NATIONAL QUALITY FORUM
Driving measurable health improvements together

Measure Worksheet

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

Brief Measure Information

NQF #: 0216
Corresponding Measures:
Measure Title: Percentage of patients who died from cancer admitted to hospice for less than 3 days
Measure Steward: American Society of Clinical Oncology
sp.02. Brief Description of Measure: Percentage of patients who died from cancer, and admitted to hospice and spent less than 3 days there
1b.01. Developer Rationale: The Institute of Medicine's report, Dying in America, advocates for measures to improve the quality and sustainability of end-of-life care, urging the federal government to "require public reporting on quality measures, outcomes, and costs regarding care near the end-of-life" (Daly et al., 2016). Hospice care is a form of palliative care for patients with a limited life expectancy of six months or less (according to Medicare hospice coverage criteria) who want to focus on quality of life and comfort rather than life-prolonging care. Most insurance plans provide coverage for hospice (ICSI Guideline, 2020). Currently the median length of stay in hospice before death for Medicare cancer patients is about 19 days (NCCN Guidelines, 2021). Approximately 28% of hospice patients died or were discharged within 7 days of admission to hospice care (NCCN Guidelines, 2021). This is despite the fact that the hospice benefit is at least six months or longer if needed. This short length of stay means that the patient, family and care team have limited time to get a plan of care in place before death and that the focus tends to be more on the care of the imminently dying patient than living life to the fullest before the final decline in function. Early referral to hospice increases the likelihood that pain and other symptoms will be managed more aggressively and therefore there will be less anxiety and distress at the end of life (ICSI Guideline, 2020).
One retrospective study of more than 64,000 patients with cancer who were admitted to hospice found that over 16% of those patients were only enrolled in the last three days of life or less (O'Connor, 2015). The rate of patients who do not have a hospice referral prior to death continues to be higher than desired with one study reporting that more than 30% of patients were not referred and of those patients, only 7% had a documented discussion on the option of palliative care (O'Connor, 2015).
Patients enrolled in hospice experience increased survival times along with a reduction in resource use such as aggressive end of life care and hospital admissions; benefits that increased the longer patients are enrolled in hospice (Lee, 2015; Langton, 2014).- Patients who use hospice, compared with those who do not use hospice, have markedly improved symptoms, less caregiver distress, reduced costs of approximately $8,700 per Medicare beneficiary, and, according to two published reports, actually live longer (ASCO Guideline, 2017).
References:
Daly, B., Hantel, A., Wroblewski, K., Balachandran, J.S., Chow, S., DeBoer, R., Fleming, G.F., Hahn, O.M., Kline, J., Liu, H., Patel, B.K., Verma, A., Witt, L.J., Fukui, M., Kumar, A., Howell, M.D., Polite, B.N. (2016). No Exit: Identifying Avoidable Terminal Oncology Intensive Care Unit Hospitalizations. J Oncol Pract,12(10), e901-e911. doi: 10.1200/JOP.2016.012823. Retrieved from https://pubmed.ncbi.nlm.nih.gov/27601514/
ICSI Health Care Guidelines: Palliative Care for Adults. (2020). Sixth edition. Retrieved from https://www.icsi.org/wp-content/uploads/2021/11/PalliativeCare_6th-Ed_2020_v2.pdf
Ferrell, B. R., Temel, J. S., Temin, S., Alesi, E. R., Balboni, T. A., Basch, E. M., Firn, J. I., Paice, J. A., Peppercorn, J. M., Phillips, T., Stovall, E. L., Zimmermann, C., & Smith, T. J. (2017). Integration of Palliative Care Into Standard Oncology Care: American Society of Clinical Oncology Clinical Practice Guideline Update. *Journal of Clinical Oncology*, *35*(1), 96–112. Retrieved from https://doi.org/10.1200/jco.2016.70.1474

O´Connor, T. L., N. Ngamphaiboon, et al. (2015). "Hospice utilization and end-of-life care in metastatic breast cancer patients at a comprehensive cancer center." J Palliat Med 18(1): 50-55.
Langton, J. M., B. Blanch, et al. (2014). "Retrospective studies of end-of-life resource utilization and costs in cancer care using health administrative data: a systematic review." Palliat Med 28(10): 1167-1196.
Lee, Y. J., J. H. Yang, et al. (2015). "Association between the duration of palliative care service and survival in terminal cancer patients." Support Care Cancer 23(4): 1057-1062.
National Comprehensive Cancer Network Clinical Practice Guidelines in Oncology. (2021). Palliative Care (version 2.2021). Retrieved from https://www.nccn.org/guidelines/guidelines-detail?category=3&id=1454.

---

sp.12. Numerator Statement: Patients who died from cancer and spent fewer than three days in hospice.
sp.14. Denominator Statement: Patients who died from cancer who were admitted to hospice
sp.16. Denominator Exclusions: None

Measure Type: Process
sp.28. Data Source: Registry Data
sp.07. Level of Analysis: Clinician: Individual; Clinician: Group/Practice

---

IF Endorsement Maintenance – Original Endorsement Date: 10/01/2007
Most Recent Endorsement Date: 10/26/2016

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

Preliminary Analysis: Maintenance of Endorsement

To maintain NQF endorsement, endorsed measures are evaluated periodically to ensure that the 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 change in evidence since the prior evaluation.

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

The developer provides the following description for this measure:

- This is a maintenance process measure at the Clinician Group and Clinician Individual level that measures the percentage of patients who died from cancer within three days of a hospice admission.
- The developer provides a logic model that depicts actions that can be taken by the accountable entity in terms of timely enrollment in palliative care and hospice services, as well as a reduction of aggressive interventions at the end-of-life (EOL). Taken together, these actions ultimately result in

improved quality of life, patient and caregiver/family satisfaction at EOL, and lower resource utilization costs.

