnosis for dementia/Alzheimer'/Parkinson's, and less likely to have a primary diagnosis for cancer or cerebrovascular accident.
    - Developers used a signal to noise ratio to assess accountable entity reliability.

- Distributive reliability estimates were provided with a mean reliability of 0.973, 4.5 percent of facilities resulted in a reliability of less than 0.9, and no facility had a reliability of less than 0.8.

Validity

- Validity testing at the Accountable Entity Level:
  - The developers calculated (Pearson's) correlation coefficients to identify associations between hospice-level measures between the submission and responses in the CAHPS Hospice Survey quality measure, NQF #2651. Developers hypothesized higher caregiver's experiences with patient's care (i.e., higher #2651 responses) and increased days of hospice service in the last three days of life.
  - The developer selected eight patient reported outcome (PRO) responses from a multi-item PRO-PM (i.e., communication with family, getting timely help, treating patient with respect, emotional and spiritual support, help for pain and symptoms, training family to care for patient, rating of the hospice, and willing to recommend the hospice) as correlates for accountable entity validity testing.
  - The measure positively correlated to eight CAHPS Hospice Survey items, ranging from 0.1739 to 0.2830, which the developers "indicate highly positive correlation between Hospice Visits in the Last Days of Life".

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?

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

Committee Pre-evaluation Comments:

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

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

- Clear denominator and numerator, no concern for implementation
- No concerns
- The measure elements are straight forward and clearly defined and thus likely to be consistently implemented
- No concerns about implementation; no data elements are not clearly defined
- Reliability seems OK. No facility with reliability < 0.8.
- The use of Medicare claims data is clearly defined. I have no concerns about reliability.
- Details of inclusion and exclusion criteria were included.
- I do not have any concerns.
- No concerns
- None

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

- No
- No concerns

- No
- No
- No
- No concerns
- Agree with moderate reliability as noted by review
- I do not have any concerns.
- No concerns
- No

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

- No
- No
- No
- No
- No. The measure positively correlated to eight CAHPS Hospice Survey items, ranging from 0.1739 to 0.2830, which the developers "indicate highly positive correlation between Hospice Visits in the Last Days of Life".
- No concerns
- Agree with high validity as noted by review
- I do not have any concerns.
- No concerns
- 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 concern
- No risk adjustment with this measure.
- NA
- not applicable- this is a process measure
- No concerns
- The exclusion of patients who received continuous or general inpatient care or were enrolled for less than 3 days are appropriate. There was no risk adjustment.
- Exclusion of patients on hospice less than 2-3 days makes sense. All patients included receiving hospice services with exception of exclusion criteria to exclude along with those patients requiring continuous care with Nurse at bedside around the clock.
- I have no concerns.
- No threats to validity noted
- No exclusions are apparent. Yes, everything appears to be present.

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 concern

- I would like the visits in last days analyzed according to data other than HCAPS, such as ED visits, hospital admissions, etc.
- TBD
- Yes
- No concerns
- No
- 2b4-7 Using claims data, 2b4 correlation with CAHPS data, 2b5 comparable results available, 2b6 N/A
- I do not have concerns about threats to validity.
- Missing data is not a threat to validity
- The measure positively correlated to Hospice Survey items, they used PROs from 8 patients for validity testing. There does not appear to be any missing data.

Criterion 3. Feasibility

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.

- The measure is available in an electronic format (i.e., electronic administrative claims billing data)
- The measure is coded by someone other than person obtaining original information.
- All data elements needed to calculate this measure are from Medicare claims records.

Questions for the Committee:

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

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

Committee Pre-evaluation Comments:

Criteria 3: Feasibility

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

- No undue burden, high
- Data is available in claims data
- All the data elements should be routinely generated and available in electronic form, so I don't have concerns about data collection
- Data elements are routinely generated, no concerns about how the data collection strategy can be put into operational uses operational use
- Data available in electronic format. Requires that someone does coding.
- The measure is available electronically, coded by someone other than the person obtaining the information and available in Medicare claims data making its use very feasible. I have no concerns about data collection.
- Data is in electronic form and able to be collected
- I have no concerns.
- Data elements are routinely generated. No concerns

- Everything seems to be in place, no concerns.

Criterion 4: Usability and Use

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

Planned use in an accountability program? ☒ Yes ☐ No

Accountability program details

- This is a newly developed measure by CMS for the Hospice Quality Reporting Program. CMS has announced in rulemaking the measure will be publicly reported no earlier than May 2022.

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 stated the measure was not implemented ye.

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

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

Improvement results

- The developer stated that this measure has not yet been implemented, and thereby improvement results are not available yet.

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

Unexpected findings (positive or negative) during implementation

- The developer did not identify any.

Potential harms

- The developer did not identify any.

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?

- New measure to be publicly reported May 2022
- Not in use yet.
- This is a new measure so has not yet been publicly reported but the plan to do so seems appropriate
- The measure is being used in Medicare Accountability reporting beginning in May of 2022. This is not a maintenance measure. Those being measured have been given performance results and assistance with interpreting the measure results and the data and have been given opportunity to provide feedback on the measure I am not aware of changes to the measure that have incorporated feedback from those being measured
- Those being measured have been given results/data and opportunity to provide feedback. Feedback considered when changes are incorporated. Not clear what is meant by "The developer stated the measure was not implemented yet."
- The measure is not in use yet but CMS plans to publicly report the data.
- CMS has this data. CAHPS publicly reported, this proposed measure not implemented yet
- The measure is not currently being publicly reported but there is a plan to do so in 2022 via the CMS Care Compare website
- Measure not yet implemented
- This measure has not been implemented and does not improvement results.

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.

- Will be useful to identify high quality, efficient, socially equitable care
- Unneeded visits might be provided to meet measure expectations.
- The rationale is credible and there are no harms anticipated with this measure
- The measure has usability in that the process being measure There may be some unintended consequences of the measure, but I do not think they outweigh the benefits. Increased service intensity for patients in the last days of life could divert care from patients who need the care more but are not in the last three days of life. If a process is performed only to achieve a high score rather than to respond to a real patient need, then the process itself loses value and interferes with resource allocation. Visits to a dying patient and family in the last days of life are extremally important and

provide a much-needed level of support for most but not all families. There may be families who would benefit from visits only one out of 3 of the final days of life as an assessment visit with an algorithm to determine if they need a visit the following day or days. Urgent needs must be prioritized and there may not be sufficient staff resources if all patients are receiving visits. There are areas of the country where geographical barriers will may it difficult for the hospice to perform at the desired level for quality reporting and result in undeserved poor ratings has been shown to correlate with improved outcomes including improved satisfaction scores on CHAPS surveys and outcomes such as decreased hospital related disenrollment's from the hospice and decreased ER visits in the last days of life

- "The developer stated that this measure has not yet been implemented, and thereby improvement results are not available yet."
- The only possible unintended consequence is that families may not want visits during the final 2-3 days of life and if quality is being measured by the team's presence this could create conflict of interest.
- Results will be utilized in the CMS Hospice Quality Reporting Program
- The developer argues that collecting this data could encourage hospices to 1) monitor the patient's status for signs of impending death; and 2) ensure two visits are made by nurses and social workers in the last three days. Considering this measure, my concern is the harm that could be done if hospices push visits on patients and caregivers when they do not want them in those final days. In other areas of health disparity research, evidence has been found that communities of color do not always voice what they want and/or feeling. It could be that the visits would not be wanted but patients and caregivers cannot express this. Patients and caregivers could also have concerns that if they say they do not want the visits, that there could be retribution such as not receiving care if they changed their minds. Also, some cultures have preferences as to who is or is not present when someone is dying. I am concerned that more evidence is not available from different communities regarding preferences for the last three days. Finally, having worked as a hospice social worker, I also know that it is not always clear when a patient has a few days left and so visits can be missed. This could definitely have an impact on the data.
- Measure not yet implemented. Developer should be more specific on when the measure will now be implemented
- No harms were identified. This measure has not been implemented and does not improvement results.

