NATIONAL QUALITY FORUM
Driving measurable health improvements together

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through National Quality Forum's (NQF) Consensus Development Process (CDP). The information submitted by the measure developers/stewards is included after the *Brief Measure Information* and *Preliminary Analysis* sections.

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

Brief Measure Information

NQF #: 1626

Measure Title: Patients Admitted to ICU who Have Care Preferences Documented
Measure Steward: RAND Corporation
Brief Description of Measure: Percentage of vulnerable adults admitted to ICU who survive at least 48 hours who have their care preferences documented within 48 hours OR documentation as to why this was not done.
Developer Rationale: The aim of this measure is to assist healthcare providers in providing care that is consistent with patient preferences.
Numerator Statement: Patients in the denominator who had their care preferences documented within 48 hours of ICU admission or have documentation of why this was not done.
Denominator Statement: All vulnerable adults admitted to ICU who survive at least 48 hours after ICU admission.
Denominator Exclusions: None
Measure Type: Process
Data Source: Paper Medical Records
Level of Analysis: Facility
IF Endorsement Maintenance – Original Endorsement Date: 2012-02-14 09:17 AM
Most Recent Endorsement Date: 10/26/2016 8:21:14 AM

Preliminary Analysis: Maintenance of Endorsement

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

Criteria 1: Importance to Measure and Report

1a. Evidence

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

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

The developer provides the following description for this measure:

- This is a maintenance process measure at the facility level that looks at the percentage of vulnerable adults admitted to the ICU who survive at least 48 hours and who have their care preferences documented within 48 hours OR documentation as to why this was not done.
- The developer provided a logic model that depicts policies and procedures that support routine communication of documenting care preferences with patients admitted to the ICU. These policies result in effective communication and goal concordant care.

The developer provides the following evidence for this measure:

- SR of the evidence specific to this measure? ☒ Yes ☐ No
- Quality, Quantity, and Consistency of evidence provided? ☐ Yes ☒ No
- Evidence graded? ☐ Yes ☒ No

Summary of prior review in 2016

- In 2016, there was no updated evidence submitted, but the developer did provide two systematic reviews during initial endorsement in 2012. The reviews linked advance care planning to better patient outcomes and provided evidence that patients want to communicate their care preferences to their physicians.
- The Committee referenced evidence from a different measure NQF #1641 *Hospice and Palliative Care – Treatment Preferences* that supports advance care planning and shared decision making.
- In 2016, the Committee noted that the evidence presented does not pertain to the documentation of the care preferences themselves as much as to the importance of care preferences and the discussion around those.

Changes to evidence from last review

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

☐ The developer provided updated evidence for this measure:

Exception to evidence

- N/A

Questions for the Standing Committee:

- *The developer attests that the underlying evidence for the measure has not changed since the last NQF endorsement review. Does the Standing Committee agree that the evidence basis for the measure has not changed and that there is no need for repeated discussion and a vote on evidence?*
- *How strong is the evidence for this relationship?*
- *Is the evidence directly applicable to the process of care being measured?*

Guidance From the Evidence Algorithm

- Not a Health Outcome or PRO (Box 1) -> Process measure based on systematic review (Box 3) -> QQC not provided (Box 4)-> Only 3 studies cited and no Randomized Control Trials (Box 6) -> Low. The highest possible rating is moderate.

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

RATIONALE: A process measure based on two systematic reviews with no QQC summary provided has the highest possible rating of moderate for evidence. In Box 6 of the evidence algorithm, the measure does not meet the described high quality evidence, resulting in a low rating for evidence for this measure.

However, the Committee may choose to apply the evidence discussion and vote from the prior review in 2016, which considered information from the Lorenz, et. al. (2007) article and evidence presented for measure #1641 (treatment preferences measure) in its application of the evidence criterion for this measure. A repeated vote and discussion are not required.

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

Maintenance measures – increased emphasis on gap and variation

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

- Current performance data and/or a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement was not provided.
- In 2012, the developer provided data from four individual studies to demonstrate opportunity for improvement. Across these studies, the sample size ranged from 6 to 349 and measure results ranged from 9% to 63.7%.
- For the current submission, the developer provided the following information:
    - 2013: Sample size=150, measure performance=63.7% .
    - 2010: Sample size=369, measure performance=46%.

Disparities

- The developer provided a 2019 study published in Palliative Medicine that analyzed patient characteristics including demographics, language, type of admission (elective vs. non-elective), SOFA score, whether mechanical ventilation was initiated in first 48 hours of ICU visit and type of ICU, and how they were associated with whether the patient passed this quality indicator or not.
- The study found patients who were female (53.8% vs. 43.4%; P < 0.001), older (84.8 v. 83, P<0.001), had an unelective admission (99.3%% v. 88.7%) and were admitted to medical ICU (64.9% v. 40.0%, p<0.001), were more likely to meet criteria for this quality indicator.

Questions for the Standing Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *If limited disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

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

RATIONALE: For NQF maintenance of endorsement, measure stewards/developers are expected to provide current performance data. Data from the literature can be considered, but were not provided. Little information is provided to determine whether there is still opportunity for improvement in documenting care preferences for ICU patients.

Criteria 2: Scientific Acceptability of Measure Properties

Complex measure evaluated by the Scientific Methods Panel (SMP)? ☐ **Yes** ☒ **No**

Evaluators:

2a. Reliability: Specifications and Testing

For maintenance measures—no change in emphasis—specifications should be evaluated the same as with new measures.

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

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

2a2. Reliability testing demonstrates whether the measure data elements are repeatable and producing the same results a high proportion of the time when assessed in the same population during the same time period, and/or whether the measure score is precise enough to distinguish differences in performance across providers.

Specifications:

- Additional clarification is needed on denominator definitions with definitions for identifying each of the 4 vulnerable populations. Specifically, it is not clear what diagnosis codes are used to define poor prognosis/terminal illness defined as life expectancy of <6 months.

Reliability Testing:

- Reliability testing conducted at the Patient/Encounter Level:
    - The developer calculated a kappa statistic for the 402 records (BIDMC) that were coded by two different individuals. They also tested an algorithm for identifying documentation of care preferences compared to human coding. The kappa statistic between the two human coders for identifying documented care preferences was 0.71. The algorithm had a sensitivity of 93.5% and a specificity of 91.0% for identifying documentation of care preferences.
    - Prior submissions used data from 3 academic medical centers (n=369 (UCLA), 22 (Hopkins), and 150 (VA) patients). Reliability testing was updated using a database of 1,141 patients (MIMIC III) who were admitted to the ICU and survived for 48 hours and died during the admission. Each of the four samples was tested for reliability using inter-rater reliability.

