

NQF #0208 Family Evaluation of Hospice Care
NATIONAL QUALITY FORUM

Measure Submission and Evaluation Worksheet 5.0

This form contains the information submitted by measure developers/stewards, organized according to NQF's measure evaluation criteria and process. The evaluation criteria, evaluation guidance documents, and a blank online submission form are available on the [submitting standards web page](#).

NQF #: 0208 NQF Project: Palliative Care and End-of-Life Care
(for Endorsement Maintenance Review) Original Endorsement Date: Aug 10, 2009 Most Recent Endorsement Date: Aug 10, 2009
BRIEF MEASURE INFORMATION
De.1 Measure Title: Family Evaluation of Hospice Care
Co.1.1 Measure Steward: National Hospice and Palliative Care Organization
De.2 Brief Description of Measure: Composite Score: Derived from responses to 17 items on the Family Evaluation of Hospice Care (FEHC) survey presented as a single score ranging from 0 to 100. Global Score: Percentage of best possible response (Excellent) to the overall rating question on the FEHC survey. Target Population: The FEHC survey is an after-death survey administered to bereaved family caregivers of individuals who died while enrolled in hospice. Timeframe: The survey measures family members perception of the quality of hospice care for the entire enrollment period, regardless of length of service.
2a1.1 Numerator Statement: Composite Score: Numerator is the hospice's composite score, which is the weighted incidence of problem scores derived from responses from 17 items on the FEHC survey. The 17 questions focus on the following aspects of hospice care: symptom management, communication, provision of information, emotional support, and care coordination. Global Score: Numerator is the number of best possible responses (excellent) to the overall rating question on the FEHC survey.
2a1.4 Denominator Statement: Composite Score: 100 (100 is the best possible composite score which indicates 0% incidence of problem scores). Global Score: Total number of responses to the overall rating of care quality on the FEHC survey, question G1.
2a1.8 Denominator Exclusions: Composite Score: If a survey respondent did not enter a response to more than 14 of the 17 FEHC survey questions included in calculation of the composite score then a composite score will not be calculated for that survey and the survey will not be included in the calculation of a composite score for the hospice. Global Score: If survey respondent has not entered a response to overall rating question (G1), the question is not included in the denominator.
1.1 Measure Type: Composite 2a1.25-26 Data Source: Patient Reported Data/Survey 2a1.33 Level of Analysis: Facility, Population : National 1.2-1.4 Is this measure paired with another measure? No
De.3 If included in a composite, please identify the composite measure (title and NQF number if endorsed): N/A

STAFF NOTES <i>(issues or questions regarding any criteria)</i>
Comments on Conditions for Consideration:
Is the measure untested? Yes <input type="checkbox"/> No <input type="checkbox"/> If untested, explain how it meets criteria for consideration for time-limited

endorsement:
1a. Specific national health goal/priority identified by DHHS or NPP addressed by the measure (check De.5): 5. Similar/related endorsed or submitted measures (check 5.1): Other Criteria:
Staff Reviewer Name(s):

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

Importance to Measure and Report is a threshold criterion that must be met in order to recommend a measure for endorsement. All three subcriteria must be met to pass this criterion. See [guidance on evidence](#).
Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. (evaluation criteria)

1a. High Impact: H M L I
(The measure directly addresses a specific national health goal/priority identified by DHHS or NPP, or some other high impact aspect of healthcare.)

De.4 Subject/Topic Areas (Check all the areas that apply): [Cancer, Infectious Diseases, Neurology, Pulmonary/Critical Care : Chronic Obstructive Pulmonary Disease \(COPD\), Renal](#)

De.5 Cross Cutting Areas (Check all the areas that apply): [Care Coordination, Palliative Care and End of Life Care, Patient and Family Engagement](#)

1a.1 Demonstrated High Impact Aspect of Healthcare: [Patient/societal consequences of poor quality, Severity of illness](#)

1a.2 If "Other," please describe:

1a.3 Summary of Evidence of High Impact (Provide epidemiologic or resource use data):
[In 2009, an estimated 1.56 million patients received services from hospice. 1,020,00 patients died under hospice care in 2009. For 2009, NHPCO estimates that approximately 41.6% of all deaths in the United States were under the care of a hospice program.](#)

1a.4 Citations for Evidence of High Impact cited in 1a.3: [NHPCO Facts and Figures: Hospice Care in America. Alexandria, VA: National Hospice and Palliative Care Organization, September 2010.](#)

1b. Opportunity for Improvement: H M L I
(There is a demonstrated performance gap - variability or overall less than optimal performance)

1b.1 Briefly explain the benefits (improvements in quality) envisioned by use of this measure:
[Use of this measure affords hospices a valid means of ensuring quality of care by providing useful, meaningful, and actionable information that can be incorporated into their Quality Assurance/Performance Improvement \(QAPI\) programs. Implementation of a QAPI program is a requirement in the Medicare Conditions of Participation for hospices. Use of the measure will facilitate improved quality in the following aspects of hospice care: symptom management, communication, provision of information, emotional support, and care coordination.](#)