Criterion 5: Related and Competing Measures

Related or competing measures

- No related or competing measures identified.

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?

- None
- Hospice Care Index is related, no concerns
- No, this fills an important existing gap
- I have not examined this completely, but I think there could be competing measures;
- No related or competing measures identified.
- None
- No competing measures
- Not that I am aware of.

- No related or competing measures noted
- No

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 01/19/22

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

Staff Scientific Acceptability Evaluation

RELIABILITY: SPECIFICATIONS

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

   Submission document: "Measure 3645 MIF" document, items S.1-S.22

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

   - None

RELIABILITY: TESTING

Submission document: "Measure 3645 MIF" document for specifications, testing attachment questions 1.1-1.4 and section 2a2

3. **Reliability testing level** ☒ **Accountable Entity** ☐ **Patient/Encounter** ☐ **Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure** ☒ **Yes** ☐ **No**
5. If accountable entity level and/or patient/encounter reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of patient-level data conducted?

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

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

   Submission document: Testing attachment, section 2a2.2

   - Developers used a signal to noise ratio. The developer uses 100% Medicare hospice data set of 4,811 Medicare-certified hospices for reliability testing detailing the hospice population with both beneficiary and facility descriptive statistics in percentages.

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

   Submission document: Testing attachment, section 2a2.3

   - Distributive reliability estimates were provided with a mean reliability of 0.973, 4.5 percent of facilities resulted in a reliability of less than 0.9, and no facility had a reliability of less than 0.8.

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.**
    - Specifications are precise, unambiguous and complete (Box 1) >>>Empirical reliability testing done (Box 2) >>> Testing at the accountable entity level conducted (Box 4) >>> The method was appropriate (Box 5) >>> There is moderate certainty that the performance measure scores are reliable (Box 6b).

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.
    - No concerns

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

    Submission document: Testing attachment, section 2b4.
    - The developer used the 100 percent Medicare sample to differentiate performance between hospice facilities. The mean score of hospices with twenty or more beneficiary stays was 63.2 percent with scores ranging from zero to 100 percent. The median was 69.4 percent, the interquartile range was 30.1 percent, and the standard deviation was 22.5 percent. Two hospices had perfect scores. The measure demonstrates a high potential for differentiating between high- and low-quality hospice 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.
    - NA. The measure has only one set of specifications that draw data from a single data source.

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

    Submission document: Testing attachment, section 2b6.
    - The developer conducted testing on 100 percent of Medicare Part A hospice data set assessing reported billed administrative claims codes. The developer states that "claims not submitted (or submitted incorrectly) would result in a denial of claim/payment", meaning they would not be included on the final claim and therefore could not be used to calculate performance.
    - No testing of missing data was conducted and stated it is not needed with a full data set.

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

- Although it is not appropriate to adjust for an access to care process measure (i.e., all patients should have access to high quality care regardless of the patient, community, or provider characteristics), NQF measure evaluation criteria state that access of care measures are stratified to detail population-specific care access gaps in need of remedy.

16c. **Social risk adjustment:**

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

16c.2 Conceptual rationale for social risk factors included? ☐ Yes ☐ No ☒ Not applicable

16c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☐ Yes ☐ No ☒ Not applicable

16d.**Risk adjustment summary:**

16d.1 All of the risk-adjustment variables present at the start of care? ☐ Yes ☒ No ☒ Not applicable

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

16d.3 Is the risk adjustment approach appropriately developed and assessed? ☐ Yes ☐ No ☒ Not applicable

16d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☐ Yes ☐ No ☒ Not applicable

16d.5.Appropriate risk-adjustment strategy included in the measure? ☐ Yes ☐ No ☒ Not applicable

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

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)** ☒ Not applicable

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

- N/A

VALIDITY: TESTING

19. **Validity testing level: ☒ Accountable entity ☐ Patient/Encounter ☐ 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**

Submission document: Testing attachment, section 2b2.2

- The developers calculated (Pearson's) correlation coefficients to identify associations between hospice-level measures between the submission and responses in the CAHPS Hospice Survey quality measure, NQF #2651. Developers hypothesized higher caregiver's experiences with patient's care (i.e., higher #2651 responses) and increased days of hospice service in the last three days of life.
- The developer selected eight patient reported outcome (PRO) responses from a multi-item PRO-PM (i.e., communication with family, getting timely help, treating patient with respect, emotional and

spiritual support, help for pain and symptoms, training family to care for patient, rating of the hospice, and willing to recommend the hospice) as correlates for accountable entity validity testing.

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

    Submission document: Testing attachment, section 2b2.3

    - The measure positively correlated to eight CAHPS Hospice Survey items, ranging from 0.1739 to 0.2830, which the developers "indicate highly positive correlation between Hospice Visits in the Last Days of Life".

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.**
    - The measure positively correlated to eight CAHPS Hospice Survey items, ranging from 0.1739 to 0.2830, which the developers "indicate highly positive correlation between Hospice Visits in the Last Days of Life".

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**

- ☒ **Not applicable** (this is not a composite measure)

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

NA - This is not a composite measure.

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.**

1. Importance to Measure and Report

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

---

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Measure and Report: Evidence section. For example:

2021 Submission:
Updated evidence information here.

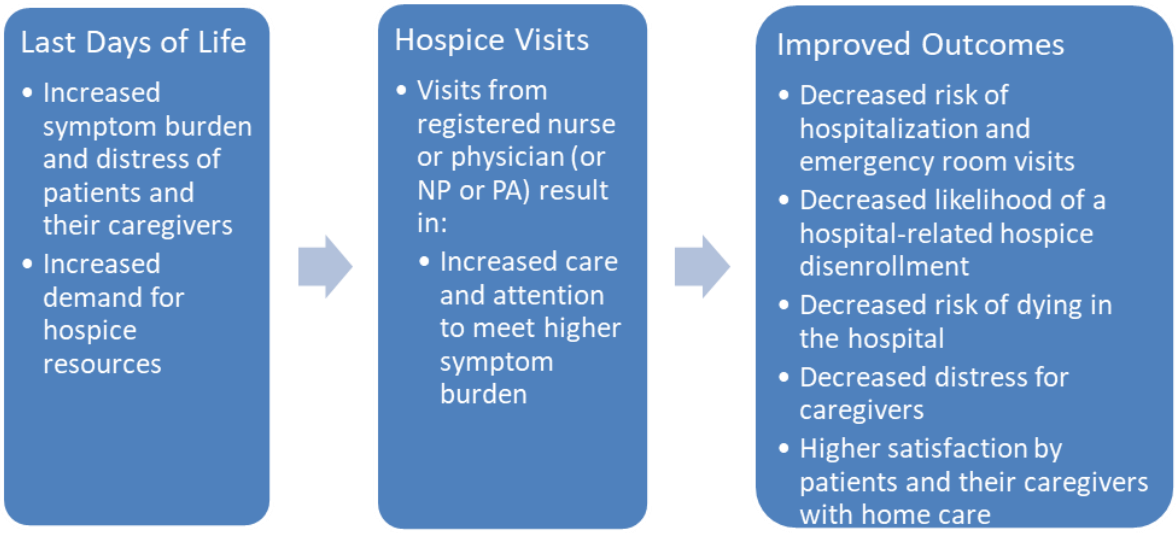
2018 Submission:
Evidence from the previous submission here.