The developer provides the following evidence for this measure:

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

Summary of prior review in 2016

- The developer provided evidence during the previous review that evaluated the impact of palliative care/hospice services on patient EOL experiences.
    - One 2013 systematic review (SR) found that home-based palliative/hospice care significantly increased the likelihood of an individual with advanced illness dying at home.
    - A 2012 clinical opinion from the American Society of Clinical Oncology (ASCO) addressed the integration of palliative care/hospice services on patient and caregiver outcomes (e.g., improved overall survival, reduced depression, enhanced quality of life, decreased resource use, and cost).
- The Standing Committee agreed that the updated evidence provided during the previous evaluation appeared to be directionally the same since the last endorsement review in 2012 and accepted the prior evaluation without further discussion.

Changes to evidence from last review

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

☒ The developer provided updated evidence for this measure:

- The developer added three additional clinical practice guidelines and one SR with varying levels of evidence:
    - The 2021 National Comprehensive Cancer Network (NCCN) Clinical Practice Guidelines in Oncology
        - Recommendation: Oncologists should integrate palliative care into general oncology care. Early consultation/collaboration with a palliative care specialist/hospice team should be considered to improve quality of life and survival. (Category 2A)
        - Quality: The developer noted that the evidence is based upon lower-level evidence and there is uniform NCCN consensus that the intervention is appropriate.
        - Quantity: The developer noted that the NCCN guidelines do not provide information on the quantity of studies.
        - Consistency: The developer noted that the NCCN guidelines do not provide this information.
    - The 2017 Integration of Palliative Care into Standard Oncology Care: American Society of Clinical Oncology Clinical Practice Guideline Update
        - Recommendation: Patients with advanced cancer should be referred to interdisciplinary palliative care teams (consultation) that provide inpatient and outpatient care early in the course of disease, alongside active treatment of their cancer. (Type: evidence based, benefits outweigh harms; Evidence Quality: Intermediate; Strength of Recommendation: Strong)
        - Quality: The developer noted the recommendation as having moderate confidence that the available evidence reflects the true magnitude and direction of the net effect.

        - Quantity: The developer noted that a total of nine new RCTS, two publications reporting on one large quasi-experimental trial, and five secondary publications based on prior published RCTs met eligibility criteria and were suggested by the Expert Panel.
        - Consistency: The developer noted that there is high confidence that the recommendation reflects best practice due to overall strong quality and quantity of evidence and Expert Panel consensus.
    - The 2020 Institute for Clinical Systems Improvement (ICSI) Guidelines
        - Recommendation 1 (Palliative Care): Palliative care discussion or referral should be considered whenever a patient develops or presents with a serious or life-threatening illness, in all care settings. (Quality of Evidence: Low; Strength of Recommendation: Strong)
        - Recommendation 2 (Hospice Referral): In a patient with serious illness, clinicians should recognize the prognosis of less than 6 months, and if in line with the patient goals of care, refer to hospice. (Quality of Evidence: Low; Strength of Recommendation: Strong)
        - Quality: The developer noted for both recommendations that the quality of evidence is low with limited effect estimates (the true effect may be substantially different from the estimate of the effect).
        - Quantity: For the first recommendation, the developer noted a total of two SR/meta-analysis, one report, one review, one summary, and one consensus report form the evidentiary basis for the guideline recommendation. For the second recommendation, the developer noted a total of two SRs, one observational study, one summary, and one controlled trial form the evidentiary basis for the guideline recommendation.
        - Consistency: The developer noted that no information was provided on the consistency across studies.
    - A 2019 SR found that EOL discussions are associated with increased use of hospice services and lower healthcare resource utilization costs in the final 30 days of life. Conversations that occur greater than 30 days before death are strongly associated with less aggressive interventions compared to discussions that occur near time of death.
        - Quality: The developer noted that the 20 studies in this SR were assessed by two authors using the Oxford Centre for Evidence-based Medicine Levels of Evidence grading guide; levels of quality varied among the studies.
        - Quantity: Twenty studies (one retrospective analysis of a randomized clinical trial (RCT); one non-randomized, intervention-based study; and 18 observational studies) were included in the SR; all studies were published between January 2012 and January 2019.
        - Consistency: The developer noted that there was wide variation among the studies on the definition of EOL discussions and how clinicians conceptualize aggressive care and overuse of healthcare services near EOL.
- The developer noted that there were no harms identified across the studies.
- The developer also provided a 2016 Cancer Care Outcomes Research and Surveillance study focused on aggressive EOL indicators using patient-and-family-centered outcomes among 1,146 cancer patients.
    - The developer noted that the study reported that patients who died within three days of a hospice admission were associated with large differences in family-reported perceptions of EOL care quality and a lower likelihood that patients receive care congruent with their preferences.
    - The developer also noted that the families were much more likely to report that patients died in their place of preference if they were enrolled in hospice greater than 3 days before their death.

Exception to evidence

- N/A

Questions for the Committee:

- *What is the relationship of this measure to patient outcomes?*
- *How strong is the evidence for this relationship?*
- *Is the evidence directly applicable to the process of care being measured?*
- *Does the Standing Committee feel that the evidence provided by the developer supports the relationship between the 3-days spent in hospice or palliative care before death to desired patient outcomes and reduced utilization?*
- *Does the Standing Committee have any concerns with the 3-day timeframe specified in the measure?*

Guidance from the Evidence Algorithm

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

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

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

Maintenance measures – increased emphasis on gap and variation

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

- For this measure, a lower score indicates better performance.
- The developer provided performance data derived from the ASCO Quality Oncology Practice Initiative (QOPI) for 2017 through 2020. The mean performance rate for 154 practices in Spring 2017 was 16.85 percent (standard deviation [SD]). The most recent data for round one 2020 across 68 practices had a mean performance rate of 22.84 percent (SD 16.78).
- The developer noted that the MIPS performance benchmark mean in Program Year (PY) 2020 was 8.88 percent. The performance benchmark mean in PY 2017 and 2018 were 10.7 and 10.83, respectively.
- The developer also noted that the MIPS performance rates may not be a national representation as participants are allowed to self-select measures and may chose those measures that will result in higher performance rates.
- The developer cited literature to demonstrate opportunities for performance improvement.
    - A 2015 retrospective study found that over 16 percent of patients were enrolled in hospice services within the last three days of life.