1b.2 Summary of Data Demonstrating Performance Gap (Variation or overall less than optimal performance across providers):
[For Maintenance – Descriptive statistics for performance results for this measure - distribution of scores for measured entities by quartile/decile, mean, median, SD, min, max, etc.]
[The final sample included 1,060 \(81.7%\) of submitting hospices, with surveys representing the care given to 172,558 \(83.0%\) hospice patients. In a two-year sample of 1231 hospices, the mean Composite Score was 86.6% with a median of 86.5%. The lowest score recorded over the two-years was 73.3% \(SD = 3.12%\) and the highest was 96.3%, demonstrating a clear and significant range of scores. The inter quartile range of scores was 84.5% and 88.5% for the 25th and 75th percentiles respectively. The very low skewness and kurtosis of the measure also indicates good normality in the distribution of responses.](#)

1b.3 Citations for Data on Performance Gap: [For Maintenance – Description of the data or sample for measure results reported]

in 1b.2 including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included]

A sample of FEHC survey submissions was collected for the 2009 calendar year. During that year 1,297 hospices across the United States, provided 208,025 completed surveys. Composite scores were calculated for each of the surveys and then averaged to produce the overall hospice score. Finally, to prevent skewing results, hospices that provided less than 25 composite scores during the two years were excluded from the sample. The final sample included 1,060 (81.7%) of submitting hospices, with surveys representing the care given to 172,558 (83.0%) hospice patients.

1b.4 Summary of Data on Disparities by Population Group: [For Maintenance – Descriptive statistics for performance results for this measure by population group]

A statistically significant but non-substantial difference in Composite Score performances between racial and ethnic groups was observed. The two-year average Composite Scores for Hispanic and non-Hispanic patients were 84.3% (95%CI: 83.9% to 85.6%) and 86.0% (95%CI: 86.1% to 86.0%) respectively. The average Composite Scores for patients who's race was identified as white verses all other races was 86.0% (95% CI: 85.9% to 86.1%) and 85.5% (95% CI: 85.4% to 86.1%).

The remarkably small confidence intervals for all race and ethnic group mean scores are indicative of the extremely large sample sizes. Since the sample is so large, it is exceptionally over-powered. This allows for the detection of statistically significant yet non-meaningful disparities between comparison groups.

1b.5 Citations for Data on Disparities Cited in 1b.4: [For Maintenance – Description of the data or sample for measure results reported in 1b.4 including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included]

The analysis for 1b.4 consisted of two years (2009 through 2010) worth of Composite Scores, comprising 1,231 hospices and over 340,000 survey responses. Of the 347,356 patients represented by the completed FEHC surveys 91.9% were identified as "White" by their primary caregiver. The remaining 8.9% of patients were identified as either Black (4.1%), other/multiracial (2.3%), Asian/Pacific Islander (0.95%) or, American Indian/Alaska Native (0.78%). Only 3.1% of the patients were identified by their primary caregivers as being of Hispanic ethnicity.

1c. Evidence (Measure focus is a health outcome OR meets the criteria for quantity, quality, consistency of the body of evidence.)

Is the measure focus a health outcome? Yes No **If not a health outcome, rate the body of evidence.**

Quantity: H M L I Quality: H M L I Consistency: H M L I

Quantity	Quality	Consistency	Does the measure pass subcriterion 1c?
M-H	M-H	M-H	Yes <input type="checkbox"/>
L	M-H	M	Yes <input type="checkbox"/> IF additional research unlikely to change conclusion that benefits to patients outweigh harms: otherwise No <input type="checkbox"/>
M-H	L	M-H	Yes <input type="checkbox"/> IF potential benefits to patients clearly outweigh potential harms: otherwise No <input type="checkbox"/>
L-M-H	L-M-H	L	No <input type="checkbox"/>

Health outcome – rationale supports relationship to at least one healthcare structure, process, intervention, or service	Does the measure pass subcriterion 1c? Yes <input type="checkbox"/> IF rationale supports relationship
--	---

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

Measure focus is a health outcome that reflects patient/family caregiver experience of hospice care. Because the measure is a composite/global measure of hospice care, multiple intermediate clinical outcomes and hospice care processes are linked in the areas of symptom management, communication, provision of information, emotional support, and care coordination.

The survey that serves as the data source for the measure is based on conceptual model of patient focused, family centered medical care. This model was developed based on expert advice, a structured review of guidelines, and focus groups with bereaved family members.

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

- Other
- Conceptual model of patient focused, family centered care

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

The model was developed by Dr. Joan Teno and colleagues at Brown University as part of a larger research project and mortality followback study that examined the quality of end-of-life care in multiple care settings. The model posits that high quality care at the end of life is obtained when health care institutions: 1) provide the desired level of symptom palliation and emotional support; 2) treat the patient with respect; 3) promote shared decision making; 4) attend to the needs of caregivers for information and skills in providing care for the patient; 5) provide emotional support to the family before and after the patient's death; and 6) coordinates care across settings of care and health care providers. The model was developed based on advice from a panel of experts in end-of-life care, a structured review of existing guidelines for provision of end-of-life care, and analysis of data from focus groups conducted with bereaved family members. This model serves as the conceptual framework for the FEHC survey, which is the data source for the measure.

Measure is also consistent with the Domains delineated in the NQF consensus report titled A National Framework and Preferred Practices for Palliative and Hospice Care Quality endorsed by NQF in 2006.