1a.01. Provide a logic model.

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

[Response Begins]
The last week of life is typically the period in the terminal illness trajectory with the highest symptom burden. Particularly during the last few days before death, patients experience many physical and emotional symptoms, necessitating close care and attention from the integrated hospice team and drawing increasingly on hospice team resources (de la Cruz et al., 2015; Dellon et al., 2010; Kehl et al., 2013). Hospice responsiveness during times of patient and caregiver need is an important aspect of care for hospice patients (Ellington et al., 2016). Clinician visits to patients at the end of life are associated with decreased risk of hospitalization and emergency room visits in the last 2 weeks of the patients' life, decreased likelihood of a hospital-related disenrollment, as well as decreased odds of dying in the hospital (Seow et al., 2010; Phongtankuel et al., 2018; Almaawiy et al., 2014).In addition, clinician visits to patients at the end of life is also associated with decreased distress for caregivers and higher satisfaction with home care (Pivodic et al., 2016).

Figure 1. Relation Between Hospice Visits in the Last Days of Life and Patient Outcome



[Response Ends]

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

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

[Response Begins]

Other (specify)

The evidence for this measure was obtained through a review of the literature (academic journals and other publicly available reports).

[Response Ends]

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

Evidence - Systematic Reviews Table (Repeatable)

Group 1 - Evidence - Systematic Reviews Table

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

[Response Begins]

N/A

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

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

[Response Begins]
N/A
[Response Ends]

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

[Response Begins]
N/A
[Response Ends]

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

[Response Begins]
N/A
[Response Ends]

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

[Response Begins]
N/A
[Response Ends]

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

[Response Begins]
N/A
[Response Ends]

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

[Response Begins]
N/A
[Response Ends]

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

[Response Begins]
N/A
[Response Ends]

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

[Response Begins]
N/A
[Response Ends]

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

[Response Begins]

The evidence for this measure was obtained through a review of the literature (academic journals and other publicly available reports).
[Response Ends]

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

[Response Begins]
Hospice Visits in the Last Days of Life is a measure of the proportion of hospice patients who have received visits from a Registered Nurse or Medical Social Worker (non-telephonically) on at least two out of the final three days of the patient's life. There is evidence available from clinical organizations and panels, as well as from individual studies, supporting the measure's basis that clinician visits to patients at the end of life are associated with improved outcomes for both the patients and their caregivers.

The last week of life is typically the period in the terminal illness trajectory with the highest symptom burden. Particularly during the last few days before death, patients experience many physical and emotional symptoms, necessitating close care and attention from the integrated hospice team and drawing increasingly on hospice team resources (de la Cruz et al., 2015; Dellon et al., 2010; Kehl et al., 2013). Hospice responsiveness during times of patient and caregiver need is an important aspect of care for hospice patients (Ellington et al., 2016). Although Medicare-certified hospices do not have any mandated minimum number of required visits for patients in routine home care (RHC), the most common level of hospice care, at the end of life, hospices should be equipped to meet the higher symptom and caregiving burdens of patients and their caregivers during this critical period (Teno et al., 2016). Clinician visits to patients at the end of life are associated with decreased risk of hospitalization and emergency room visits in the last 2 weeks of the patients' life, decreased likelihood of a hospital-related disenrollment, as well as decreased odds of dying in the hospital (Seow et al., 2010; Phongtankuel et al., 2018; Almaawiy et al., 2014). In addition, clinician visits to patients at the end of life are associated with decreased distress for caregivers and higher satisfaction with home care (Pivodic et al., 2016). Visits by staff who can assess symptoms and make changes to the plans of care as well as work with the patient and the primary caregiver to provide the appropriate palliation and emotional support (nurses, social workers, and physicians) are important to the quality-of-care hospices deliver, as noted by the NQF's preferred practices on the recognition and management of the actively dying patient (Teno et al., 2016). During the development of the Family Evaluation of Hospice Care survey, families voiced the importance of visits by these staff in the last days of life (Teno et al., 2016).

Several organizations and panels have identified care of the imminently dying patient as an important domain of palliative and hospice care and established guidelines and recommendations related to this high priority aspect of healthcare that affects many people. The NQF 2006 report "A Framework and Preferred Practices for Palliative and Hospice Care Quality" recommends that signs and symptoms of impending death are recognized, communicated, and educated, and care appropriate for the phase of illness is provided (National Quality Forum 2006). The National Consensus Project (NCP) Clinical Practice Guidelines for Quality Palliative Care also acknowledges that "care of the patient at the end of life is time- and detail intensive, requiring expert clinical, psychological, social, and spiritual attention to the process as it evolves", and recommends that hospices continually evaluate symptom management issues in anticipation of higher levels of staff support at this time (2018). Further, the American College of Physicians Clinical Practice Guidelines recommend that clinicians regularly assess pain, dyspnea, and depression for patients with serious illness at the end of life (Qaseem et al., 2008). Highly specific physical signs associated with death were identified within 3 days of death (Hui et al., 2014)

Further, there is evidence of disparity in visits among patients at their end of life. Greater increase in visits was associated with patients who were younger, male, Caucasian, had a spouse caregiver, and shorter lengths of stay (Ellington 2016, Teno 2016). Thus, measuring these visits may also reduce disparities in care.

Finally, two conferences presentations found that visits by a nurse or social worker in at least two of the last three days of life was positively correlated with improved CAHPS Hospice outcomes (Christian et al., 2020).

[Response Ends]

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

[Response Begins]

We identified evidence from literature searches using PubMed, and in reviews of references cited in the relevant identified studies.
[Response Ends]

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

[Response Begins]
Almaawiy, U., et al. (2014). Are family physician visits and continuity of care associated with acute care use at end-of-life? A population-based cohort study of homecare cancer patients. *Palliative Medicine*, 28(2), 176−183. doi: 10.1177/0269216313493125

Christian et al. (2020). Trends in the Receipt of Consistent Hospice Professional Visits at the End of Life and Ratings of Hospice Care Quality. Podium presentations at the AcacdemyHealth (June) and Gerontological Society of America (November) annual research conferences.

de la Cruz, M., et al. (2015). Delirium, agitation, and symptom distress within the final seven days of life among cancer patients receiving hospice care. *Palliative & Supportive Care*, 13(2): 211-216. doi: 10.1017/S1478951513001144

Dellon, E. P., et al. (2010). Family caregiver perspectives on symptoms and treatments for patients dying from complications of cystic fibrosis. *Journal of Pain & Symptom Management*, 40(6): 829-837. doi: 10.1016/j.jpainsymman.2010.03.024

Ellington, L., et al. (2016). Interdisciplinary team care and hospice team provider visit patterns during the last week of life. *Journal of Palliative Medicine*, 19(5), 482-487. doi: 10.1089/jpm.2015.0198

Hui D et al. (2014). Clinical signs of impending death in cancer patients. *The Oncologist*, 19(6):681-687. doi:10.1634/theoncologist.2013-0457.