Disparities

- The developer provided citations from literature to demonstrate that disparities across different racial and ethnic groups and Medicaid status exist in this area of healthcare.
    - A longitudinal cohort study in 2010 found that compared to White patients, African American patients (OR 0.37, 95% CI 0.19 to 0.73) and Hispanic patients (OR 0.29, 95% CI 0.12 to 0.73) were less likely to experience end-of-life (EOL) discussions.
    - A 2020 retrospective cohort study supported the hypothesis that a higher proportion of patients with cancer who were enrolled in Medicaid received aggressive EOL care compared to patients with Medicare or privately insured populations. Additionally, the developer noted

compared to White patients, African American patients had higher chances of receiving any aggressive EOL care (odds ratio [OR], 1.87; 95% CI, 1.07 to 3.26).

- It is unclear from the literature if these findings address disparities in care related to the specific measure focus (i.e., patients who died from cancer and spent fewer than three days in hospice).

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Are you aware of evidence that disparities exist in this area of healthcare?*
- *Does the Standing Committee have any concerns with the level of analysis for the performance data provided (i.e., individual clinician, group, both)?*

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

Committee Pre-evaluation Comments:

1a. Evidence

- While one study cited by the developer provides support for the 3-day time frame, I wonder if the findings would have been similar for a 4-day, 5-day, or less than a week timeframe. I do have concerns about limiting the time to 3 days and wish there was clearer support for limiting to 3 days.
- The developers present strong evidence that supports the connection between this process measure and outcomes although that evidence seems to be mostly tangential. However, the volume of the evidence and its consistency supports the conclusion drawn. The process measure tested here clearly has a relationship to some desirable outcomes.
- Evidence is adequate

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

- A gap is demonstrated. In regard to disparities, the studies did not focus on patients dying from cancer who spent fewer than three days receiving hospice care. The studies cited did note that compared to white patients, AFrican American patiens and Hispanic patients were less likely to experience EOL discussions, and African American patients more likely to receive aggressive EOL care. Studies show that the percentage of African American and other non-white communities are incredibly small overall regardless of diagnosis. Some factors for these results are a lack of trust on the part of the African American community to have EOL conversations with healthcare professionals (particularly when they are white) and a lack of trust that they will receive the same amount of care as whites so they seek to have all the care they possibly can.
- The developers provide data demonstrating several disparities in the performance date.
- Yes, a gap exists

Criteria 2: Scientific Acceptability of Measure Properties

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

Evaluators: Staff

2a. Reliability: Specifications and Testing

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

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

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

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

Specifications:

- Measure specifications are clear and precise.
- The developer noted that there was a minor update to the measure title and description, replacing *Proportion* with the term *Percentage.*

Reliability Testing:

- Reliability testing conducted at the Patient/Encounter Level:
    - The developer conducted inter-rater reliability testing on patient-level data across 264 patient records submitted from 44 practices using the 2008 Quality Oncology Practice Initiative (QOPI) dataset.
    - Trained, independent nurse abstractors served as the "gold-standard" against which practice abstractions were compared for accuracy.
    - The developer provided an inter-rater reliability kappa of 55.13 percent for all data elements and data element combinations assessed. Kappa values ranged from low of 0.51846 for the denominator to a high of 0.58414 for the numerator.
- Reliability testing conducted at the Accountable Entity Level:
    - The developer indicates that updated reliability testing was conducted at the clinician: group/practice level yet they were unable to determine from the rolled-up data sample the number unique NPIs (n= 215) who reported to the 2017 Physician Quality Reporting System (PQRS) registry as an individual or a group. Therefore, the developer recommends that the measure should be considered for endorsement at the group/practice level using a potential group size of one.
    - The developer conducted a signal to noise analysis using a beta-binomial model on 215 individual clinicians with results ranging from 0.1099 to 1 (mean= 0.7921; IQR= 0.4698; SD= 0.3082).
    - The developer noted that half of providers reporting on this measure have a reliability of 100 percent.

Questions for the Committee regarding reliability:

- *Do you have any concerns that the measure cannot be consistently implemented (i.e., are measure specifications adequate)?*
- *Is it likely this measure can be consistently implemented?*
- *Does the Standing Committee have any concerns with the age of the data used for the reliability testing?*

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 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 Accountable Entity Level:
    - Face validity testing was conducted in 2016
        - ASCO-led focus groups and structured interviews were conducted with patients diagnosed with terminal cancer and receiving end-of-life care and their bereaved caregivers
        - Surveys were performed to solicit patient preferences for care, desire to avoid overly aggressive.
        - An expert panel of 12 cancer subject matter experts asked to provide an accurate reflection of the quality actions and if the scores obtained from the measure as specified can be used to distinguish good from poor quality of care.
        - Agreement of the validity of the measure was high
            - One hundred percent of the respondents “agreed” or “strongly agreed” that this measure provides an accurate reflection of quality that can be used to distinguish good and poor quality.
            - One hundred percent of the experts “strongly agreed” or “agreed’ that the measure specifications are appropriate and align with current evidence.
            - Ninety-two percent of subject matter experts 'strongly agreed' or 'agreed' that the performance score for the measure is meaningful, understandable, and useful for public reporting.
    - Concurrent validity conducted in 2022
        - The developer conducted concurrent bivariate correlation analysis using 2017 PQRS datasets for two correlated measures, NQF #0216 Percentage Admitted to Hospice for Less Than 3 Days and NQF #0210 Percentage Receiving Chemotherapy in the Last 14 Days of Life.
            - The developer hypothesized that a positive association exists between both measure due to the similarities in both domain of the quality action and patient populations.
            - The developer calculated a Pearson correlation coefficient to evaluate the association across 12 provider scores on both measures (r=0.9158).
            - The developer noted the results of the correlation indicate a strong positive relationship between measures.