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

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

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

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

N/A

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

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

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

1c.12 If other, identify and describe the grading scale with definitions: No formal system

1c.13 Grade Assigned to the Body of Evidence: Grade not assigned

1c.14 Summary of Controversy/Contradictory Evidence: N/A

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

N/A

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

N/A

1c.17 Clinical Practice Guideline Citation: N/A

1c.18 National Guideline Clearinghouse or other URL: N/A

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

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation

and any disclosures regarding bias:

1c.21 System Used for Grading the Strength of Guideline Recommendation: [Other](#)

1c.22 If other, identify and describe the grading scale with definitions: [No formal system used](#)

1c.23 Grade Assigned to the Recommendation: [No grad assigned.](#)

1c.24 Rationale for Using this Guideline Over Others: [No Guideline used](#)

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

1c.25 Quantity: [High](#) 1c.26 Quality: [High](#) 1c.27 Consistency: [High](#)

Was the threshold criterion, *Importance to Measure and Report*, met?

(1a & 1b must be rated moderate or high and 1c yes) Yes No

Provide rationale based on specific subcriteria:

For a new measure if the Committee votes NO, then STOP.

For a measure undergoing endorsement maintenance, if the Committee votes NO because of 1b. (no opportunity for improvement), it may be considered for continued endorsement and all criteria need to be evaluated.

2. RELIABILITY & VALIDITY - SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (**evaluation criteria**)

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate field. Supplemental materials may be referenced or attached in item 2.1. See [guidance on measure testing](#).

S.1 Measure Web Page (*In the future, NQF will require measure stewards to provide a URL link to a web page where current detailed specifications can be obtained*). Do you have a web page where current detailed specifications for this measure can be obtained? [Yes](#)

S.2 If yes, provide web page URL: www.nhpco.org/fehcc

2a. RELIABILITY. Precise Specifications and Reliability Testing: H M L I

2a1. Precise Measure Specifications. (*The measure specifications precise and unambiguous.*)

2a1.1 Numerator Statement (*Brief, narrative description of the measure focus or what is being measured about the target population, e.g., cases from the target population with the target process, condition, event, or outcome*):

[Composite Score: Numerator is the hospice's composite score, which is the weighted incidence of problem scores derived from responses from 17 items on the FEHC survey. The 17 questions focus on the following aspects of hospice care: symptom management, communication, provision of information, emotional support, and care coordination.](#)

[Global Score: Numerator is the number of best possible responses \(excellent\) to the overall rating question on the FEHC survey.](#)

2a1.2 Numerator Time Window (*The time period in which the target process, condition, event, or outcome is eligible for inclusion*):
[Time period eligible for inclusion is the entire length of service the patient was enrolled in hospice.](#)

2a1.3 Numerator Details (*All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, codes with descriptors, and/or specific data collection items/responses*):

[Composite Score: Responses to the following questions on the FEHC survey:](#)

[B2 \(How much medicine did the patient receive for his/her pain?\)](#)

[B4 \(Did you want more information than you got about the medicines used to manage the patient's pain?\)](#)

[B6 \(How much help in dealing with his/her breathing did the patient receive while under the care of hospice?\)](#)

B8 (Did you want more information than you got about what was being done for the patient's trouble with breathing?)
 B10 (How much help in dealing with these feelings did the patient receive?)(refers to feelings of anxiety and sadness)
 D3 (How confident did you feel about doing what you needed to do in taking care of the patient?)
 D4 (How confident were you that you knew as much as you needed to about the medicines being used to manage the patient's pain, shortness of breath, or other symptoms?)
 D5 (How often did the hospice team keep you or other family members informed about the patient's condition?)
 D7 (Would you have wanted more information about what to expect while the patient was dying?)
 D8 (How confident were you that you knew what to expect while the patient was dying?)
 D9 (How confident were you that you knew what to do at the time of death?)
 E2 (Did you have as much contact of that kind as you wanted?)(refers to spiritual care)
 E3 (How much emotional support did the hospice team provide to you prior to the patient's death?)
 E4 (How much emotional support did the hospice team provide to you after the patient's death?)
 F1 (How often did someone from the hospice team give confusing or contradictory information about the patient's medical treatment?)
 F2 (While under the care of hospice, was there always one nurse who was identified as being in charge of the patient's overall care?)
 F3 (Was there any problem with hospice doctors or nurses not knowing enough about the patient's medical history to provide the best possible care?)

Global Score: Number of responses of "Excellent" to the overall rating of care quality on the FEHC survey, question G1 (Overall, how would you rate the care the patient received while under the care of hospice?)

2a1.4 Denominator Statement (Brief, narrative description of the target population being measured):

Composite Score: 100 (100 is the best possible composite score which indicates 0% incidence of problem scores).

Global Score: Total number of responses to the overall rating of care quality on the FEHC survey, question G1.

2a1.5 Target Population Category (Check all the populations for which the measure is specified and tested if any): Adult/Elderly Care, Populations at Risk

2a1.6 Denominator Time Window (The time period in which cases are eligible for inclusion):

Time period eligible for inclusion is the entire length of service the patient was enrolled in hospice.

2a1.7 Denominator Details (All information required to identify and calculate the target population/denominator such as definitions, codes with descriptors, and/or specific data collection items/responses):

Composite Score: 100 (100 is the best possible composite score which indicates 0% incidence of problem scores).