Kehl, K. A., et al. (2013). A systematic review of the prevalence of signs of impending death and symptoms in the last 2 weeks of life. *American Journal of Hospice & Palliative Care*, 30(6): 601-616. doi: 10.1177/1049909112468222

National Consensus Project for Quality Palliative Care. (2018). *Clinical Practice Guidelines for Quality Palliative Care, 4th edition*. Richmond, VA: National Coalition for Hospice and Palliative Care. https://www.nationalcoalitionhpc.org/ncp

National Quality Forum. (2006). *A National Framework and Preferred Practices for Palliative and Hospice Care Quality*. http://www.qualityforum.org/Publications/2006/12/A_National_Framework_and_Preferred_Practices_for_Palliative_and_Hospice_Care_Quality.aspx

Phongtankuel, V., et al. (2018). Association between nursing visits and hospital-related disenrollment in the home hospice population. *American Journal of Hospice & Palliative Medicine*, 35(2): 316-323. doi: 10.1177/1049909117697933

Pivodic, L., Harding, R., Calanzani, N., McCrone, P., Hall, S., Deliens, L., & Gomes, B. (2016). Home care by general practitioners for cancer patients in the last 3 months of life: An epidemiological study of quality and associated factors. *Palliative Medicine*, 30(1), 64−74. doi:10.1177/0269216315589213

Qaseem, A., et al. (2008). Evidence based interventions to improve the palliative care of pain, dyspnea, and depression at the end of life: A clinical practice guideline from the American College of Physicians. *Annals of Internal Medicine*, 148(2): 141-146.

Seow, H., Barbera, L., Howell, D., & Dy, S. M. (2010). Using more end-of-life homecare services is associated with using fewer acute care services: A population-based cohort study. *Medical Care*, 48(2): 118−124. doi: 10.1097/MLR.0b013e3181c162ef

Teno, J. M., et al. (2016). Examining variation in hospice visits by professional staff in the last 2 days of life. *JAMA Internal Medicine*, 176(3): 364-370. doi: 10.1001/jamainternmed.2015.7479

[Response Ends]

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

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

[Response Begins]
This measure addresses a high-priority area by assessing hospice staff visits to patients and caregivers during the final days of life when patients and caregivers typically experience higher symptom and caregiving burdens, and therefore a higher need for care. Collecting information about hospice staff visits for measuring quality of care will encourage hospices to visit patients and caregivers and provide services that will address their care needs and improve quality of life during the patients' last days of life.

This measure focuses on the provision of consistent professional staff visits during the last three days of life during this time of increased symptom burden. The disciplines included in this measure – registered nurse and medical social worker service visits – were selected not only because they are the focus of CMS's payment incentive policy for end-of-life visits (the "Service Intensity Add-On") but also empirical testing found these visits had the highest correlation with CAHPS Hospice scores. Additionally, CAHPS hospice ratings were found to be greatest when visits were provided on two of the last three days of life. Two days would seemingly be preferable to one day for providing more supportive services; as to why it would be preferable to three days, it may be that family prefers at least some privacy during an end-of-life vigil (and not requiring a visit every day of the last three days of life also grants reprieve to hospice providers to still achieve success at this measure).
[Response Ends]

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

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

[Response Begins]
We undertook analysis at the hospice level using nationwide Medicare Part A hospice claims with dates of death ending in Federal Fiscal Years 2018 through 2019. The performance scores below represent fee-for-service claims data from 1,561,465 hospice patients receiving service from 3,997 hospice entities (with measure denominators of 20+, the steward's minimum standard for public reporting). Performance scores are listed below, calculated at the hospice-level. The nationwide average is 63.2% (standard deviation is 22.5%; interquartile range 50.2% - 80.5%). Score in the 1st decile averaged 15.5% and in the 10th decile averaged 90.7%.

Distribution Statistic	Estimate
Hospice Count	3,997
Mean	63.2%
Standard Deviation	22.5%
1st Percentile	3.8%
10th Percentile	28.0%
25th Percentile	50.2%
50th Percentile (Median)	69.6%
75th Percentile	80.5%
90th Percentile	87.2%
99th Percentile	94.4%
Minimum	0.0%
Maximum	100.0%
Interquartile Range	50.2% - 80.5%
1st Decile Mean Score	15.5%
2nd Decile	36.5%
3rd Decile	50.1%
4th Decile Mean Score	59.7%
5th Decile Mean Score	66.5%
6th Decile Mean Score	71.9%
7th Decile Mean Score	76.3%
8th Decile Mean Score	80.4%

Distribution Statistic	Estimate
9th Decile Mean Score	84.8%
10th Decile Mean Score	90.7%

[Response Ends]

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

[Response Begins]
N/A, performance data reported above.
[Response Ends]

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

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

[Response Begins]
We calculated disparity statistics at the *beneficiary-level*. Beneficiary disparity characteristics were obtained from the Medicare Enrollment Database. We compared measure scores by beneficiaries' race/ethnicity, Medicare/Medicaid dual eligibility status (to proxy for socioeconomic status), and urban/rural location (whether the beneficiary resided in a U.S. Census-defined metropolitan area or not). Using an Analysis of Variance model, we confirmed the statistical dissimilarity of scores among the groups (p-value<0.001 for all three groups). There are notable disparities among race and ethnic groups: white patients are more likely to achieve success in the measure (68.5%) vs. other groups (black 61.0%, Asian 57.2%, Hispanic 57.5%, Other/Unknown 63.7%) and by dual status (Medicare-only 69.1% vs. 63.2%). Prior to the analysis, there had been concerns among urban/rural disparities, if rural hospices would have more difficulty accessing patients who would thereby have poorer rates of success for the measure; however, rural patients actually have higher rates of success for this measure (rural 71.0% vs. urban 67.0%), so any access issues do not seem to affect this measure's scores
.

Beneficiary Characteristics	Estimate
(Race)	*
White	68.5%
Black	61.0%
Asian	57.2%
Hispanic	57.5%
Other/Unknown	63.7%
(Medicare/Medicaid Dual Eligibility)	*
Medicare-Only	69.1%
Medicare/Medicaid Dual Eligibility	63.2%
(Urban/Rural Status)	*
Rural	71.0%
Urban	67.0%

Cells left intentionally left blank

[Response Ends]

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

[Response Begins]
N/A, performance data reported above.
[Response Ends]

2. Scientific Acceptability of Measure Properties

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

---

sp.01. Provide the measure title.