Exclusions

- The measure does not use exclusions.

Risk-Adjustment

- The measure is not risk adjusted or stratified.

Meaningful Differences

- Provider performance across 215 providers ranged from 100 (minimum) to 0 (maximum) with a median percentage score of 0 percent (interquartile range [IQR] 25, mean 16.88 percent, SD 28.14)
- The developer noted the distribution of performance scores across 215 providers is highly skewed, with the largest number of providers reporting a perfect score of 0 percent.

Missing Data

- The developer noted that due to PQRS data completeness requirements, the dataset did not contain missing data.

Comparability

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

Questions for the Committee regarding validity:

- *Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?*
- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*
- *Does the Standing Committee have any concerns with the level of analysis (clinician/group practice; clinician/individual) used for validity testing?*

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

Committee Pre-evaluation Comments:

2a1. Reliability-Specification

- I have no concerns about reliability specifications.
- No new data provided. No concerns.
- there is good reliability

2a2. Reliability - Testing

- No concerns
- No concerns
- I do not have any concerns at this time

2b. Validity

- My only concern is that with the advances in cancer treatment such immunotherapy it is becoming difficult to make prognostic projections and people may benefit from these treatments and so understandably defer hospice, since current hospice policy doesn't pay for both. That's a problem with hospice policy but can make this measure misleading as something trying to evaluated good care.
- No concerns
- I have no concerns. I appreciate that the developer has conducted concurrent validity testing in 2022, as this is stronger than the face validity originally conducted in 2016.

2b2-2b6. Potential threats to validity

- I have no concerns about exclusions or risks.
- The measure is not risk adjusted.
- Yes, there is a link and risk adjustment may make sense

2b4-7. Threats to Validity

- No threats to validity reported.
- There are difference in hospice use across demographic groups so it would be good to have this measure reported on that basis and consider things like risk adjustment where gaps are significant.

- I have no concerns

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 developer noted that data elements needed to compute the measure score can be abstracted by someone other than the person obtaining the original information.
- The developer indicates all data elements are in defined fields in electronic clinical data.

Questions for the Committee:

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

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

Committee Pre-evaluation Comments:

3. Feasibility

- The data elements are routinely generated and are available in electronic form. I have no concerns about putting the data collection strategy into operational use.
- Data collection is in operational use. All elements are available.
- The measure is feasible.

Criterion 4: Use and Usability

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

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

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

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

Current uses of the measure

Publicly reported? ☒ Yes ☐ No

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

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

Accountability program details

- The measure is used in the following programs:
    - The CMS PPS -Exempt Cancer Hospital Quality Reporting (PCHQR) Program: Eligible PPS-Exempt Cancer Hospitals (PCHs) report on best practices for their respective facilities and types of care.
    - MIPS Program: MIPS eligible providers may earn performance-based payment adjustments for the services provided to Medicare patients.
    - Polaris: A CMS-approved Qualified Clinical Data Registry (QCFR) hosted by FIGmd that offers a simplified approach to MIPS reporting
    - ASCO's QOPI: An oncologist-let, practice-based quality assessment and improvement program available to all practices with at least one active ASCO member
- The measure is included in the following measure sets:
    - The Core Quality Measures Collaborative (CQMC) 2020 Medical Oncology Core Set: measure set that promotes a patient-centered assessment of quality

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

Feedback on the measure by those being measured or others

- The developer notes that CMS publicly reports MIPS program performance rates and benchmarks annually for all actively participating eligible providers m and offers support for those using the measure through the Quality Payment Program (QPP) Service Center and webinars.
- The PCHQR program publishes care data publicly on a rolling quarter basis and offers assistance to those using the measure through their QualityNet Service Center.
- The developer notes that the ASCO's measurement team is available to receive comments and questions from measure implementers and clinicians reporting ASCO measures by email and notes that no specific feedback has been received on this measure.

***Questions for the Committee**:*

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

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

RATIONALE: [Rationale for voting low or insufficient]

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

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

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

Improvement results

- The developer provided a 2013-2015 mean performance rate of 17.57 percent. For 2017-2020, the mean performance rate was 19.24 percent. This performance data were derived from the ASCO Quality Oncology Practice Initiative (QOPI).
    - The developer noted that the number of practices reporting on this measure has increased since 2013.
    - The developer further notes that the available performance data indicates continued performance at lower levels ranging from 61.54 percent to 100 percent.
- The developer noted the 2017-2020 mean performance rate of 10.83 to 8.48 percent using MIPS performance data.
- It is unclear from these findings what level of performance the data is for (i.e., individual clinician, groups, both).

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

Unexpected findings (positive or negative) during implementation

- The developer noted there were no unexpected findings.

Potential harms

- The developer noted there were no potential harms.

***Questions for the Committee**:*

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

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

Committee Pre-evaluation Comments:

4a. Use

- The measure is publicly reported and currently used in an accountability program. Also, feedback is generated.
- The measure is publicly reported, and available. Data is widely shared with users and feedback received.
- No use concerns

4a. Usability

- I think the results can serve as evidence for the need to understand why people are being admitted so late to hospice such that effective interventions can be developed and implemented. I also feel the same about the disparities that exist. I do not think there is any harm or unintended consquences.
- Data can continue to be used to increase quality of care at EOL. No unintended consequences noted.
- No harms

Criterion 5: Related and Competing Measures

Related measures

- NQF #0210 Percentage of patients who died from cancer receiving chemotherapy in the last 14 days of life
- NQF #0213 Percentage of patients who died from cancer admitted to the Intensive Care Unit (ICU) in the Last 30 Days of Life
- NQF #2651 CAHPS® Hospice Survey (experience with care)
- NQF #3235 Hospice and Palliative Care Composite Process Measure—Comprehensive Assessment at Admission

Harmonization

- The developer indicates that measures are harmonized to the extent possible.