Global Score: All responses to overall rating of care question on the FEHC survey (G1) are included. If survey respondent has not entered a response, the question is not included in the denominator.

2a1.8 Denominator Exclusions (Brief narrative description of exclusions from the target population):

Composite Score: If a survey respondent did not enter a response to more than 14 of the 17 FEHC survey questions included in calculation of the composite score then a composite score will not be calculated for that survey and the survey will not be included in the calculation of a composite score for the hospice.

Global Score: If survey respondent has not entered a response to overall rating question (G1), the question is not included in the denominator.

2a1.9 Denominator Exclusion Details (All information required to identify and calculate exclusions from the denominator such as definitions, codes with descriptors, and/or specific data collection items/responses):

Composite Score: If a survey respondent did not enter a response to more than 3 of the 17 FEHC survey questions included in calculation of the composite score then a composite score will not be calculated for that survey and the survey will not be included in the calculation of a composite score for the hospice.

Global Score: If survey respondent has not entered a response to overall rating question (G1), the question is not included in the denominator.

2a1.10 Stratification Details/Variables (All information required to stratify the measure results including the stratification variables, codes with descriptors, definitions, and/or specific data collection items/responses):

N/A

2a1.11 Risk Adjustment Type (Select type. Provide specifications for risk stratification in 2a1.10 and for statistical model in 2a1.13): No risk adjustment or risk stratification **2a1.12 If "Other," please describe:**

2a1.13 Statistical Risk Model and Variables (Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development should be addressed in 2b4.):

N/A

2a1.14-16 Detailed Risk Model Available at Web page URL (or attachment). Include coefficients, equations, codes with descriptors, definitions, and/or specific data collection items/responses. Attach documents only if they are not available on a webpage and keep attached file to 5 MB or less. NQF strongly prefers you make documents available at a Web page URL. Please supply login/password if needed:

2a1.17-18. Type of Score: Weighted score/composite/scale

2a1.19 Interpretation of Score (Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score): Better quality = Higher score

2a1.20 Calculation Algorithm/Measure Logic (Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.):

1. Obtain data (responses to questions) for the 17 questions from the FEHC survey that comprise the Composite Score
2. Dichotomize all constituent questions into a) most desirable response; and b) all other responses for each question. "No answer" or non-valid responses = null.
3. Calculate composite score for each of the 17 questions for each survey
4. Calculate composite score for hospice by averaging the composite scores for each survey

2a1.21-23 Calculation Algorithm/Measure Logic Diagram URL or attachment:

URL

www.nhpc.org/fehcc

2a1.24 Sampling (Survey) Methodology. If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey and guidance on minimum sample size (response rate):

1. Download FEHC survey: Hospice downloads the survey from NHPCO web site (www.nhpc.org/fehcc)
2. Administer FEHC survey: Hospice mails survey (on a rolling basis) to caregivers of all patients who died while enrolled in hospice services. NHPCO recommends mailing the surveys from 1 to 3 months post-death.
3. Data collection and submission: Surveys are returned to hospice. As soon as surveys are returned, data submission can begin. Data submission is done online on a quarterly schedule through the FEHC web-based data submission system. The web-based system is accessed through the NHPCO Web site.

A hospice may also use a vendor for survey administration.

2a1.25 Data Source (Check all the sources for which the measure is specified and tested). If other, please describe:

Patient Reported Data/Survey

2a1.26 Data Source/Data Collection Instrument (*Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.*): Family Evaluation of Hospice Care (FEHC) survey. The survey is based on structured literature review, (1) cognitive testing, (2) pre-test, (2) and national survey of the quality of end of life care. (3) The conceptual model is patient focused, family centered care (1) that posits that high quality care at the end of life is obtained when health care institutions: 1) provide the desired level of symptom palliation and emotional support; 2) treat the patient with respect; 3) promote shared decision making; 4) attend to the needs of caregivers for information and skills in providing care for the patient; 5) provide emotional support to the family before and after the patient's death; and 6) coordinates care across settings of care and health care providers.

2a1.27-29 Data Source/data Collection Instrument Reference Web Page URL or Attachment: [URL www.nhpco.org/fehc](http://www.nhpco.org/fehc) (Survey Materials section of the web page)

2a1.30-32 Data Dictionary/Code Table Web Page URL or Attachment:

2a1.33 Level of Analysis (*Check the levels of analysis for which the measure is specified and tested*): Facility, Population : National

2a1.34-35 Care Setting (*Check all the settings for which the measure is specified and tested*): Hospice

2a2. Reliability Testing. (*Reliability testing was conducted with appropriate method, scope, and adequate demonstration of reliability.*)

2a2.1 Data/Sample (*Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included*):

The initial work on reliability and validity of the FEHC survey was completed on a sample of 156 bereaved family members who died receiving care from hospice, nursing home, and hospital. This was published in JPSM in 2001. (1) In this first test, we examined short-term test-retest (4 to 8 weeks after the original interview) and internal consistency was examined with Cronbach's alpha among the entire sample of 156 interviews for each of the proposed composite scores. This analysis was updated for the 2004 publication of the mortality followback survey published in JAMA (2) with the result published in an online appendix on a Brown University web site (http://www.chcr.brown.edu/dying/MEASURES_JAMA_PAPER_LAST_PLACE_OF_CARE.PDF).