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

[Response Begins]
Hospice Visits in the Last Days of Life
[Response Ends]

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

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

[Response Begins]
The proportion of hospice patients who have received visits from a Registered Nurse or Medical Social Worker (non-telephonically) on at least two out of the final three days of the patient's life.
[Response Ends]

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

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Surgery: General*

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

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

[Response Begins]
Care Coordination
[Response Ends]

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

Select only those target populations which can be stratified in the reporting of the measure's result.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Populations at Risk: Populations at Risk*

[Response Begins]
Elderly (Age >= 65)
Populations at Risk: Dual eligible beneficiaries of Medicare and Medicaid
[Response Ends]

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

Check ONLY the levels of analysis for which the measure is SPECIFIED and TESTED.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

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

[Response Begins]
Facility
[Response Ends]

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

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

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

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

[Response Begins]
None available
[Response Ends]

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

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

sp.12. State the numerator.

Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome).

DO NOT include the rationale for the measure.

[Response Begins]
The numerator of this measure is the number of patient stays in the denominator in which the patient and/or caregiver received visits from registered nurses or medical social workers on at least two of the final three days of the patient's life, as captured by hospice claims records.
[Response Ends]

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

All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets.

Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]
Registered nurse visits are identified by revenue code 055x (with the presence of HCPCS code G0299); Non-telephone visits by medical social workers are identified by revenue code 056x (other than 0569; HCPCP code G0155).
[Response Ends]

sp.14. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]
The denominator for the measure includes all hospice patient enrollments in hospice with the patient discharged to death except those meeting exclusion criteria outlined below.
[Response Ends]

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

All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets.

Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]
The denominator for the measure includes all hospice patient stays where the patient expired in hospice except for those with exclusions as identified below. Patients that expired in hospice care are indicated by reason for discharge code on the claim (PTNT_DSCHRG_STUS_CD equals [40, 41, or 42]]). Hospice patient dates of death must occur during the target period (in the development data, pooled Federal Fiscal Years 2018-2019).
[Response Ends]

sp.16. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]

Patient stays are excluded from the measure if the patient (1) received any continuous home care, respite care, or general inpatient care in the final three days of life or (2) if the patient was enrolled in hospice fewer than three calendar days.
[Response Ends]

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

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

[Response Begins]
The exclusion criteria are:

1. Patient received any continuous home care, respite care or general inpatient care in the final three days of life (exclude if revenue codes = [0652, 0655, or 0656])
2. Patient was enrolled in hospice one or two calendar days, only.

The rationale for these exclusions is provided below (in section 2b.16).

[Response Ends]

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

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

[Response Begins]
N/A; no stratification for this process measure.
[Response Ends]

sp.19. Select the risk adjustment type.

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

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

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

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

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

[Response Ends]

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

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

[Response Begins]

1. The data are all Medicare hospice fee-for-service claims within the relevant time period; this measure is calculated over two pooled years (the measure development time period was Federal Fiscal Years 2018-2019; 10/1/17 – 9/30/19)
2. Identify all Medicare hospice decedents discharged to death within the time period of data as identified by the claims discharge stats code, PTNT_DSCHRG_STUS_CD equals [40, 41, or 42].
3. The exclusion criteria are that the:
    1. Patient received any continuous home care, respite care or general inpatient care in the final three days of life (exclude if revenue codes = [0652, 0655, or 0656])
    2. Patient was enrolled in hospice one or two days, only
4. Cases meeting the target process are identified as the number of patient stays in the denominator for which the patient and/or caregiver received visits from registered nurses or medical social workers on at least two days of the final three days of life
    1. Registered nurse visits are identified by revenue code 055x (with the presence of HCPCS code G0299)
    2. Non-telephone visits by medical social workers are identified by revenue code 056x (other than 0569; HCPCP code G0155)
5. The rates of patients meeting the target process are calculated for each hospice provider with at least 20 patients in the denominator during the time period of date
    1. For each hospice, divide the total number of patients in the numerator (Step 4) by the total number of patients in the denominator (Steps 2 and 3) and multiply by 100
    2. The measure is not calculated for hospices with fewer than 20 patients in the denominator
6. For this process measure there are no risk adjustments to measure scores

[Response Ends]

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

[Response Begins]
Data were from nationwide Medicare hospice fee-for-service claims records, Federal Fiscal Years 2018-2019.
[Response Ends]

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

[Response Begins]
Claims
[Response Ends]

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

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

[Response Begins]

Data are obtained from Medicare Part A Hospice Fee-For-Service Claims with dates of discharge ending in Federal Fiscal Years 2018-2019; access was through the CMS Research Data Assistance Center (ResDAC) Chronic Conditions Warehouse (CCW).
[Response Ends]

sp.30. Provide the data collection instrument.

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

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

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

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.
2a. Reliability testing demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.
2b1. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.
2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure;
AND
If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).
2b3. For outcome measures and other measures when indicated (e.g., resource use):

- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; 14,15 and has demonstrated adequate discrimination and calibrationOR
- rationale/data support no risk adjustment/ stratification.

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful 16 differences in performance;

OR
there is evidence of overall less-than-optimal performance.
2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.
2b6. Analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.
2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:
2c1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and
2c2. the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.
(if not conducted or results not adequate, justification must be submitted and accepted)

Definitions
Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).
Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.
Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.
Risk factors that influence outcomes should not be specified as exclusions.
With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v.$5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Scientific Acceptability sections. For example:

2021 Submission:
Updated testing information here.

2018 Submission:
Testing from the previous submission here.

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

[Response Begins]
Claims
[Response Ends]

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

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

[Response Begins]
The dataset used for testing is the 100% Medicare Part A hospice fee-for-service claims with dates of discharge in Federal Fiscal Years (FY) 2018-2019 (October 1, 2017 through September 30, 2019).
[Response Ends]

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

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

[Response Begins]
10-01-2017 - 09-30-2019
[Response Ends]

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

Testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

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

[Response Begins]
Facility
[Response Ends]

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

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

[Response Begins]
Data for these analyses corresponds to 4,811 Medicare-certified hospices providing service during Federal Fiscal Years 2018-2019. In the table below, we also separate statistics by whether the hospice did or did not meet CMS public reporting criteria (of having 20+ eligible cases in the denominator during the reporting period). These hospices were identified using 100% Medicare Part A (hospice) claims during this time period. These providers correspond to the nationwide set of hospices; no sampling was used. Characteristics of these providers are as follows:

Facility Information	All hospices (4,811)	Hospices Meetings Public Reporting Criteria, Only (n=3,997)	Excluded Hospices Not Meeting Public Reporting Criteria (n=814)
Facility Type			
Facility-Based	15.5%	17.6%	5.2%
Freestanding	84.5%	82.4%	94.8%
Profit Status			
Non-profit	24.2%	5.4%	25.9%

Facility Information	All hospices (4,811)	Hospices Meetings Public Reporting Criteria, Only (n=3,997)	Excluded Hospices Not Meeting Public Reporting Criteria (n=814)
For-profit	62.6%	90.9%	62.6%
Government-owned	13.2%	3.7%	11.5%
Census Regions			
Northeast	10.7%	2.8%	9.7%
Midwest	23.1%	6.3%	16.8%
South	38.5%	27.8%	37.3%
West	26.7%	63.0%	36.2%
Outlying territories	1.2%	0.1%	0.1%
Urban/Rural			
Urban	79.3%	6.5%	13.7%
Rural	20.7%	93.5%	86.2%
Hospice size:	Cells intentionally left blank		

The number of eligible stays per hospice ranged from 0 (among 53 of 4,811 hospices, or 1.1%) to 8,118. The mean hospice size was 326 eligible stays, and the median size was 153 eligible stays. Among those 3,997 hospices included for public reporting, the average number of stays was 391; among the 814 hospices not meeting the public reporting criteria, the average number of eligible stays was just 7.5.