- The developer tested inter-rater reliability for 47 of the 369 patients (UCLA) using re-abstraction of records at the measure level for the denominator and numerator. The denominator level kappa statistic was 0.95 and the numerator level kappa statistic was 0.87.
- The developer tested inter-rater reliability for a 4% sample of charts (Hopkins) using a kappa statistic for 41 quality indicators. The overall denominator kappa was 0.87 and the numerator kappa was 0.86. It is unclear how these 41 quality indicators relate to measure NQF #1626. It is unclear if this kappa can be inferred as reliability for this specific measure.
- The developer tested a 5% abstraction sample and a kappa statistic was presented for the overall measure set (VA), but not at the measure level. The denominator level kappa was 0.92 and the numerator kappa was 0.68. It is unclear if this kappa can be inferred as reliability for this specific measure.

Questions for the Standing Committee regarding reliability:

- *Do you have any concerns that the measure cannot be consistently implemented (i.e., are measure specifications adequate)?*

Guidance From the Reliability Algorithm

The developer provided complete specifications that can be consistently implemented by users (Box 1) > The empirical reliability was tested using statistical tests with the measure as specified (Box 2) > The empirical reliability testing was conducted at the accountable entity level for each level of anayalsis (Box 4) > The developer conducted inter-rater reliability to assess varability (Box 5) > There is moderate confidence that the accountable entity levels are reliable (Box 6b) > Moderate. The highest possible rating is high.

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

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

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

2b2. Validity testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

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

Validity Testing

- Validity testing conducted at the patient/encounter level:
  - The developer provided patient/encounter empirical testing of reliability which also serves as a demonstration of validity at the patient/encounter level. The developer calculated a kappa statistic for the 402 records (BIDMC) that were coded by two different individuals. They also tested an algorithm for identifying documentation of care preferences compared to human coding. The kappa statistic between the two human coders for identifying documented care preferences was 0.71. The algorithm had a sensitivity of 93.5% and a specificity of 91.0% for identifying documentation of care preferences.
- Validity testing conducted at the Accountable Entity Level:
  - A sample of 21,310 older persons from 19 California counties was tested using the UCLA-ACOVE quality indicator set (41 measures) and examining mortality and functional status. The developer did not test NQF #1626 specifically. The developer asserts: "Although validity has

not been tested empirically for this measure alone, the process-outcome link of the set of quality measures including this measure has been tested. Process of care measured using the ACOVE quality indicator set is linked to patient function and survival."

- The developer notes that face validity was conducted and asserts that the measure has face validity and is supported by expert consensus. Demonstration of face validity using NQF criteria requires the description of a systematic and transparent process and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor performance. The degree of consensus and any areas of disagreement must be provided/discussed. The degree of consensus was not provided in this submission.
- The developer hypothesized that medical ICUs will have higher scores than surgical ICUs because palliative care is used less often in surgical services than medical services. A 2020 study (BIDMC) found that Coronary Care Units and Medical ICU units have significantly higher scores compared to surgical ICUs in a generalized linear mixed effects regression model controlling for other patient factors.

Exclusions

- The measure does not use exclusions.

Risk-Adjustment

- The measure is not risk adjusted or stratified.

Meaningful Differences

- Performance varies between the four medical centers: Hopkins 9%, UCLA 46%, VA 63.7%, BIDMC 64.7%. The BIDMC study found variation ranging from 30-75% in ICUs and was able to detect significant differences between medical/coronary care ICUs and other ICUs.
- Patients in the BIDMC study all died during their ICU stay, but after at least 48 hours. However, this measure is intended to be applied to all ICU patients and not limited to those patients who died during the ICU admission.

Missing Data

- The developer states that missing data analysis was not performed for this measure. It is unclear from the developer's response if there is missing data. The developer should be prepared to clarify if there is missing data.

Comparability

- The measure only uses one set of specifications.

Questions for the Standing Committee regarding validity:

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

- *The developer did not provide validity testing on missing data. The Standing Committee should consider asking the developer to provide a rationale for this. Does the Standing Committee have concerns about the lack of missing data information and testing?*
- *The developer attests that additional validity testing was not conducted. Does the Standing Committee agree that the measure is still valid and that there is no need for repeated discussion and a vote on validity?*

Guidance From the Validity Algorithm

Threats to validity including missing data were not empirically assessed (Box 1).

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

All threats to validity relevant to the measure, including missing data, must be empirically assessed.

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

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

- The current method of collection for this measure is through manual chart abstraction. The developers recognize that care preferences have not traditionally lent themselves to electronic data collection and they note a key challenge currently is the time-consuming nature of manual chart abstraction required for this measure.
- The developers are currently testing methods that allow for computer assisted abstractions to improve efficiency. They intend to have the results of this testing in the next year.

Questions for the Standing 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**

Criterion 4: Use and Usability

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

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

4a. Use evaluates the extent to which audiences (e.g., consumers, purchasers, providers, and 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 they are not in use at the time of initial endorsement, then a credible plan for implementation within the specified time frames is provided.

Current uses of the measure

Publicly reported? ☐ Yes ☒ No

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

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

Accountability program details

- The measure is not currently publicly reported and is not used in an accountability program which are both required for maintenance measures following initial endorsement.
- The developer stated that "this is currently being measured in the in a national sample and in an academic cancer center using a computer-assisted abstraction methodology that may be easier to implement for rapid quality improvement and feedback, but results are not ready".
- The developer included mention of a 2017 article published in Health Affairs where an expert panel suggested CMS include this measure in the Medicare Hospital Inpatient Quality Reporting Program.
- Additionally, the measure was referenced in the VA´s Life Sustaining Treatment Decisions Handbook as a quality standard that clinicians should consider adopting.

4a.2. Feedback on the measure provided 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; and (3) This feedback has been considered when changes are incorporated into the measure.

Feedback on the measure provided by those being measured or others

- Overtime the developers have received feedback that the manual abstraction is not effective.
- The measure is currently being tested using a compter assisted abstraction method in a national VA sample and in an academic center to increase efficiency.

***Questions for the Standing 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

RATIONALE: Although the developer has considered efficiency improvements based on feedback for the measure, as a maintenance measure it does not meet NQF criteria requirements by being neither used in at least one accountability application (within three years after initial endorsement), nor publicly reported (within six years after initial endorsement).