Committee Pre-evaluation Comments:

5: Related and Competing Measures

- My answer to all these questions is 'none that I am aware of.'

some related measures but nothing that addresses this specific issue

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

Member Expression of Support

- One NQF member submitted an expression of "support" for the measure.

Comments

Comment 1 by: Anna Kim, American Geriatrics Society; Submitted by Anna Kim

While the American Geriatrics Society (AGS) is generally supportive of the measure, we are concerned that practices may be disincentivized to refer patients to hospice in order to evade the patient spending less than three days in hospice and reduce the percentage of referrals. We believe that a later referral to hospice would be more helpful than none at all, particularly as caregivers may benefit from bereavement support and patients may die at home more peacefully. Further, it was not clear from the materials provided whether patients who died from cancer but were never admitted to hospice were accounted for in the measure development. The AGS recommends further consideration of the concerns raised and how they can be addressed.

Comment 2 by: Lela Durakovic, American Society of Clinical Oncology; Submitted by Lela Durakovic, American Society of Clinical Oncology

Measures #0210, #0213, and #0216 were tested using a 2017 PQRS/MIPS registry dataset, and the signal-to-noise analysis was performed at the provider NPI level. Since CMS de-identified all practice and provider ID's in the registry dataset, we were initially unable to determine whether these NPIs belong to individual clinicians or provider organizations. Therefore, the initial level of analysis was set as Clinician-Group as ASCO could not confidently state that the dataset contained only individual clinician NPIs. However, after the initial signal-to-noise analysis, ASCO received feedback from CMS that only individual clinician NPIs are eligible for the MIPS program. Hence, we can now confidently state that the 2017 PQRS/MIPS registry dataset contains only individual clinician NPIs and that the analysis meets the specificity requirements for the Clinician-Individual level. Additionally, since the signal-to-noise reliability results at the NPI level were high for all three measures, ASCO feels confident in recommending that the NPI level analysis be used to prove reliability at the Clinician-Group level. Calculating group-level reliability by combining patient scores under individual NPIs into larger groupings according to organizations' TINs will increase the sample sizes of patient scores and produce more reliable results with greater precision and power. Therefore, performing a group-level analysis will introduce no potential

threats to the measures' reliability. The reliability of measure scores at group-level analysis can only increase.

Scientific Acceptability Evaluation

RELIABILITY: SPECIFICATIONS

1. **Have measure specifications changed since the last review? ☒ Yes ☐ No**
2. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented? ☒ Yes ☐ No**
3. **Briefly summarize any changes to the measure specifications and/or concerns about the measure specifications.**
   - The developer provided minor updates to the measure title and description, but the measure intent and calculation remains the same.
     - Proportion was replaced by Percentage.

RELIABILITY: TESTING

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

   4a. If no, summarize the Standing Committee's previous feedback:
   - N/A

   4b. If yes, describe any differences between the new and old testing and summarize any relevant Standing Committee's feedback from the previous review:
   - The developer did not provide updated reliability or validity testing during the previous evaluation.
   - During the previous review, the Standing Committee noted that the measure specified both claims and registry data. When questioned about identifying cancer deaths from claims data, the developer clarified that the denominator is derived from registry data (e.g., a death registry or other cancer registry that includes information on cancer deaths) while the numerator is derived from claims data or the ASCO QOPI Initiative.
   - The Standing Committee accepted the prior evaluation of the reliability criterion without further discussion during the 2016 evaluation.
   - The developer noted that since the last review, the Physician Quality Reporting System (PQRS) has transitioned to the Merit-based Incentive Payment System (MIPS).
   - The developer provided new reliability testing using 2017 CMS MIPS performance data.
5. **Reliability testing level: ☒ Accountable-Entity Level ☒ Patient/Encounter Level ☐ Neither**
6. **Reliability testing was conducted with the data source and level of analysis indicated for this measure:**

   ☒ Yes ☐ No
7. If accountable-entity level and/or patient/encounter level reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of patient-level data conducted?

   ☐ Yes ☐ No
8. **Assess the method(s) used for reliability testing:**
   - The developer conducted inter-rater reliability testing from the Quality Oncology Practice Initiative (QOPI) dataset (2008 Virginia Quality Health Center Quality Improvement Organization's case report).
   - The developer conducted a signal to noise analysis of 215 NPIs using 2017 PQRS registry performance data.
9. **Assess the results of reliability testing**
   - Nurse abstractors conducted testing across 44 practices on a total of 264 records (i.e., six records per practice) with a kappa result of 0.551.

- The developer used the beta-binomial model to assess the signal-to-noise analysis with overall reliability ranging from 0.1099 to 1 (mean= 0.7921) indicating high reliability.

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

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

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

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

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

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

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

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

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

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

    Measure specifications precise, unambiguous, and complete (Box 1) -> Empirical reliability testing conducted with the measure as specified (Box 2) -> Empirical testing at the accountable entity level (Box 4) -> Reliability testing method described and appropriate (Box 5) -> Moderate certainty or confidence that the levels are reliable (Box 6b) -> Moderate rating

VALIDITY: TESTING

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

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

    - N/A

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

    - During the previous evaluation, the Standing Committee did not express concern with the lack of risk-adjustment.
    - The developer did not provide updated testing during the previous evaluation.
    - The Standing Committee agreed the previous validity testing demonstrated the scientific acceptability of the measure and passed with a rating of moderate.
    - The developer provided accountable entity level testing and empirical level testing in addition to the face validity submitted during the previous submission.