[Response Ends]

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

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

[Response Begins]
Analyses were based upon those hospice elections ending in the patient's death during the target period (Federal Fiscal Years 2018-2019). A total of 2,265,876 elections were identified. Among these, discharges were excluded from the denominator if the patient received any general inpatient care, continuous home care, or respite care during the final three days of life; or, if the patient elected hospice fewer than three days (more detail on exclusion criteria is provided below).

Patient characteristics are as follows (for all identified beneficiaries, those without any exclusions, and those excluded for at least one reason):

*	All Beneficiaries (n=2,265,876)	Beneficiaries without Exclusions (n=1,567,536)	Excluded Beneficiaries (n=698,340)
Race/Ethnicity			
White	86.9%	87.5%	85.7%
Black	8.0%	7.6%	9.0%
Asian	1.2%	1.3%	1.2%
Hispanic	1.9%	1.7%	2.2%
Other/Unknown	2.0%	2.0%	2.0%
Gender			
Male	42.4%	47.7%	48.2%
Female	57.6%	52.3%	51.8%
Primary Diagnosis			
Cancer	28.9%	28.4%	15.1%
Dementia/Alzheimer'/Parkinson's	23.1%	10.8%	26.8%
Cerebrovascular Accident	8.9%	10.9%	6.8%

*	All Beneficiaries (n=2,265,876)	Beneficiaries without Exclusions (n=1,567,536)	Excluded Beneficiaries (n=698,340)
Heart Disease (excluding Congestive Heart Failure)	9.5%	8.3%	9.1%
Congestive Heart Failure	8.6%	7.9%	8.4%
Chronic Obstructive Pulmonary Disease	6.4%	6.3%	6.4%
Pneumonia & Infections	2.5%	7.9%	4.7%
Chronic Kidney Disease	2.2%	3.6%	2.7%
Chronic Liver Disease	1.3%	1.8%	1.8%
None of the Above	8.7%	12.4%	9.8%

Cells intentionally left blank

Mean Age: 82.6 years (at the time of death) for all patients (83.4 included patients, 80.9 excluded patients)

[Response Ends]

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

[Response Begins]
Whereas the characteristic estimates above reflect both included and excluded denominator cases, measure scores - and reliability and validity performance testing for those scores - are limited to only those hospice elections which are included in the denominator. Additionally, reliability and validity testing only include those hospices which meet the minimum CMS public reporting threshold of at least 20 denominator cases (to generate statistically meaningful results).

Analyses of excluded cases focus on those discharges that are excluded from the denominator.

No risk adjustment was conducted for this process measure.
[Response Ends]

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

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

[Response Begins]
We performed socioeconomic disparity analyses at the patient-level. Medicaid status was used as a proxy measure of low socioeconomic status as patient-level income was not available. We examined the rate of receiving a visit by registered nurse or social worker in at least two of the patient's final three days of life between patients with full Medicaid dual eligibility and those without. We also examined differences in rates by patients' race & ethnicity and rural/urban residential locations. We compared rates calculated among each of these groups to determine whether performance scores varied by subgroup.
[Response Ends]

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

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

Choose one or both levels.
[Response Begins]

Accountable Entity Level (e.g., signal-to-noise analysis)
[Response Ends]

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

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

[Response Begins]
We followed the methodological approach to reliability testing outlined in "The Reliability of Provider Profiling: A Tutorial" by John Adams of the RAND Corporation (2009) which was featured in the NQF document "What Good Looks Like" for measure submission examples. The actual approach entails using a hierarchical model to obtain an estimate of provider-to-provider variance, and then applying that estimate to the reliability formula (along with estimates of individual provider error).

This approach calculates what's known as the "signal-to-noise ratio"; which as Adams (2009) writes, presents "...the proportion of variability in measured performance that can be explained by real differences in performance." I.e., it indicates the extent it can confidentially be ascertained the measure distinguishes performance of one facility.

Reliability scores - as expressed as signal-to-noise ratios - vary from 0 to 1, with 0 indicating all variation is attributable to noise/measurement and 1 indicating that all variation owes to real differences in performance across facilities.
[Response Ends]

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

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

[Response Begins]
Our approach produces facility-level reliability scores; with 3,997 facilities meeting CMS reporting thresholds (denominator of at least 20), we present summary statistics below.

In general, the mean average reliability score was 0.973 and the median score was 0.986. Just 4.5% of facilities had signal-to-noise ratio values below 0.9, and no facility had a ratio below 0.8.

Reliability Statistic	Estimate
Facility Count	3,997
Mean	0.974
Standard Deviation	0.031
1st Percentile	0.854
10th Percentile	0.931
25th Percentile	0.967
50th Percentile (Median)	0.986
75th Percentile	0.994
90th Percentile	0.997
99th Percentile	0.999
Minimum	0.823
Maximum	1.000

[Response Ends]

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

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

[Response Begins]
Although there is not a definite threshold that is considered reliable, per Adams (2009), values above 0.7 are sufficient to confidentially detect differences between facilities. The reliability scores calculated here are almost all above 0.9 (and all completely are above 0.8), so reliability can be considered very good.

Clinic-specific reliability is consistently greater than 0.9, and thus can be considered to be very good.

There is not a clear cut-off for minimum reliability level. Values above 0.7, however, are considered sufficient to see differences between some physicians (or clinics) and the mean, and values above 0.9 are considered sufficient to see differences between pairs of physicians (in this case clinics) (see RAND tutorial, 2009).
[Response Ends]

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

[Response Begins]
Empirical validity testing
[Response Ends]

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

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

[Response Begins]
We calculated (Pearson's) correlation coefficients for the association between hospice-level measure scores and responses from the CAHPS Hospice Survey. CAHPS scores are an existing endorsed QM (NQF #2651) and best capture the caregiver's experiences of a patient's care. We are looking for concordance between two: higher CAHPS scores indicate better quality hospices, so if we find facilities with higher rates of Hospice Visits in the Last Days of Life also having higher CAHPS scores, that would validate our process measure towards capturing an important measure of care.
[Response Ends]

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

Examples may include correlations or t-test results.

[Response Begins]
Correlation estimates are presented below. It should be noted CAHPS Hospice data were only available for 2,895 facilities (due to CAHPS suppression) of which we had reportable measure scores. Coefficient estimates in the table below were all (positively) statistically significant with p-values<0.001.