(1) Teno JM, Clarridge B, Casey V, Edgman-Levitan S, Fowler J. Validation of Toolkit After-Death Bereaved Family Member Interview. *J Pain Symptom Manage.* 2001 Sep 2001;22(3):752-758.

(2) Teno JM, Clarridge BR, Casey V, et al. Family perspectives on end-of-life care at the last place of care. *JAMA.* 2004 Jan 7 2004;291(1):88-93.

Composite Score:

A sample of FEHC survey response submissions was collected through the 2009 calendar year. During that year 1,297 hospices across the United States, provided 208,025 completed surveys. Composite scores were calculated for each of the surveys and then averaged to produce the overall hospice score. Finally, to prevent skewing results, hospices that provided less than 25 composite scores during the two years were excluded from the sample. The final sample included 1,060 (81.7%) of submitting hospices, with surveys representing the care given to 172,558 (83.0%) hospice patients.

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

1. Short term reliability over 4-8 weeks was done among 29 family members with all the key items of the survey. The reliability was examined among a Kappa statistic or an intraclass correlation. This was published in 2001 JPSM article. (1)

(1) Teno JM, Clarridge B, Casey V, Edgman-Levitan S, Fowler J. Validation of Toolkit After-Death Bereaved Family Member Interview. *J Pain Symptom Manage.* 2001 Sep 2001;22(3):752-758.

2. The internal consistency of all proposed composite scores was examined with Cronbach's alpha with items dropped. This has been done for the initial validation study and national mortality follow back survey in 2004 (2). This was published in the 2001 JPSPM article (1) and online with the 2004 JAMA publication.(2)

(1) Teno JM, Clarridge B, Casey V, Edgman-Levitan S, Fowler J. Validation of Toolkit After-Death Bereaved Family Member Interview. J Pain Symptom Manage. 2001 Sep 2001;22(3):752-758.

(2) Teno JM, Clarridge BR, Casey V, et al. Family perspectives on end-of-life care at the last place of care. JAMA. 2004 Jan 7 2004;291(1):88-93.

Composite Score:

Internal consistency reliability was tested using the Cronbach's alpha statistic. This method was chosen because the measure is based on responses to 17 questions on the FEHC survey and internal consistency is an appropriate method for multi-item scales.

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

Based on reliability testing of the survey given 4-8 weeks apart, we decided to drop 4 items which have a Kappa under 0.4. Two items that we retained have skewed marginals with a high percent agreement of 79% and 82%. The remaining items all had a Kappa or ICC above 0.49 and higher.

Examination of internal consistency was done with the Cronbach's alpha. We relied on a national mortality follow back survey published in JAMA in 2004 to report the internal consistency of NQF evaluation. As we have previously stated, there are 3 composite scores. For the composite score examining physician communication, the Cronbach's Alpha was 0.67 while the other two composite scores exceed 0.70.

Based on reliability testing of the survey given 4-8 weeks apart, we decided to drop 4 items which have a Kappa under 0.4. Two items that we retained have skewed marginals with a high percent agreement of 79% and 82% which is a known problem with the Kappa Statistic. The remaining items all had a Kappa or ICC above 0.49 and higher.

Composite Score:

The internal consistency of the Composite score was examined with Cronbach's alpha. Composite Score Cronbach's alpha = .7977

2b. VALIDITY. Validity, Testing, including all Threats to Validity: H M L I

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

Measure focus is a health outcome that reflects patient/family caregiver experience of hospice care. The measure is a composite/global measure of hospice care that includes multiple key areas of hospice care: symptom management, communication, provision of information, emotional support, and care coordination. The unit of care for hospice is the patient AND family caregiver. Therefore, measures of quality of hospice care should include the perspective and the experience of family caregivers of hospice patients.

The development of the FEHC survey, which is the data source for the measure, was informed by analysis of data from focus groups of bereaved family members and consequently reflects the values and preferences for end-of-life care of family caregivers. There are no differences between the measure specifications and the evidence cited for the measure focus.

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

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

A two year sample of FEHC survey submissions was collected from 2009 through 2010. During those two years period 1,404 hospices across the United States, provided 433,410 completed surveys. Composite scores were calculated for each of the surveys and then averaged to produce the overall hospice score. Finally, to prevent skewing results, hospices that provided less than 25 composite scores during the two years were excluded from the sample. The final sample included 1,231 (87.7%) of submitting hospices, with surveys representing the care given to 357,008 (82.4%) hospice patients.

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

To assess Composite Score criterion (predictive) validity, we statistically compared the composite score performance with the primary caregivers overall evaluation of care quality (rated either “excellent”, “very good”, “good”, “fair”, and “poor”). While caregiver perception of quality is subjective and can be influenced by many external, non-care related factors, it is expected that the perception of quality should correlate well to the composite score if the composite score is, in fact, measuring quality. To test this hypothesis an analysis of variance (ANOVA) was performed modeling the composite score against the caregiver’s quality of rating. A Tukey’s test was also performed to identify statistically significant differences in mean composite scores between rating groups. Finally, a non-parametric Spearman’s correlation analysis was performed to identify the correlation of caregiver rating to composite score.