4b. Usability (4b1. Improvement; 4b2. Benefits of measure)

4b. Usability evaluates the extent to which audiences (e.g., consumers, purchasers, providers, and 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

- Data to support progress on improvement or trends trends in performance results were not provided. The developer stated that as organizations strive to match patient preferences with care received, measures such as this one will be needed to evaluate their success.

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

Unexpected findings (positive or negative) during implementation

- The developer is not aware of any unitended consequences or unexpected benefits.

Potential harms

- The developer is not aware of any potential harms.

***Questions for the Standing 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: ☐ **High** ☐ **Moderate** ☐ **Low** ☒ **Insufficient**

RATIONALE: There are no data on current performance or improvement in performance over time.

Criterion 5: Related and Competing Measures

Related Measures

- NQF #0326 Advance Care Plan
- NQF #1641 Hospice and Palliative Care-Treatment Preferences
- NQF #3665 Ambulatory Palliative Care Patients' Experience of Feeling Heard and Understood

Harmonization

The developer indicated that the measures are harmonized to the extent possible, however, if the measure is recommended for continued endorsement, the Committee may be asked to provide recommendations for harmonizing the measures.

Criteria 1: Importance to Measure and Report

1a. Evidence

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

---

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

[Response Begins]

No

[Response Ends]

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

Current Submission:

Updated evidence information here.

Previous (Year) 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]

Goal	Measure	Relevant NQF Preferred Practices	Structure	Process	Outcome
Effective Communication and goal concordant care	Patients admitted to ICU who have care preferences documented	#7,18	Policies and procedures that support routine communication to ensure patients preferences guide care upon admission to the ICU	Timely goals of care communication upon admission to the ICU	Patient receives goal concordant care

[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)

[Other (specify) Please Explain]

There is no clinical trial directly linking the care process in this measure with outcomes. However, elicitation of preferences is one important step in advance care planning and/or serious illness communication to ensure matching care with patient goals. The ACOVE expert panel and the ASSIST expert panel, based on a clinically informed understanding of the medical literature identified this care process important for providing care to seriously ill patients receiving intensive care in the hospital.

Lorenz KA, Rosenfeld K, Wenger N. Quality indicators for palliative and end-of-life care in vulnerable elders. J Am Geriatr Soc. 2007;55 Suppl 2:S318-26.

Walling A, Lorenz KA, Dy SM, et al. Evidence-based recommendations for information and care planning in cancer care. J Clin Oncol 2008;26(23):3896-3902.

[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]

see above

[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]

see above

[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]

There is no clinical trial directly linking the care process in this measure with outcomes. However, elicitation of preferences is one important step in advance care planning and/or serious illness communication to ensure matching care with patient goals. The ACOVE expert panel and the ASSIST expert panel, based on a clinically informed understanding of the medical literature identified this care process important for providing care to seriously ill patients receiving intensive care in the hospital.

Lorenz KA, Rosenfeld K, Wenger N. Quality indicators for palliative and end-of-life care in vulnerable elders. J Am Geriatr Soc. 2007;55 Suppl 2:S318-26.

Walling A, Lorenz KA, Dy SM, et al. Evidence-based recommendations for information and care planning in cancer care. J Clin Oncol 2008;26(23):3896-3902.

[Response Ends]

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

[Response Begins]

Elicitation of preferences is one important step in advance care planning and/or serious illness communication to ensure matching care with patient goals, especially upon admission to the ICU when intensive and potentially burdensome treatments are often offered. The ACOVE expert panel and the ASSIST expert panel, based on a clinically informed understanding of the medical literature identified this care process important for providing care to seriously ill patients receiving intensive care in the hospital. A recent measure scan and synthesis of palliative and end-of-life process quality measures for advanced cancer identified 59 advance care planning measures across 12 sources or measure sets and was the domain under which the most number of palliative care measures were identified, highlighting their importance. Experts have identified admission to the ICU as an important time to ensure that goals of care are addressed.

Lorenz KA, Rosenfeld K, Wenger N. Quality indicators for palliative and end-of-life care in vulnerable elders. J Am Geriatr Soc. 2007;55 Suppl 2:S318-26.

Walling A, Lorenz KA, Dy SM, et al. Evidence-based recommendations for information and care planning in cancer care. J Clin Oncol 2008;26(23):3896-3902.

O'Hanlon CE, Lindvall C, Lorenz K et al. Measure Scan and Synthesis of Palliative and End-of-Life Process Quality Measures for Advanced Cancer. JCO Oncology Practice. 2020, 17:e140-e148.

[Response Ends]

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

[Response Begins]

Please note that the below direct quotes are taken from the methods section of the referenced papers:

The original literature search is detailed in Lorenz et al from 2007: "A total of 299 articles were considered in this review: 74 identified using a Web search, 120 through reference mining, 50 through the Assessing Care of Vulnerable Elders (ACOVE)-3 literature searches, and 55 from reference mining the ACOVE-1 monograph."

Another search published in 2008 focused on cancer: "Articles were identified through MEDLINE, the Cochrane Database of Systematic Reviews, Cochrane Database of Abstracts of Reviews of Effects, and the Cochrane Register of Clinical Trials (1996 to 2006). We limited searches to English language, using terms combined with each topic area (Appendix, online only). To enrich the 10-year literature review, we searched for earlier citations selectively. We searched broadly and later included studies that included patients other than oncology patients because there are few formal trials of ACP specific to oncology.[35] It is also important to note that much of the information underlying excellent practice in this area is theoretical literature based on ethical and professional principles rather than empiric experience."

The literature reviews were reviewed by expert panels in both cases.

[Response Ends]

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

[Response Begins]

Lorenz KA, Rosenfeld K, Wenger N. Quality indicators for palliative and end-of-life care in vulnerable elders. J Am Geriatr Soc. 2007;55 Suppl 2:S318-26.

Walling A, Lorenz KA, Dy SM, et al. Evidence-based recommendations for information and care planning in cancer care. J Clin Oncol 2008;26(23):3896-3902.

O'Hanlon CE, Lindvall C, Lorenz K et al. Measure Scan and Synthesis of Palliative and End-of-Life Process Quality Measures for Advanced Cancer. JCO Oncology Practice. 2020, 17:e140-e148.

[Response Ends]

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

[Response Begins]

The aim of this measure is to assist healthcare providers in providing care that is consistent with patient preferences.