CAHPS® Quality measure	**Correlation coefficient**
Communication with family	0.2635
Getting timely help	0.2489
Treating patient with respect	0.2830
Emotional and spiritual support	0.2842
Help for pain and symptoms	0.1997
Training family to care for patient	0.1739
Rating of this hospice	0.2792
Willing to recommend this hospice	0.2608

[Response Ends]

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

[Response Begins]
These results indicate highly positive correlation between Hospice Visits in the Last Days of Life. Although there is not a minimum threshold, CMS had previously developed another measure using the Hospice Item Set data to measure end-of-life visits (with a slightly different specification), and the correlations of this new, claims-based specification with CAHPS hospice exceeds that of the measure previously developed.

Generally speaking, these findings (of positive, statistically significant correlation) establish consistent rankings of facilities in terms of CAHPS Hospice scores and measure performance rates. It indicates the hospices which most often provide visits at the end of life by registered nurses and social workers also have the highest ratings of patience experience as reported by caregivers. This validates our process measure, in that it suggests that high-quality hospices are most often performing this important care process.
[Response Ends]

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

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

[Response Begins]
We analyze the distribution of the hospice-level QM scores to assess the measure's variability, to examine how well the measure distinguished between high- and low-quality hospices. We assess the QM's variability by focusing on the percentage of hospices with perfect scores (i.e., a clinical visit provided in the final 3 days for 100 percent of patients), and empirical/distribution of the QM scores.
[Response Ends]

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

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

[Response Begins]
The mean score (keeping with a 20-stay minimum denominator size), was 63.2% with scores ranging from 0% to 100%. The median was 69.4%, the interquartile range was 30.1%, and the standard deviation was 22.5%. Just two hospices had perfect scores.

The full distribution results using Federal Fiscal Years 2018-2019 are displayed below:

Distribution Statistic	Estimate
Hospice Count	3,997
Mean	63.2%
Standard Deviation	22.5%
1st Percentile	3.8%
10th Percentile	28.0%
25th Percentile	50.2%
50th Percentile (Median)	69.6%
75th Percentile	80.5%
90th Percentile	87.2%

Distribution Statistic	Estimate
99th Percentile	94.4%
Minimum	0.0%
Maximum	100.0%

[Response Ends]

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

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

[Response Begins]
Mean scores nationwide are 63.2%, and thereby are far from being "topped-out"; only two hospices out of 3,997 had perfect scores. Moreover, the interquartile range of 30.1% represents a substantial range, indicating this measure has high potential for ability to differentiate hospices between high- and low-quality providers.
[Response Ends]

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

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

[Response Begins]
There are no missing data in the data set for the data elements used to calculate this measure. The calculation for the measure is based on claims for payment submitted by Medicare-certified hospices. Claims not submitted (or submitted incorrectly) would result in a denial of claim/payment.
[Response Ends]

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

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

[Response Begins]
N/A - as noted Medicare claims data are fully complete and as such there are no missing data in the data set for the data elements used to calculate this measure.
[Response Ends]

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

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

[Response Begins]

N/A - as noted Medicare claims data are fully complete and as such there are no missing data in the data set for the data elements used to calculate this measure.
[Response Ends]

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

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

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

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

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

[Response Begins]
[Response Ends]

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

Examples may include correlation, and/or rank order.

[Response Begins]
[Response Ends]

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

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

[Response Begins]
[Response Ends]

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

[Response Begins]
Yes, the measure uses exclusions.
[Response Ends]

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

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

[Response Begins]
The specifications for Hospice Visits in the Last Days of Life include two exclusions:

1. Beneficiaries with enrollments less than three days
2. Beneficiaries utilizing services other than routine home care (general inpatient, continuous home care, or respite care)

To test the impact of these exclusions, we compared measure performance for individuals included and excluded under the specifications (for each respective exclusion individually) and assessed the differences in rates.

Of note, the rationale for these particular exclusions included some practical considerations which impacts the approach for this analysis. For example: Hospice Visits in the Last Days of Life identifies patients receiving visits in at least two of the last three days of life. It would be impossible/non-sensical to assess among patients enrolled in hospice one day, how many received visits on two hospice election days. For this exclusion analysis, for the enrollment length exclusion, we focus therefore on those individuals enrolled in hospice two days (and calculate measure success as those receiving visits on two days).

An important rationale for the exclusions for non-routine home care service was to align this measure with CMS's Service Intensity Add-On payment policy, which incentivizes visits by a registered nurse or social worker on routine home care service days (only) at the end of life. For calculation purposes, visits are not recorded in instances of patients receiving general inpatient care in a hospice inpatient unit. For this analysis, because the patient was in a facility, we assume the visit took place in this circumstance. However, this exemplifies another rationale for excluding (non-routine home care) higher levels of hospice care: because the patient is in a facility or requiring additional visits to justify the service, by definition the visit is being received by definition of expectations of the service type, so it is not interesting or helpful for a quality measurement perspective.

[Response Ends]

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

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

[Response Begins]
Overall, 264,027 (13.2%) individuals were excluded for having hospice enrollments fewer than three days and 591,166 (26.1%) of individuals were excluded for having at least one day of general inpatient, continuous home care, or inpatient respite hospice service during the last three days of life.

- Among individuals eligible for the denominator as specified, the rate of receiving visits in two of the last three days of life was 67.4%
- Among individuals enrolled in hospice fewer than three days (in this analysis, two days only, as mentioned above), the rate of receiving visits in two of the last three days of life was 72.8%
- Among individuals with at least one day of general inpatient, continuous home care, or inpatient respite hospice service during the last three days of life, the rate of receiving visits in two of the last three days of life was 81.5%

[Response Ends]

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

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

[Response Begins]

There were some differences in rates among excluded individuals, but as noted the rationale for exclusions were more about correct interpretation of the measure. If the measure as specified checks for visits in at least two of the last three days of life, it is only fair to hospices to only include those individuals with at least three days enrollment. Additionally, because patients on higher levels of hospice service (general inpatient, continuous home care, or inpatient respite) receive more intensive care by definition, it is not useful to include such patients in the measure. As such, we believe both exclusions are appropriate.
[Response Ends]

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

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

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

[Response Begins]
N/A; no risk adjustment or stratification for this process measure.
[Response Ends]

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

[Response Begins]
Hospice Visits in the Last Days of Life is a process measure and is not risk-adjusted.
[Response Ends]

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

[Response Begins]
Other (specify)
N/A; this process measure is not risk-adjusted.
[Response Ends]

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

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

[Response Begins]
N/A; this process measure is not risk-adjusted or stratified.
[Response Ends]

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

[Response Begins]
N/A; this process measure is not risk-adjusted.
[Response Ends]

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

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

[Response Begins]
N/A; this process measure is not risk-adjusted.
[Response Ends]

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

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

[Response Begins]
N/A; this process measure is not risk-adjusted.
[Response Ends]

2b.27. Provide risk model discrimination statistics.