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

The ANOVA results indicated that significant differences in mean composite scores existed between caregiver rating groups (F=43165.8; P<0.001). The Tukey’s test for difference between caregiver groups found that significant differences existed between the mean scores of all rating groups.

However, univariate analysis found that the mean composite score was substantially higher as the caregiver perception of care improved. The mean composite score for those rating the care as “excellent” was 90.3%, as “very good” was 76.2%, as “good” was 60.8%, as “fair” was 43.5%, and as poor was 30.5%. The Spearman’s correlation coefficients found a significant direct correlation between the composite score and the caregivers rating of care quality (Spearman’s Rho = .444: P<0.0001).

The above results clearly demonstrate a direct and significant relationship between a hospice’s composite score and the overall rating of care quality given by the family caregiver.

The expected relationship showing an increasing mean composite score as caregiver rating also increased was observed in univariate analysis and confirmed as statistically significant in the analysis of variance with Tukey’s test. The slightly lower than what might be desired correlation coefficient was expected since it is known that external non-care related factors confound the relationship between perception of care quality and the actual quality of care received.

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

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

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

To analyze the impact of missing data, we tested for significant differences in the frequency of missing responses between racial and ethnic groups. An analysis of variance was performed to determine if there was a significant difference in missing Overall Score response or Composite Score responses in racial groups. Where a significant difference was found, a Tukey’s test was performed to identify which racial groups had a significantly higher or lower likelihood of having either score. A t-test was performed to examine the difference in likelihood that surveys relating to Hispanic patients were more likely to have a Composite Score or Overall Score compared to those of non-Hispanic patients. Statistical significance was considered at a P-Value of 0.05 or lower.

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

To analyze the impact of missing data, we tested for significant differences in the frequency of missing responses between racial and ethnic groups. An analysis of variance was performed to determine if there was a significant difference in missing Overall Score response or Composite Score responses in racial groups. Where a significant difference was found, a Tukey’s test was performed to identify which racial groups had a significantly higher or lower likelihood of having either score. A t-test was performed to examine the difference in likelihood that surveys relating to Hispanic patients were more likely to have a Composite Score or Overall Score compared to those of non-Hispanic patients. Statistical significance was considered at a P-Value of 0.05 or lower.

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

In the analysis of variance a statistically significant difference was found in the likelihood of having a Composite Score based on racial category (F value = 18.5: P<0.001). The Tukey’s test revealed that a significantly greater proportion of patients identified as Asian had surveys with Composite Scores (89.9%: 95%CI= 88.9% to 90.9%) compared to those identified as White (88.3%: 95% CI = 88.2% to 88.4%) and Black (86.1%: 95%CI= 85.5% to 86.7%). There was also a significantly greater proportion of surveys with

Composite Scores from those patients who were reported as other/multiracial (88.6%: 95% CI = 88.0% to 89.3%) compared to those reported as Black. No other significant differences between racial groups were identified.

The differences in the percent of available Composite Scores between racial groups, while statistically significant, were not substantial. The observed differences were 3.8% (between Asian patients and Black patients) at the highest, with most significant differences within two percentage points. This indicates that no substantial differences exist for in the likelihood of Composite Scores being present based on race.

Similar results were observed in the analysis for the differences in the presence of the Overall Score for based on racial category. Only surveys for those patients, whose caregivers identified as other/multiracial, had a significantly different likelihood of have an Overall Score (97.3% 95% CI: 96.9% to 97.6%) compared to those identified as Asian (98.1%: 95%CI = 67.6% to 98.5%), Black (97.8%: 95% CI = 97.7% to 97.8%), American Indian/Alaska Native (97.2%: 95% CI = 96.5% to 97.8%), and White (97.78%: 95% CI = 97.72% to 97.82). Again, while statistically significant, the differences in the proportion without Overall Scores (all less than 1.0%) could not have major influence on results.

T-T test results for the difference in proportion of present Composite Scores by ethnicity also show a significant difference between those patients identified as Hispanic versus those not (Satterthwaite t-value = 6.01: P<0.0001). Much like the difference between racial groups, the absolute difference between ethnic groups was only marginally substantial (90.2% for Hispanic patients verses 88.5% for non-Hispanic). Similarly, the t-test showed a significant difference in the proportion available Overall Scores for surveys from Hispanic patients compared to non-Hispanic patients (Satterthwaite t-value = 2.41: P-value = 0.012). The proportion of present Overall Scores for Hispanic patients was 98.3%. When compared to the proportion of non-Hispanic patients with an Overall Score (97.9%) the differences are again non-substantial.

In conclusion, this analysis shows that there are significant difference exist in the proportional presence of Composite and Overall scores based on race and ethnicity. However, in all cases the absolute difference in proportions was non-substantial. This indicates that while differences exist, their potential effect on the validity of the results is minimal, if any.

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

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

N/A

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

N/A

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

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment: Usual reasons for risk adjustment do not apply in the context of hospice care. Measure applies equally across all patient characteristics regardless of case-mix.

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

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

A two year sample of FEHC survey submissions was collected from 2009 through 2010. During this two years period 1,404 hospices across the United States, provided 433,410 completed surveys. Composite scores were calculated for each of the surveys and then averaged to produce the overall hospice score. Finally, to prevent skewing results, hospices that provided less than 25 composite scores during the two years were excluded from the sample. The final sample included 1,231 (87.7%) of submitting hospices, with surveys representing the care given to 357,008 (82.4%) hospice patients.