[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]

N, % measure performance
Assessing Symptoms Side Effects and Indicators of Supportive Treatment (ASSIST): Walling 2013, Inpatients in a national VA sample, N=150, 63.7%

N, % measure performance
Assessing Care of Vulnerable Elders (ACOVE3)(Walling 2010): Inpatient decedents, N=369, 46%

Assessing Symptoms Side Effects and Indicators of Supportive Treatment (ASSIST) (Dy 2010): Inpatient decedents, N=22, 9%

ACOVE (Wenger 2003): Vulnerable elders, N=6, 17%

[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]

Walling AM, et al. The Quality of Supportive Cancer Care in the Veterans Affairs Health System and Targets for Improvement. JAMA IM. 2013:173:2071-2079.

Dy SM, Asch SM, Lorenz KA, et al. Quality of end-of-life care for patients with advanced cancer in an academic medical center. J Palliat Med 2011;14(4)"451-457

Walling AM, Asch AM, Lorenz KA, et al. The quality of care provided to hospitalized patients at the end of life. Arch Intern Med 2010;170(12):1057-1063

Wenger NS, Solomon DH, Roth CP, et al. The quality of medical care provided to vulnerable community-dwelling older patients. Ann Intern Med 2003;139(():740-E759

[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]

See 2a.08

[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]

See 2a.08

[Response Ends]

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

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

---

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

[Response Begins]

No

[Response Ends]

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

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

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

[Response Begins]

N/A

[Response Ends]

sp.01. Provide the measure title.

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

[Response Begins]

Patients Admitted to ICU who Have Care Preferences Documented

[Response Ends]

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

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

[Response Begins]

Percentage of vulnerable adults admitted to ICU who survive at least 48 hours who have their care preferences documented within 48 hours OR documentation as to why this was not done.

[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]

Cancer

Critical Care

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)

[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]

Inpatient/Hospital

[Response Ends]

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

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

[Response Begins]

none available

[Response Ends]

sp.12. 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.13. State the numerator.

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

DO NOT include the rationale for the measure.

[Response Begins]

Patients in the denominator who had their care preferences documented within 48 hours of ICU admission or have documentation of why this was not done.

[Response Ends]

sp.14. 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]

Edits indicated by [brackets]

atients whose medical record includes documentation of care preferences within 48 hours of admission to ICU. Care preferences may include any of the following:

Code status, preferences for general aggressiveness of care, mechanical ventilation, hemodialysis, transfusion, or permanent feeding tube, OR

Documentation that a care preference discussion was attempted and/or reason why it was not done

[Simply having an advance directive or other advance care planning document or POLST in the medical record does not satisfy this criterion. However, a notation in the record during the allotted time period referring to preferences or decisions within such a document satisfies this requirement.]

[Response Ends]

sp.15. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]

All vulnerable adults admitted to ICU who survive at least 48 hours after ICU admission.

[Response Ends]

sp.16. 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]

All vulnerable adults admitted to ICU who survive at least 48 hours after ICU admission.
"Vulnerable" is defined as any of the following:
- >74 years of age
- Vulnerable Elder Survey-13 (VES-13) score >2 (Saliba 2001)
- Poor prognosis/terminal illness defined as life expectancy of <6 months
- Stage IV cancer

[Response Ends]

sp.17. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]

None

[Response Ends]

sp.18. Provide details needed to calculate the denominator exclusions.

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

[Response Begins]

N/A

[Response Ends]

sp.19. Provide all information required to stratify the measure results, if necessary.

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

[Response Begins]

N/A

[Response Ends]

sp.20. Is this measure adjusted for socioeconomic status (SES)?

[Response Begins]

No

[Response Ends]

sp.21. 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.22. Select the most relevant type of score.

Attachment: If available, please provide a sample report.

[Response Begins]

Rate/proportion

[Response Ends]

sp.23. 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.24. 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. Identify all vulnerable adults admitted to ICU who survive at least 48 hours after ICU admission
2. Examine the medical record for evidence of a statement of patient care preferences OR attempt to elicit these or other reason why this was not done within 48 hours of ICU admission.

[Response Ends]

sp.27. If measure testing is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.

Examples of samples used for testing:

- *Testing may be conducted on a sample of the accountable entities (e.g., hospital, physician). The analytic unit specified for the particular measure (e.g., physician, hospital, home health agency) determines the sampling strategy for scientific acceptability testing.*
- *The sample should represent the variety of entities whose performance will be measured. The 2010 Measure Testing Task Force recognized that the samples used for reliability and validity testing often have limited generalizability because measured entities volunteer to participate. Ideally, however, all types of entities whose performance will be measured should be included in reliability and validity testing.*
- *The sample should include adequate numbers of units of measurement and adequate numbers of patients to answer the specific reliability or validity question with the chosen statistical method.*
- *When possible, units of measurement and patients within units should be randomly selected.*

[Response Begins]

No sampling.

[Response Ends]

sp.30. Select only the data sources for which the measure is specified.

[Response Begins]

Electronic Health Records

Paper Medical Records

[Response Ends]

sp.31. 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]

Medical record abstraction tool

[Response Ends]

sp.32. Provide the data collection instrument.

[Response Begins]

Available in attached appendix in Question 1 of the Additional Section

[Response Ends]

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

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

Current Submission:

Updated testing information here.

Previous Submission:

Testing from the previous submission here.

[Response Begins]

Yes

[Response Ends]

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

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

Current Submission:

Updated testing information here.

Previous Submission:

Testing from the previous submission here.

[Response Begins]

No

[Response Ends]

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

[Response Begins]

No

[Response Ends]

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

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

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

[Response Begins]

No additional risk adjustment analysis included

[Response Ends]

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

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

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

2a. Reliability testing demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.

2b1. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure;

AND

If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

2b3. For outcome measures and other measures when indicated (e.g., resource use):

- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; 14,15 and has demonstrated adequate discrimination and calibration
- 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 Scientific Acceptability sections. For example:

Current Submission:

Updated testing information here.

Previous (Year) Submission:

Testing from the previous submission here.

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

[Response Begins]

Electronic Health Records

Paper Medical Records

[Response Ends]

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

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

[Response Begins]

An existing data set was not used for the evaluation of this measure in the traditional sense. In prior NQF endorsement submissions, we report on chart abstractions that occurred as a result of three quality measurement efforts that have been previously reported to NQF from Johns Hopkins Sidney Kimmel Comprehensive Cancer Center, the Veterans Health Administration (national study), and UCLA Health.