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

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

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

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

    ☐ **Yes**

    ☐ **No**

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

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

☒ **Face validity**

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

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

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

☒ **Yes**

☐ **No**

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

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

- The developer conducted concurrent validity testing using bivariate correlation analysis on a subset of 12 providers from the 2017 PQRS registry performance dataset who reported on measures NQF #0216 Proportion Admitted to Hospice for Less Than 3 Days and NQF #0210 Proportion Receiving Chemotherapy in the Last 14 Days of Life.
- In 2016, the developer performed face validity, convening focus groups interviewing patients diagnosed with terminal cancer and receiving end-of-life care, along with their bereaved caregivers and vetted by oncology providers.

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

- The developer calculated a Pearson correlation coefficient to evaluate the association between NPI scores on both measures (n=12 providers, r= 0.9158).
- A majority of the expert panel (97%) "agreed" or "strongly agreed" that this measure provides an accurate reflection of quality that can be used to distinguish good and poor quality.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

- The developer stated there were no exclusions.

22. **Risk Adjustment**

22a. **Risk-adjustment method**

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

☐ Other method assessing risk factors (please specify)

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

☐ Yes ☐ No ☒ Not applicable

22c. **Social risk adjustment:**

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

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

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

22d.**Risk adjustment summary:**

22d.1 All of the risk-adjustment variables present at the start of care? ☐ Yes ☐ No

22d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐ Yes ☐ No

22d.3 Is the risk adjustment approach appropriately developed and assessed? ☐ Yes ☐ No

22d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration)
☐ Yes ☐ No

22d.5.Appropriate risk-adjustment strategy included in the measure? ☐ Yes ☐ No

22e. **Assess the risk-adjustment approach**

- N/A

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

For cost/resource use measures, does this measure identify meaningful differences about cost and resource use between the measured entities?

- The developer acknowledges that while more than half of the providers had 100 percent performance on the measure, there remains wide variation among the remaining providers (n=215, range 0-1, interquartile range [IQR] 0.25, mean 0.1688, standard deviation [SD] 0.2814, 95 percent confidence interval [CI] for mean performance 0.13-0.21)

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

- The developer stated there is only one data source for this measure.

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

- There is concern that the data set provided to the developer did not contain missing data, therefore missing data testing was not conducted.

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

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

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

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

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

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

Threats to validity empirically assessed (Box 1) -> Empirical validity testing conducted using the measure as specified (Box 2) -> Empirical validity conducted at the accountable entity level (Box 5) -> Validity testing method described and appropriate (Box 6) -> Moderate certainty or confidence (Box 7b) -> Moderate rating

ADDITIONAL RECOMMENDATIONS

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

Criteria 1: Importance to Measure and Report

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

---

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

[Response Begins]
Yes
[Yes Please Explain]
There are new clinical guidelines, systematic reviews, and relevant research studies that continue to support the measure since the last NQF submission in 2016.

[Response Ends]

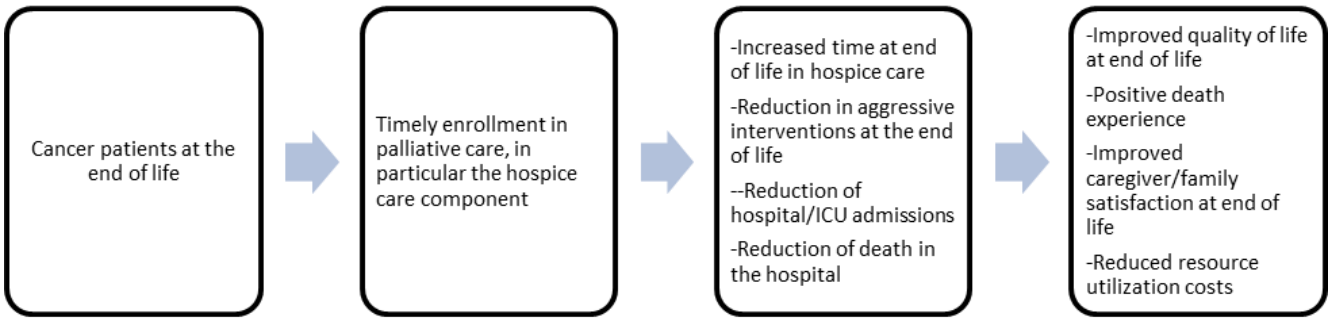
Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Measure and Report: Evidence section. For example:
2021 Submission:
Updated evidence information here.
2018 Submission:
Evidence from the previous submission here.

1a. Evidence

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]



Benefits of Timely Enrollment in Hospice for Cancer Patients at the End of Life

Timely enrollment in palliative care, specifically the hospice care component, for the cancer population can lead to reduction of aggressive interventions at the end of life and increased benefits from hospice care, which ultimately can improve quality of life at end of life, provide a positive death experience, improve family satisfaction at end of life, and reduce resource utilization costs.

[Response Ends]

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

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

[Response Begins]

Clinical Practice Guideline recommendation (with evidence review)

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

Other (specify)

[Other (specify) Please Explain]

One multi-regional, prospective, observational study included in evidence (1a.13).

[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]

2022 Submission:

1. National Comprehensive Cancer Network Clinical Practice Guidelines in Oncology. (2021). Palliative Care (version 2.2021). Retrieved from https://www.nccn.org/guidelines/guidelines-detail?category=3&id=1454.
    a. Date: February 12, 2021
    b. Page no (for recommendation): PAL 3 and PAL 6
    c. URL: https://www.nccn.org/guidelines/category_3
2. Ferrell, B. R., Temel, J. S., Temin, S., Alesi, E. R., Balboni, T. A., Basch, E. M., Firn, J. I., Paice, J. A., Peppercorn, J. M., Phillips, T., Stovall, E. L., Zimmermann, C., & Smith, T. J. (2017). Integration of Palliative Care Into Standard Oncology Care: American Society of Clinical Oncology Clinical Practice Guideline Update. *Journal of Clinical Oncology*, *35*(1), 96–112. Retrieved from https://doi.org/10.1200/jco.2016.70.1474
    a. Date: January 1, 2017
    b. Page no (for recommendation): 97 and 99
    c. URL: https://ascopubs.org/doi/10.1200/JCO.2016.70.1474
3. ICSI Health Care Guidelines: Palliative Care for Adults. (2020). Sixth edition. Retrieved from https://www.icsi.org/wp-content/uploads/2021/11/PalliativeCare_6th-Ed_2020_v2.pdf
    a. Date: January 2020
    b. Page no (for recommendation): 12 and 39
    c. URL: https://www.icsi.org/wp-content/uploads/2021/11/PalliativeCare_6th-Ed_2020_v2.pdf
4. Starr, L. T., Ulrich, C. M., Corey, K. L., & Meghani, S. H. (2019). Associations Among End-of-Life Discussions, Health-Care Utilization, and Costs in Persons With Advanced Cancer: A Systematic Review. The American journal of hospice & palliative care, 36(10), 913–926. https://doi.org/10.1177/1049909119848148