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

[Response Begins]
N/A; this process measure is not risk-adjusted.
[Response Ends]

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

[Response Begins]
N/A; this process measure is not risk-adjusted.
[Response Ends]

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

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

[Response Begins]
N/A; this process measure is not risk-adjusted.
[Response Ends]

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

[Response Begins]

N/A; this process measure is not risk-adjusted.
[Response Ends]

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

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

[Response Begins]
N/A; this process measure is not risk-adjusted.
[Response Ends]

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

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

[Response Begins]
N/A (no risk adjustment)
[Response Ends]

3. Feasibility

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

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

[Response Begins]
Coded by someone other than person obtaining original information (e.g., DRG, ICD-10 codes on claims)
[Response Ends]

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

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

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

[Response Begins]
All data elements needed to calculate this measure are from Medicare claims records.
[Response Ends]

3.04. Describe any efforts to develop an eCQM.

[Response Begins]
In that all needed to calculate this measure were already available from Medicare claims records, there were no further efforts taken to develop an eCQM.
[Response Ends]

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

[Response Begins]
Claims records are submitted by hospices in order to receive payment for hospice services. No further data collection was needed beyond information already being collected, so this measure did not impose any further burden to patients, caregivers, or providers, and there was no additional data collection costs. Because hospices must submit to receive payment, missing data is not an issue (as noted above). Typically providers submit claims monthly (but not always); in publicly reporting measure scores the steward will allow at least several months' lag to allow claims records to be submitted.

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

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

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

Attach the fee schedule here, if applicable.

[Response Begins]
The measure steward (CMS) will provide a freely available, publicly accessible manual explaining the calculation of this measure from data elements allowing anyone with claims data to be able to compute it. Measure scores for every hospice will be displayed on CMS's public website.
[Response Ends]

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.

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

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

4a.01.

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

Name of program and sponsor

URL

Purpose

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

Level of measurement and setting

[Response Begins]

Not in use
This is a newly developed measure by CMS for the Hospice Quality Reporting Program. CMS has announced in rulemaking the measure will be publicly reported no earlier than May 2022.
[Response Ends]

4a.02. Check all planned uses.

[Response Begins]
Public reporting
[Response Ends]

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

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

[Response Begins]
This measure is newly-developed. It was submitted to NQF as a measure under consideration in 2019. In recent rulemaking, the measure steward, CMS, has indicated they plan to begin publicly reporting this measure not sooner than 2022 via the CMS Care Compare website (https://www.medicare.gov/care-compare/).
[Response Ends]

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

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

[Response Begins]
This measure is newly-developed. It was submitted to NQF as a measure under consideration in 2019. In recent rulemaking, the measure steward, CMS, has indicated they plan to begin publicly reporting this measure not sooner than 2022 via the CMS Care Compare website (https://www.medicare.gov/care-compare/).
[Response Ends]

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

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

[Response Begins]
This measure has not yet been fully implemented. However, CMS has expressed in rulemaking that this measure will be publicly-reported no earlier than 2022. Individual hospice scores benchmarked to national averages were released confidentially to hospices in the fall of 2021.
[Response Ends]

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

[Response Begins]
The confidential hospice scores benchmarked to national averages have only been released one time so far. The measure steward (CMS) provided an educational webinar to assist hospices in understanding the report format and also how to interpret their score (relative to the national average). CMS also released a manual detailing the specifications of the measure and how it could be calculated by individual hospices (or anyone with access to claims data).
[Response Ends]

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

[Response Begins]
N/A, this measure is not yet fully implemented.
[Response Ends]

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

[Response Begins]
N/A, this measure is not yet fully implemented.
[Response Ends]

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

[Response Begins]
N/A, this measure is not yet fully implemented.

[Response Ends]

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

[Response Begins]
N/A, this measure is not yet fully implemented.
[Response Ends]

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

[Response Begins]
At this time of initial endorsement, this measure has not yet been implemented, and thereby no improvement could yet be detected. However, collecting information about hospice staff visits near death will encourage hospices to better monitor patients' status for signs of impending death, and, at the appropriate time, visit patients and caregivers and provide services that will address their care needs. This will lead to improved quality of care for patients and their caregivers during the patients' last days of life.
[Response Ends]

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

[Response Begins]
N/A – this measure is not yet implemented
[Response Ends]

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

[Response Begins]
N/A – this measure is not yet implemented
[Response Ends]

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.

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

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

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

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

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

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

[Response Begins]
CMS has developed a new measure concept, the "Hospice Care Index", a composite of ten claims-based indicators, implemented in the FY 2022 Final Rule (https://www.cms.gov/medicaremedicare-fee-service-paymenthospicehospice-regulations-and-notices/cms-1754-f). One of the ten indicators deals with visits by a nurse or social worker in the last days of life (although there are differences: the index indicator includes non-RN nurses such as Licensed Practical Nurses, it only measures receipt of visits in at least one day in the last three, and the index itself is more concerned with categorizing hospices into the upper 90th and lower 10th percentiles). CMS anticipates submitting for endorsement in a future cycle.
[Response Ends]

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

[Response Begins]
Yes
[Response Ends]

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

[Response Begins]
N/A
[Response Ends]

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

Provide analyses when possible.

[Response Begins]
N/A
[Response Ends]

Appendix

Supplemental materials may be provided in an appendix.: No appendix

Contact Information

Measure Steward (Intellectual Property Owner) : CMS - DCPAC
Measure Steward Point of Contact: Natanov, Rebekah, rebekah.natanov@cms.hhs.gov
Abdur-Rahman, Ihsan, ihsan.abdur-rahman@cms.hhs.gov

Measure Developer if different from Measure Steward: Abt Associates
Measure Developer Point(s) of Contact: Harrison, Zinnia, zinnia_harrison@abtassoc.com
Christian, T.J., thomas_christian@abtassoc.com

Additional Information

1. **Provide any supplemental materials, if needed, as an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be collated one file with a table of contents or bookmarks. If material pertains to a specific criterion, that should be indicated.**

- **[Response Begins]**
- No appendix
- **[Response Ends]**

2. ***List the workgroup/panel members' names and organizations.**

- *Describe the members' role in measure development.*
- **[Response Begins]**
- The steward and developer relied on internal subject matter experts to develop this measure.
- **[Response Ends]**

3. **Indicate the year the measure was first released.**

- **[Response Begins]**
- Measure is not yet publicly reported; however, it was discussed in FY 2022 rulemaking and noted it would be publicly reported no earlier than May, 2022. However, hospice providers have had access to their own scores (obtained in their confidential QM reports) since September, 2021. Data available to hospices now is the same as used for development of the quality measure (FY2018-2019), but later releases would have less of a time lag: hypothetically if data were released in May 2022 the release would include data into the third quarter of 2021.
- **[Response Ends]**

4. **Indicate the month and year of the most recent revision.**

- **[Response Begins]**
- Measure is not yet endorsed.
- **[Response Ends]**

5. **Indicate the frequency of review, or an update schedule, for this measure.**

- **[Response Begins]**
- Anticipating three-year maintenance of endorsement review beginning with endorsement.
- **[Response Ends]**

6. **Indicate the next scheduled update or review of this measure.**

- **[Response Begins]**
  Assuming NQF endorsement in 2022, this measure would next be submitted for maintenance of endorsement three years after, or, in the Fall 2025 review cycle.
  [Response Ends]

7. **Provide a copyright statement, if applicable. Otherwise, indicate "N/A".**

[Response Begins]
N/A
Response Ends]

8. **State any disclaimers, if applicable. Otherwise, indicate "N/A".**

[Response Begins]
N/A
[Response Ends]

9. Provide any additional information or comments, if applicable. Otherwise, indicate "N/A"

[Response Begins]
N/A
[Response Ends]