For this submission, we are reporting on a chart abstraction that was completed using a publicly available ICU database, Multi Parameter Intelligent Monitoring of Intensive Care (MIMIC) III, developed by Massachusetts Institute of Technology (MIT) and Beth Israel Deaconess Medical Center (BIDMC). This data base is unique in that it includes free text medical record notes for patients admitted to the ICU. We looked at ICU admissions in MIMIC III between 2008-2012. The denominator was 1141 inpatients admitted to an ICU between 2008-2012 that are 75 years or older.

References

Johnson AE, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016;3:160035.

Chan A, Chien I, Moseley E, et al. Deep learning algorithms to identify documentation of serious illness conversations during intensive care admissions. 2019;33:187-196.

[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]

Hopkins study (ASSIST): 2003-2005 diagnosis with 15 month follow-up; VA National Study (ASSIST): 2008 diagnosis with 3 year follow up; UCLA study (ACOVE3): decedents with hospitalization between April 2005 and April 2006; MIMIC-III study: ICU admissions between 2008-2012

[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]

UCLA Health (ACOVE 3 study, Walling 2010): inpatient decedents at an academic medical center (n=369)

Johns Hopkins Sidney Kimmel Comprehensive Cancer Center (ASSIST study, Dy 2010, 2011): patients with advanced cancer who died between 2-15 months after diagnosis (diagnosed between 2003-2005), n=22

Veterans Health Administration (national study, Walling 2013): inpatient in a national VA sample, n=150

MIMIC III data base (Chan, 2019): inpatients admitted to an ICU between 2008-2012 that are 75 years or older (n=1141)

[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]

UCLA Health (ACOVE 3 study, Walling 2010): inpatient decedents at an academic medical center,n=369. This study identified all patients 18 years or older who died during admission to one medical center between April 2005 and April 2006 following a hospitalization of at least 3 days. Decedents were on average 62.3 years, 47% women, 62% White, 16% Hispanic, 6% African American, 10% Asian and 5% Other. Fifty-six percent of patients had end stage disease on admission (21% advanced cancer, 11% end stage pulmonary disease, 6% end stage heart failure, 16% end stage liver disease, 9% end stage renal disease, 1% AIDS, and 1% dementia. The total sample is 496 patients and 369 of the patients were included in this measure because they were admitted to the ICU and met denominator criteria.

Johns Hopkins Sidney Kimmel Comprehensive Cancer Center (ASSIST study, Dy 2010, 2011): patients with advanced cancer who died between 2-15 months after diagnosis (diagnosed between 2003-2005), n=22. The study used the hospital cancer registry to sample patients diagnosed with stage IV (and stage IIIB for lung) lung, pancreatic, breast, and colorectal in 2003–2005. The sample was also limited to patients who died between 2 and 15 months after diagnosis and had at least three outpatient visits at Johns Hopkins or at least two outpatient visits and a cancer-related hospitalization between 3–30 days in length after the diagnosis. The overall sample included 238 patients and were on average 65.9 years, 53% female, 63% Caucasion, 32% African American, 5% Other. Three percent of patients had breast cancer, 10% had colorectal cancer, 55% had lung cancer, and 32% had pancreatic cancer. The total sample included 238 patients and 22 of the patients were included in this measure because they were admitted to the ICU and met demonimator criteria.

Veterans Health Administration (national study, Walling 2013): inpatient in a national VA sample, n=150. This study selected a national cohort of Veterans with stage IV pancreatic, lung and colorectal cancers to represent comment solid tumors with varying prognostic and clinical features. The sampling frame included the following patients diagnosed from the 2008 VA Comprehensive Cancer registry: 424 Veterans with pancreatic cancer, 3184 with lung, and 628 with colorectal cancer. Veterans were randomly sampled from each cancer in equal proportions for a total of 719 Veterans. These Veterans were on average 66.2 years old, 97.2% male, 74.3% White, 19.9% Black, 12% Asian, 30% Hispanic. The primary cancer for Veterans was 37% colorectal, 33.2% lung and 29.8% pancreatic. The total sample is 719 patients and 150 of the patients were included in this measure tbecausehey were admitted to the ICU and met denominator criteria.

MIMIC III data base (Chan, 2019): inpatients admitted to an ICU between 2008-2012 that are 75 years or older (n=1141). The study population included adult patients who were admitted to the medical, surgical, coronary care or cardiac surgery ICU at Beth Israel Deaconess Medical Center and survived 48 hours, and died during the admission. These patients had a mean age of 84.3 years, 50.8% were female, and 6.7% were African American. The top three ICU admitting diagnoses were pneumonia, congestive heart failure and sepsis.

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

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

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

[Response Begins]

For the Chan 2019 study published in Palliative Medicine, there is an associated article published in Journal of Pain and Symptom Management. This study included 1350 ICU admissions for 1141 patients from the same data set and looked at patient level factors and how they were associated with whether or not the patient passed this quality indicator. Patient characteristics that were studied included demographics, language, type of admission (elective vs. non-elective), SOFA score, whether mechanical ventilation was initiated in first 48 hours of ICU visit and type of ICU. Table 1 of the paper shows that patients who were female (53.8% vs. 43.4%; *P* < 0.001), older (84.8 v. 83, P<0.001), had an unelective admission (99.3%% v. 88.7%), and were admitted to medical ICU (64.9% v. 40.0%, p<0.001) were more likely to meet criteria for this quality indicator. Whereas, patients who were married (38.3% vs. 47.6%, p<0.001) and who received mechanical ventilation within 48 hours (33.3% v. 43.0%, p<0.001) were less likely to meet criteria for this quality indicator. In adjusted analysis, rates of care preference documentation were higher for older patients, females, unmarried patients, nonelective admissions, patients who did not receive mechanical ventilation in the first 48 hours, and admissions to the medical vs. the cardiac or surgical ICUs.

Udelsman BV, et al. Deep Natural Language Processing Identifies Variation in Care Preference Documentation. Journal of Pain and Symptom Management. 2020. 59:1186-1194.

[Response Ends]

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.09 check patient or encounter-level data; in 2a.010 enter "see validity testing section of data elements"; and enter "N/A" for 2a.11 and 2a.12.

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

Choose one or both levels.

[Response Begins]

Patient or Encounter-Level (e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements)

[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]

ACOVE3 (Walling 2010). For this study there were 369 inpatient decedents that met criteria for the denominator and the pass rate was 46%. We tested inter-rater reliability using a kappa statistic for a random sample of 47 re-abstraction records at the measure level for the denominator and the numerator.