2b5.2 Analytic Method (Describe methods and rationale to identify statistically significant and practically/meaningfully differences

in performance):

To show statistically significant differences in hospice's mean composite scores an analysis of variance was performed modeling hospice 2-year performance. Also, univariate analysis was utilized to demonstrate the distribution and variance in scores among hospices over the course of two years.

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

The analysis of variance showed a statistically significant difference exists in composite scores between hospices (F=7.63; P<0.0001). Univariate analysis showed the average 2-year mean hospice composite score was 86.6% (STD = 0.031) with a maximum mean score of 96.3% and a minimum mean score of 73.3%.

The wide range of hospice scores demonstrates that there is capacity within the measure for hospices with lower mean composite score to substantially improve their results. The analysis of variance demonstrates that statistically significant differences occur between hospices. The two results combined indicate that a hospice that improves their score can be assured that substantial increases in results are likely also significant from both statistical and clinical perspectives.

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

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

N/A

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

N/A

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

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

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

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

N/A

2.1-2.3 Supplemental Testing Methodology Information:

Steering Committee: Overall, was the criterion, Scientific Acceptability of Measure Properties, met? (Reliability and Validity must be rated moderate or high) Yes No

Provide rationale based on specific subcriteria:

If the Committee votes No, STOP

3. USABILITY

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

C.1 Intended Purpose/Use (Check all the purposes and/or uses for which the measure is intended): Public Reporting, Quality Improvement with Benchmarking (external benchmarking to multiple organizations)

3.1 Current Use (Check all that apply; for any that are checked, provide the specific program information in the following questions): **Quality Improvement with Benchmarking** (external benchmarking to multiple organizations), **Quality Improvement (Internal to the specific organization)**

3a. Usefulness for Public Reporting: H M L I
 (The measure is meaningful, understandable and useful for public reporting.)

3a.1. Use in Public Reporting - disclosure of performance results to the public at large (If used in a public reporting program, provide name of program(s), locations, Web page URL(s)). If not publicly reported in a national or community program, state the reason AND plans to achieve public reporting, potential reporting programs or commitments, and timeline, e.g., within 3 years of endorsement: **[For Maintenance** – If not publicly reported, describe progress made toward achieving disclosure of performance results to the public at large and expected date for public reporting; provide rationale why continued endorsement should be considered.]

Composite Score is not currently used for public reporting, but results from several questions that are used in the Composite Score and the Global Score are in use. The American Hospice Foundation has a comparative hospice report in testing in several states with the goal of widespread public reporting. The state of Florida also has a state level hospice reporting structure. Both of these initiatives include results from several questions from the FEHC survey that are part of the Composite Score in addition to the Global Score.

The annual AHRQ Quality and Disparity also include results from FEHC survey questions.

AHA research in developing a comparative hospice report has found lack of knowledge of the basic aspects of hospice care to be a major barrier to public understanding comparative performance results and how to evaluate quality in hospice in general. This unfamiliarity with hospice implies that more work needs to be done to ensure that disclosure of performance results to the public can be understandable, meaningful, and useful.

3a.2. Provide a rationale for why the measure performance results are meaningful, understandable, and useful for public reporting. If usefulness was demonstrated (e.g., focus group, cognitive testing), describe the data, method, and results: Extensive multi-round cognitive testing has been conducted by Dr. Shoshanna Sofaer of Baruch College on the American Hospice Foundation comparative hospice report. Dr. Sofaer extensively tested the labels and navigation and the "drill downs" with respect to how data were collected, composite items, etc. She found that what required the most work was the educational aspect of the report; it proved very difficult to overcome the misunderstandings that people had about hospice; consequently an additional 20 interviews were conducted on the educating the public on the basics of hospice care.

3.2 Use for other Accountability Functions (payment, certification, accreditation). If used in a public accountability program, provide name of program(s), locations, Web page URL(s):

3b. Usefulness for Quality Improvement: H M L I
 (The measure is meaningful, understandable and useful for quality improvement.)

3b.1. Use in QI. If used in quality improvement program, provide name of program(s), locations, Web page URL(s): **[For Maintenance** – If not used for QI, indicate the reasons and describe progress toward using performance results for improvement].

1,317 hospice participated in NHPCO's FEHC survey administration and reporting program in 2010 and submitted data from a total of 225,385 surveys. FEHC survey results, including the Composite Score and the Global Score, provide information that hospices use in their Quality Assessment and Performance Improvement (QAPI) programs. Implementation of QAPI programs is mandated in the Medicare Conditions of Participation for hospices.

3b.2. Provide rationale for why the measure performance results are meaningful, understandable, and useful for quality improvement. If usefulness was demonstrated (e.g., QI initiative), describe the data, method and results: Hospices indicated that the measure results are useful in focus groups and an end-user survey conducted in 2009. Analysis of the qualitative data from these two initiatives showed widespread use of the results by hospices in multiple performance measure programs. Hospices use the quarterly measure results provided by NHPCO to set performance improvement goals and to monitor progress toward those those goals.