[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]
2022 Submission:

1. NCCN Clinical Guidelines in Oncology: Palliative Care. 2021.
    a. Recommendation: Oncologists should integrate palliative care into general oncology care. Early consultation/collaboration with a palliative care specialist/hospice team should be considered to improve quality of life and survival (Category 2A).
2. Integration of Palliative Care Into Standard Oncology Care: ASCO Clinical Practice Guideline. 2017.
    a. Recommendation: Patients with advanced cancer should be referred to interdisciplinary palliative care teams (consultation) that provide inpatient and outpatient care early in the course of disease, alongside active treatment of their cancer (type: evidence based, benefits outweigh harms; evidence quality: intermediate; strength of recommendation: strong).
3. ICSI Health Care Guidelines: Palliative Care for Adults. 2020.
    a. Recommendation 1 (Palliative Care): Palliative care discussion or referral should be considered whenever a patient develops or presents with a serious or life threatening illness, in all care settings. (Quality of Evidence: Low; Strength of Recommendation: Strong)
    b. Recommendation 2 (Hospice Referral): In a patient with serious illness, clinicians should recognize the prognosis of less than 6 months, and if in line with the patient goals of care, refer to hospice. (Quality of Evidence: Low; Strength of Recommendation: Strong)
4. Associations Among End-of-Life Discussions, Health-Care Utilization, and Costs in Persons With Advanced Cancer: A Systematic Review. 2019.
    a. End-of-life discussions are associated with lower healthcare costs in the last 30 days of life (median $1,048 vs. $23,482; $p < .001$); greater use of hospice (ORs ranging 1.79 to 6.88). Earlier EOL discussions (30+ days before death) are more strongly associated with less aggressive care outcomes than conversations occurring near death.

[Response Ends]

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

[Response Begins]
2022 Submission:

1. NCCN Clinical Guidelines in Oncology: Palliative Care. 2021.
    a. Category 2A definition-Based upon lower-level evidence, there is uniform NCCN consensus that the intervention is appropriate.
2. Integration of Palliative Care Into Standard Oncology Care: ASCO Clinical Practice Guideline. 2017.
    a. Type: evidence based, benefits outweigh harms;
    b. Evidence quality: intermediate;
        i. Definition of intermediate- Moderate confidence that the available evidence reflects the true magnitude and direction of the net effect. Further research is unlikely to alter the direction of the net effect; however, it might alter the magnitude of the net effect.
3. ICSI Health Care Guidelines: Palliative Care for Adults. 2020.
    a. Quality of Evidence (for both recommendations): Low
        i. Definition of low- Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect.
4. Associations Among End-of-Life Discussions, Health-Care Utilization, and Costs in Persons With Advanced Cancer: A Systematic Review. 2019.
    a. No grading provided as it is not a clinical guideline recommendation. However, the quality of the 20 studies in this systematic review was assessed by two authors using the Oxford Centre for Evidence-based Medicine Levels of Evidence grading guide.
        i. Definition of Level 2b (13 studies in this systematic review): Individual cohort study (including low quality RCT; e.g., <80% follow-up)
        ii. Definition of Level 4 (6 studies in this systematic review): Case-series (and poor quality cohort and case-control studies)
        iii. Definition of Level 1b (1 study in this systematic review): Individual RCT (with narrow Confidence Interval")

[Response Ends]

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

[Response Begins]
2022 Submission:

1.NCCN Clinical Guidelines in Oncology: Palliative Care. 2021.

The guideline grading system is based on the NCCN Categories of Evidence and Consensus.
Category 1- Based upon high-level evidence, there is uniform NCCN consensus that the intervention is appropriate
Category 2A -Based upon lower-level evidence, there is uniform NCCN consensus that the intervention is appropriate.
Category 2B- Based upon lower-level evidence, there is NCCN consensus that the intervention is appropriate.
Category 3-Based upon any level of evidence, there is major NCCN disagreement that the intervention is appropriate.

2. Integration of Palliative Care Into Standard Oncology Care: ASCO Clinical Practice Guideline. 2017.

The guideline recommendations were crafted, in part, using the GuideLines Into DEcision Support (GLIDES) methodology and accompanying BRIDGE-Wiz software™
Quality of Evidence Definitions:
High- High confidence that the available evidence reflects the true magnitude and direction of the net effect (i.e., balance of benefits v harms) and that further research is very unlikely to change either the magnitude or direction of this net effect.
Intermediate- Moderate confidence that the available evidence reflects the true magnitude and direction of the net effect. Further research is unlikely to alter the direction of the net effect; however, it might alter the magnitude of the net effect.
Low- Low confidence that the available evidence reflects the true magnitude and direction of the net effect. Further research may change either the magnitude and/or direction this net effect.

Insufficient- Evidence is insufficient to discern the true magnitude and direction of the net effect. Further research may better inform the topic. The use of the consensus opinion of experts is reasonable to inform outcomes related to the topic.