ASSIST (Dy 2010, 2011) For this study there were 22 inpatient decedents that met criteria for the denominator and the pass rate was 9%. The team tested 78 quality indicators and determined that 41 met criteria for feasibility, reliability and validity. To evaluate reliability, the team calculated inter-rater reliability for a 4% sample of charts using a kappa statistic. The minimum criteria for inter-rater reliability to be retained was kappa >0.8 for eligibility (denominator) or <0.6 for adherence (numerator).

ASSIST (Walling 2013) For this study there were 150 patients that met criteria for the denominator and the pass rate was 63.7%. A 5% reabstraction sample was studied for inter-rater reliability and the kappa statistic is presented for the overall measure set, not at the measure level.

MIMIC III data base (Chan, 2019): For this study 1141 patients had 1350 ICU admissions and the measure scores were presented at the event level. Care preference documentation within 48 hours of ICU admission occurred in 64.7% of all

ICU admissions. For 402 patients that were each coded by two different humans, we calculated the kappa statistic for abstracting documentation of care preferences. We also calculated sensitivity and specificity for the performance of a deep learning algorithm in identifying documentation of care performance compared to a gold standard human coding that was dual coded with a tie break with a third coder.

Walling AM, Asch SM, Lorenz KA, et al. The quality of care provided to hospitalized patients at the end of life. Arch Intern Med. 2010;170:1057-63.

Dy SM, Asch AM, Lorenz KA, et al. Quality of end-of-life care for patients with advanced cancer in an academic medical center. J Pall Med 2011;14(4):451-59.

Dy SM, Lorenz KA, O'Neill S, et al. Cancer quality-ASSIST supportive oncology quality indicator set. Feasibility, reliability, and validity. Cancer 2010;116:3267-75.

Chan A, Chien I, Moseley E, et al. Deep learning algorithms to identify documentation of serious illness conversations during intensive care admissions. 2019;33:187-196.

[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]

ACOVE3 (Walling 2010). For this study there were 369 inpatient decedents that met criteria for the denominator and the pass rate was 46%. We tested inter-rater reliability using a kappa statistic for a random sample of 47 re-abstraction records. We calculated the inter-rater reliability at the measure level for the denominator and the numerator. The measure level kappa statistic for the denominator or the eligibility kappa was 0.95, whereas the measure level kappa statistic for the numerator or the specified care kappa was 0.87.

ASSIST (Dy 2010, 2011) For this study there were 22 inpatient decedents that met criteria for the denominator and the pass rate was 9%. Overall, the team tested 78 quality indicators and determined that 41 met criteria for feasibility, reliability and validity. To evaluate reliability, the team calculated inter-rater reliability for a 4% sample of charts using a kappa statistic. The minimum criteria for inter-rater reliability to be retained was kappa >0.8 for eligibility (denominator) or <0.6 for adherence (numerator). The overall eligibility or denominator kappa for all 41 retained measures was 0.87. The overall specified kappa (numerator) was 0.86.

ASSIST (Walling 2013)For this study there were 150 patients that met criteria for the denominator and the pass rate was 63.7%. A 5% reabstraction sample was studied for inter-rater reliability and the kappa statistic is presented for the overall measure set, not at the measure level. The overall eligibility (denominator) kappa=0.92. The overall specified (numerator) kappa=0.68.

MIMIC III data base (Chan, 2019): For this study 1141 patients had 1350 ICU admissions and the measure scores were presented at the event level. Care preference documentation within 48 hours of ICU admission occurred in 64.7% of all ICU admissions. For 402 patients that were each coded by two different humans, we calculated the inter-rater reliability using a kappa statistic for abstracting documentation of care preferences. We also calculated sensitivity and specificity for the performance of a deep learning algorithm in identifying documentation of care performance compared to a gold standard human coding that was dual coded with a tie break with a third coder. The kappa statistic between the two human coders for identifying documented care preferences was 0.71. The deep learning algorithm had a sensitivity of 93.5% and a specificity of 91.0% for identifying documentation of care preferences within 48 hours of ICU admission compared to a triple coded human gold standard.

Walling AM, Asch SM, Lorenz KA, et al. The quality of care provided to hospitalized patients at the end of life. Arch Intern Med. 2010;170:1057-63.

Dy SM, Asch AM, Lorenz KA, et al. Quality of end-of-life care for patients with advanced cancer in an academic medical center. J Pall Med 2011;14(4):451-59.

Dy SM, Lorenz KA, O'Neill S, et al. Cancer quality-ASSIST supportive oncology quality indicator set. Feasibility, reliability, and validity. Cancer 2010;116:3267-75.

Chan A, Chien I, Moseley E, et al. Deep learning algorithms to identify documentation of serious illness conversations during intensive care admissions. 2019;33:187-196.

[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]

Statistics that measure inter-rater reliability of this measure show that it can be reliably abstracted from chart abstraction. Recent research shows that this measure was also able to be abstracted from an electronic medical record using advanced computer-assisted abstraction techniques with high sensitivity and specificity compared to a triple coded human gold standard.

[Response Ends]

2b. Validity

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

[Response Begins]

Accountable Entity Level (e.g. hospitals, clinicians)

Empirical validity testing

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

[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]

Although validity has not been tested empirically for this measure alone, the process - outcome link of the set of quality measures including this measure has been tested. Process of care measured using the ACOVE quality indicator set is related to two important outcomes in vulnerable elders and persons 75 years and older: mortality and functional status. In 372 vulnerable elders there was a graded positive relationship between quality score and 3 -year survival. After

adjustment for sex, health status, and health service use, quality score was not associated with mortality for the first 500 days, but a higher quality score was associated with lower mortality after 500 days (hazard ratio, 0.64 [95% CI, 0.49 to 0.84] for a 10% higher quality score).(Higashi 2005) Using an administrative data implementation of a subset of these measures, 21,310 older persons from 19 California counties had their quality of care measured and outcomes followed over the next year. After accounting for number of measures triggered, baseline function and other covariates, better quality was associated with better function at follow-up. Ten percent better quality was associated at follow-up with 0.21 lower ADL need score [95% confidence interval (CI), 0.25-0.17], 0.022 lower IADL need score (95% CI, 0.032-0.013), and lower odds of death (0.91; 95% CI, 0.89 to 0.93).(Zingmond 2011) Validity of the process-outcome link was explicitly evaluated by the ACOVE, ACOVE3, and ASSIST expert panels that reviewed the relevant literature and used a modified Delphi panel of voting on the validity of the measure. (Shekelle 2001; Wenger 2007; Lorenz 2009) Although validity has not been tested empirically for this measure alone, the process-outcome link of the set of quality measures including this measure has been tested. Process of care measured using the ACOVE quality indicator set is linked to patient function and survival. (Higashi 2007).