Overall, to what extent was the criterion, Usability, met? H M L I
 Provide rationale based on specific subcriteria:

4. FEASIBILITY

Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (**evaluation criteria**)

4a. Data Generated as a Byproduct of Care Processes: H M L I

4a.1-2 How are the data elements needed to compute measure scores generated? (Check all that apply).

Data used in the measure are:

Other

Needed data elements are obtained through administration of the FEHC survey

4b. Electronic Sources: H M L I

4b.1 Are the data elements needed for the measure as specified available electronically (Elements that are needed to compute measure scores are in defined, computer-readable fields): *Some data elements are in electronic sources*

4b.2 If ALL data elements are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources:

4c. Susceptibility to Inaccuracies, Errors, or Unintended Consequences: H M L I

4c.1 Identify susceptibility to inaccuracies, errors, or unintended consequences of the measurement identified during testing and/or operational use and strategies to prevent, minimize, or detect. If audited, provide results:

Incorrectly capturing survey responses is the primary error that has been identified for this measure. Many hospice providers utilize manual data entry, whereby completed surveys responses are manually entered into the FEHC online data submission system. Each organization must implement their own data quality control procedures to prevent this type of error, as it is nearly impossible to detect once the data has been entered. However, there are several resources available to assist hospices perform data quality checks. The most prominent of this is the availability of a data download, whereby hospices may download an Excel file containing all of their data entered during that quarter. Hospices may then audit their entered responses against the hardcopy surveys. Finally, many hospices contract with vendors to perform survey administration and data submission on their behalf. Technologic innovations, such as the ability to scan returned surveys, are available to these vendors greatly reduce the risk of data entry.

4d. Data Collection Strategy/Implementation: H M L I

A.2 Please check if either of the following apply (regarding proprietary measures):

4d.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues (e.g., fees for use of proprietary measures):

NHPCO maintains ongoing support (in the form of written materials and one-on-one guidance) for hospice providers who use the measure for all aspects of the FEHC survey process, ranging from survey administration to results interpretation. Monitoring of support requests has not shown any trends in problems or issues that indicated the need for modifications in the approach to data collection. Hospices vary in size and resources, and data collection strategies employed tend to vary with the individual characteristics of the hospices.

Overall, to what extent was the criterion, Feasibility, met? H M L I

Provide rationale based on specific subcriteria:

OVERALL SUITABILITY FOR ENDORSEMENT

Does the measure meet all the NQF criteria for endorsement? Yes No

Rationale:

If the Committee votes No, STOP.

If the Committee votes Yes, the final recommendation is contingent on comparison to related and competing measures.

5. COMPARISON TO RELATED AND COMPETING MEASURES

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the

same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure before a final recommendation is made.

5.1 If there are related measures (either same measure focus or target population) or competing measures (both the same measure focus and same target population), list the NQF # and title of all related and/or competing measures:

5a. Harmonization

5a.1 If this measure has EITHER the same measure focus OR the same target population as [NQF-endorsed measure\(s\)](#): Are the measure specifications completely harmonized?

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

5b. Competing Measure(s)

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

CONTACT INFORMATION

Co.1 Measure Steward (Intellectual Property Owner): [National Hospice and Palliative Care Organization, 1731 King Street, Alexandria, Virginia, 22314](#)

Co.2 Point of Contact: [Carol, Spence, PhD, cspence@nhpco.org, 703-837-3137-](#)

Co.3 Measure Developer if different from Measure Steward: [National Hospice and Palliative Care Organization, 1731 King Street, Alexandria, Virginia, 22314](#)

Co.4 Point of Contact: [Carol, Spence, PhD, cspence@nhpco.org, 703-837-3137-](#)

Co.5 Submitter: [Carol, Spence, PhD, cspence@nhpco.org, 703-837-3137-, National Hospice and Palliative Care Organization](#)

Co.6 Additional organizations that sponsored/participated in measure development:
[Brown University, Center for Gerontology and Healthcare Research, Brown Medical School, Providence, RI. Contact: Dr. Joan Teno](#)

Co.7 Public Contact: [Carol, Spence, PhD, cspence@nhpco.org, 703-837-3137-, National Hospice and Palliative Care Organization](#)

ADDITIONAL INFORMATION

Workgroup/Expert Panel involved in measure development

Ad.1 Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.

[Joan Teno, MD Brown University, Center for Gerontology and Healthcare Research, Brown Medical School. Dr. Teno conducted the original research that served as the basis for the FEHC survey that is the data source for the measure. She also developed the measure in collaboration with the following NHPCO staff:](#)

[Carol Spence, PhD](#)
[Matthew Hasking, MPH](#)

Ad.2 If adapted, provide title of original measure, NQF # if endorsed, and measure steward. Briefly describe the reasons for adapting the original measure and any work with the original measure steward:

Measure Developer/Steward Updates and Ongoing Maintenance Ad.3 Year the measure was first released: 2003 Ad.4 Month and Year of most recent revision: 10,2010 Ad.5 What is your frequency for review/update of this measure? Annual Ad.6 When is the next scheduled review/update for this measure? 10,2011
Ad.7 Copyright statement: Copyright holder of the FEHC survey is Brown University which makes the survey available for use free of charge with the provision it is not modified or sold.
Ad.8 Disclaimers:
Ad.9 Additional Information/Comments:
Date of Submission (MM/DD/YY): 06/13/2011