Face validity was tested in the panels for each measure including this one as well as the strength of the process-outcome link. Data has shown an association between timely goals of care conversations and improved quality of end of life care. Despite limited evidence, the importance of goals of care conversations, the provision of tailored prognostic information, as well as information provision to ensure informed decision making was also highlighted in recent 2017 ASCO guidelines for patient-physician communication largely based on expert consensus.

Lastly, accountable entity level or score level testing was tested in the Udelsman 2020 publication. Evidence suggests that palliative care (inclusive of goals of care conversations) is used less often in surgical services than medical services (Olmsted 2014 and Rodriguez 2015). Therefore we would hypothesize that medical ICU's would have higher scores than surgical ICU's on this quality measures. As expected, the MIMIC III study showed variation in pass rate of this indicator by the type of ICU ranging from 30%-75% with Coronary Care Units and Medical ICU units having higher scores compared to surgical ICUs. These differences were found to be statistically significant in a generalized linear mixed effects regression model controlling for other patient factors.

Referemces:

1. Higashi T, Shekelle PG, Adams J, et al. Quality of care is associated with survival in vulnerable older patients. Ann Intern Med 2005;143:274-281
2. Lorenz KA, Dy SM, Naeim A, et al. Quality measures for supportive cancer care: the cancer quality-ASSIST project. J Pain Symptom Manage 2009;37(6):943-964
3. Shekelle PG, MacLean CH, Morton SC, et al. Assessing care of vulnerable elders: Methods for developing quality indicators. Ann Intern Med 2001;135:647-652
4. Wenger NW, Roth CP, Shekelle P, et al. Introduction to the assessing care of vulnerable elders-3 quality indicator measurement set. J Am Geriatr Soc 2007;55:S247-S252
5. Zingmond DS, Ettner SL, Wilber KH, et al. Association of claims-based quality of care measures with outcomes among community-dwelling vulnerable elders. Med Care 2011;49:553-559
6. Wright AA, Zhang B, Ray A, et al. Associations between end-of-life discussions, patient mental health, medical care near death, and caregiver bereavement adjustment. JAMA. 2008;300:1665-73.
7. Gilligan T, Coyle N, Frankel R et al. Patient-Clinician Communication: American Society of Clinical Oncology Consensus Guideline. Journal of Clinical Oncology. 2017:35:3618-3634.
8. Chan A, Chien I, Moseley E, et al. Deep learning algorithms to identify documentation of serious illness conversations during intensive care admissions. 2019;33:187-196.
9. Udelsman BV, et al. Deep Natural Language Processing Identifies Variation in Care Preference Documentation. Journal of Pain and Symptom Management. 2020. 59:1186-1194.
10. Olmsted CL, Johnson AM, Kaboli P, et al. Use of palliative care and hospice among surgical and medical specialties in the Veterans Health Administration. *JAMA Surg* 2014;149:1169–1175.

11. Rodriguez R, Marr L, Rajput A, et al. Utilization of palliative care consultation service by surgical services. *Ann Palliat Med* 2015;4:194–199.

[Response Ends]

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

Examples may include correlations or t-test results.

[Response Begins]

As mentioned above, accountable entity level or score level testing was tested in the Udelsman 2020 publication. Evidence suggests that palliative care (inclusive of goals of care conversations) is used less often in surgical services than medical services (Olmsted 2014 and Rodriguez 2015). Therefore we would hypothesize that medical ICU's would have higher scores than surgical ICU's on this quality measures. As expected, the MIMIC III study showed variation in pass rate of this indicator by the type of ICU ranging from 30%-75% with Coronary Care Units and Medical ICU units having higher scores compared to surgical ICUs. These differences were found to be statistically significant in a generalized linear mixed effects regression model controlling for other patient factors.

Chan A, Chien I, Moseley E, et al. Deep learning algorithms to identify documentation of serious illness conversations during intensive care admissions. 2019;33:187-196.

Udelsman BV, et al. Deep Natural Language Processing Identifies Variation in Care Preference Documentation. Journal of Pain and Symptom Management. 2020. 59:1186-1194.

Olmsted CL, Johnson AM, Kaboli P, et al. Use of palliative care and hospice among surgical and medical specialties in the Veterans Health Administration. *JAMA Surg* 2014;149:1169–1175.

Rodriguez R, Marr L, Rajput A, et al. Utilization of palliative care consultation service by surgical services. *Ann Palliat Med* 2015;4:194–199.

[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]

This measure has face validity and is supported by evidence and expert consensus. Further, the accountable entity level or score level validity is supported by the data published in the Udelsman 2020 publication that shows variation in pass rate of this indicator by the type of ICU ranging from 30%-75% with Coronary Care Units and Medical ICU units having higher scores compared to ICUs. These differences align with prior data that shows that palliative care (inclusive of goals of care conversations) is used less often in surgery compared to medical care and the hypothesis that medical ICUs would have higher scores than surgical ICUs.

[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]

The awareness of patient preferences is vital to facilitate matching end-of-life care with that which the patient would want. Failure to attempt to elicit patient preferences, if unknown, when a patient is in the ICU is significant. Performance is generally low (Hopkins 9%, UCLA 46%, VA national 63.7%, MIMI-III 64.7%. This MIMIC III study also showed variation in pass rate of this indicator by the type of ICU ranging from 30%-75% with Coronary Care Units and Medical ICU units having higher scores compared to surgical ICUs.

[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]

As noted in 2b4.1 above, the MIMIC III study showed variation in pass rate of this indicator by the type of ICU ranging from 30%-75% with Coronary Care Units and Medical ICU units having higher scores compared to surgical ICUs.

[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]

This measure can detect statistical and clinically meaningful differences that can guide quality improvement.

[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]

N/A

[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

[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

[Response Ends]

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

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

[Response Begins]

No, there is only one set of specifications for this measure

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

Examples may include correlation, and/or rank order.

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

[Response Begins]

N/A or no exclusions

[Response Ends]

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

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

[Response Begins]

N/A

[Response Ends]

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

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

[Response Begins]

N/A

[Response Ends]

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

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

[Response Begins]

N/A

[Response Ends]

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

[Response Begins]

No risk adjustment or stratification

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

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

[Response Begins]

N/A, this is a process measure

[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)

[Other (specify) Please Explain]

N/A-process measure

[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

[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

[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

[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

[Response Ends]

2b.27. Provide risk model discrimination statistics.

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

[Response Begins]

N/A

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

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

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

[Response Begins]

N/A

[Response Ends]

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

[Response Begins]

N/A

[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

[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

[Response Ends]

Criterion 3. Feasibility

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

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

[Response Begins]

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

[Response Ends]

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

In other words, indicate whether data elements that are needed to compute the performance measure score are in defined, computer-readable fields.

[Response Begins]

No data elements are in defined fields in electronic sources

[Response Ends]

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

[Response Begins]

Traditionally, most preference information and discussions by their nature, do not lend themselves to electronic data capture. This is true for other aspects of geriatric care as well. (MacLean 2006) However, the data elements are discrete and could be delineated in an EHR. However, we are currently testing a new methodology to use computer assisted abstraction to efficiently abstract this quality measure using EHR data.

MacLean GH, Louie R, Shekelle PG, et al. Comparison of administrative data and medical records to measure quality of medical care provided to vulnerable older patients. Med Care 2006;44(2):141-148

[Response Ends]

3.04. Describe any efforts to develop an eCQM.

[Response Begins]

We are currently testing methods to allow for computer assisted abstraction for this quality measure which would help with efficiency of abstraction. Manual chart abstraction can be time consuming and often not practical for rapid quality improvement.

[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]

A key challenge for this measure is the time consuming nature of manual chart abstraction. We are currently overcoming this methodological challenge by applying computer assisted abstraction methods and should have these results in the next year.

[Response Ends]

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

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

Attach the fee schedule here, if applicable.

[Response Begins]

N/A

[Response Ends]

Criterion 4: Use and Usability

4a. 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]

Quality Improvement (Internal to the specific organization)

[Quality Improvement (Internal to the specific organization) Please Explain]

This measure is referenced in the VA´s Life Sustaining Treatment Decisions Handbook 1004.02 in section related to ICU admissions. It is not being used to mandate a time frame for conversations, but the metric is referenced as a quality standard that clinicians should consider adopting.

Reference: VHA Handbook 1004.03: https://www.ethics.va.gov/docs/policy/VHA_Handbook_1004_03_LST.pdf, LAST Accessed 11/7/2022.

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]

Quality Improvement (internal to the specific organization)

[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]

While the measure is not currently publicly reported or used in an accountability program, it has been used in VA National Policy as follows:

This measure is referenced in the VA´s Life Sustaining Treatment Decisions Handbook 1004.02 in section related to ICU admissions. It is not being used to mandate a time frame for conversations, but the metric is referenced as a quality standard that clinicians should consider adopting.

Reference: VHA Handbook 1004.03: https://www.ethics.va.gov/docs/policy/VHA_Handbook_1004_03_LST.pdf, LAST Accessed 11/7/2022.

Manual chart abstraction is time consuming and we are currently testing an approach to more efficiently abstract this measure using computer assisted abstraction methods.

Further, in 2017, an article published in Health Affairs stated that an expert panel suggested that CMS should add this measure into the Medicare Hospital Inpatient Quality Reporting Program. New methods currently being evaluated to improve efficiency of abstraction using chart abstraction methods will support this.

Corrigan J, Rising J, Valuck T. Building Additional Serious Illness Measures into Medicare Programs. Health Affairs. 2017: https://www.healthaffairs.org/do/10.1377/forefront.20170525.060256.

[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]

In 2017, an article published in Health Affairs stated that an expert panel suggested that CMS should add this measure into the Medicare Hospital Inpatient Quality Reporting Program. New methods currently being evaluated to improve efficiency of abstraction using chart abstraction methods will support this.

Corrigan J, Rising J, Valuck T. Building Additional Serious Illness Measures into Medicare Programs. Health Affairs. 2017: https://www.healthaffairs.org/do/10.1377/forefront.20170525.060256.

[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 is currently being measured now within the VA in a national sample and in an academic cancer center using a computer-assisted abstraction methodology that may be easier to implement for rapid quality improvement and feedback, but results are not ready.

[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]

This is currently being measured now within the VA in a national sample and in an academic cancer center using a computer-assisted abstraction methodology that may be easier to implement for rapid quality improvement and feedback, but results are not ready.

[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]

Over time, we have received feedback that manual abstraction for quality improvement is not efficient and this is also a concern of our research team. We are currently testing the implementation of a computer-assisted abstraction methodology in a national VA sample and at an academic cancer center that would improve the ability to efficiently use this quality measure.

[Response Ends]

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

[Response Begins]

N/A see above

[Response Ends]

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

[Response Begins]

Over time, we have received feedback that manual abstraction for quality improvement is not efficient and this is also a concern of our research team. We are currently testing the implementation of a computer-assisted abstraction methodology in a national VA sample and at an academic cancer center that would improve the ability to efficiently use this quality measure.

[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]

The measure itself has not been modified, but new methods will make the implementation of the measure more efficient.

[Response Ends]

4b. Usability

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

[Response Begins]

As organizations strive to improve advance care planning to match patient preferences with care received, measures such as this will be needed to measure their success.

[Response Ends]

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

[Response Begins]

Documentation of patient preferences or an attempt to elicit them is not a care process that is likely to produce unintended consequences. We are not aware of unintended consequences.

[Response Ends]

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

[Response Begins]

We are not aware of unexpected benefits. Expected benefits would include the encouragement of ensuring high quality communication with patients upon admission to an ICU.

[Response Ends]

Criterion 5: Related and Competing Measures

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

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

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

(Can search and select measures.)

[Response Begins]

1641: Hospice and Palliative Care – Treatment Preferences

3665: Ambulatory Palliative Care Patients' Experience of Feeling Heard and Understood

0326: Advance Care Plan

[Response Ends]

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

(Can search and select measures.)

[Response Begins]

[Response Ends]

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

[Response Begins]

Please see a recent Measure Scan: O'Hanlon Claire E, et al. Measure Scan and Synthesis of Palliative and End-of-Life Process Quality Measures for Advanced Cancer. JCO Oncology Practice. 2020.17:e140-e148.

[Response Ends]

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

[Response Begins]

Yes

[Response Ends]

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

[Response Begins]

n/a

[Response Ends]

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

Provide analyses when possible.

[Response Begins]

This measure was part of the National Palliative Care Research Center (NPCRC) Key Palliative Measures Bundle during the original submission. At that time, a NPCRC cover letter and table of bundle measures for description of the selection and harmonization of the Key Palliative Measures Bundle was provided.

[Response Ends]