3. ICSI Health Care Guidelines: Palliative Care for Adults. 2020.

The guideline recommendations were crafted, in part, using GRADE methodology.
Quality of Evidence Definitions:
High = We are very confident that the true effect lies close to that of the estimate of the effect.
Moderate = We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low = Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect.
Very low = We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect.

4. Associations Among End-of-Life Discussions, Health-Care Utilization, and Costs in Persons With Advanced Cancer: A Systematic Review. 2019.

a. No grading provided as it is not a clinical guideline recommendation. However, the quality of the 20 studies in this systematic review was assessed by two authors using the Oxford Centre for Evidence-based Medicine Levels of Evidence grading guide.

Therapy/Prevention/Etiology/Harm Level of Evidence Definitions:
1a: Systematic reviews (with homogeneity) of randomized controlled trials
1b: Individual randomized controlled trials (with narrow confidence interval)
1c: All or none randomized controlled trials
2a: Systematic reviews (with homogeneity) of cohort studies
2b: Individual cohort study or low quality randomized controlled trials ((including low quality RCT; e.g., <80% follow-up)
2c: "Outcomes" Research; Ecological studies
3a: SR (with homogeneity*) of case-control studies
3b: Individual Case-Control Study
4: Case-series (and poor quality cohort and case-control studies)
5: Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"

[Response Ends]

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

[Response Begins]
2022 Submission:

1. NCCN Clinical Guidelines in Oncology: Palliative Care. 2021.
    a. Category 2A definition-Based upon lower-level evidence, there is uniform NCCN consensus that the intervention is appropriate.
2. Integration of Palliative Care Into Standard Oncology Care: ASCO Clinical Practice Guideline. 2017.
    a. Type: evidence based, benefits outweigh harms;
    b. Strength of recommendation: strong
        1. Definition of strong-There is high confidence that the recommendation reflects best practice. This is based on (1) strong evidence for a true net effect (eg, benefits exceed harms); (2) consistent results, with no or minor exceptions; (3) minor or no concerns about study quality; and/or (4) the extent of Expert Panelists' agreement. Other compelling considerations (discussed in the guideline's literature review and analyses) may also warrant a strong recommendation.
3. ICSI Health Care Guidelines: Palliative Care for Adults. 2020.
    a. Strength of recommendation (for both recommendations): strong
        i. Definition of strong- In recommendations for an intervention, the desirable effects of an intervention outweigh its undesirable effects. In recommendations against an intervention, the undesirable effects of an intervention outweigh its desirable effects. All or almost all informed people would make the recommended choice for or against an intervention.

4. Associations Among End-of-Life Discussions, Health-Care Utilization, and Costs in Persons With Advanced Cancer: A Systematic Review. 2019.
    a. No grading provided as it is not a clinical guideline recommendation.

[Response Ends]

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

[Response Begins]
2022 Submission:
1.NCCN Clinical Guidelines in Oncology: Palliative Care. 2021.

The guideline grading system is based on the NCCN Categories of Evidence and Consensus.
Category 1- Based upon high-level evidence, there is uniform NCCN consensus that the intervention is appropriate.
Category 2A -Based upon lower-level evidence, there is uniform NCCN consensus that the intervention is appropriate.
Category 2B- Based upon lower-level evidence, there is NCCN consensus that the intervention is appropriate.
Category 3-Based upon any level of evidence, there is major NCCN disagreement that the intervention is appropriate.

2. Integration of Palliative Care Into Standard Oncology Care: ASCO Clinical Practice Guideline. 2017.

The guideline recommendations were crafted, in part, using the GuideLines Into DEcision Support (GLIDES) methodology and accompanying BRIDGE-Wiz software™
Strength of Recommendation Definitions:
Strong- There is high confidence that the recommendation reflects best practice. This is based on (1) strong evidence for a true net effect (eg, benefits exceed harms); (2) consistent results, with no or minor exceptions; (3) minor or no concerns about study quality; and/or (4) the extent of Expert Panelists' agreement. Other compelling considerations (discussed in the guideline's literature review and analyses) may also warrant a strong recommendation.
Moderate- There is moderate confidence that the recommendation reflects best practice. This is based on (1) good evidence for a true net effect (eg, benefits exceed harms); (2) consistent results, with minor and/or few exceptions; (3) minor and/or few concerns about study quality; and/or (4) the extent of Expert Panelists' agreement. Other compelling considerations (discussed in the guideline's literature review and analyses) may also warrant a moderate recommendation.
Weak- There is some confidence that the recommendation offers the best current guidance for practice. This is based on (1) limited evidence for a true net effect (eg, benefits exceed harms); (2) consistent results, but with important exceptions; (3) concerns about study quality; and/or (4) the extent of Expert Panelists' agreement. Other considerations (discussed in the guideline's literature review and analyses) may also warrant a weak recommendation.

3. ICSI Health Care Guidelines: Palliative Care for Adults. 2020.

The guideline recommendations were crafted, in part, using GRADE methodology.
Strength of Recommendation Definitions:
Strong- In recommendations for an intervention, the desirable effects of an intervention outweigh its undesirable effects. In recommendations against an intervention, the undesirable effects of an intervention outweigh its desirable effects. All or almost all informed people would make the recommended choice for or against an intervention.
Weak- In recommendations for an intervention, the desirable effects probably outweigh the undesirable effects, but appreciable uncertainty exists. In recommendations against an intervention, the undesirable effects probably outweigh the desirable effects, but appreciable uncertainty exists. Most informed people would choose the recommended course of action, but a substantial number would not.

4. Associations Among End-of-Life Discussions, Health-Care Utilization, and Costs in Persons With Advanced Cancer: A Systematic Review. 2019.
    a. No grading provided as it is not a clinical guideline recommendation.

[Response Ends]

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

[Response Begins]
2022 Submission:

1. NCCN Clinical Guidelines in Oncology: Palliative Care. 2021.
    a. Quantity: The NCCN guidelines do not provide this information.
    b. Quality: Guidelines utilized the PubMed database to obtain key literature on palliative care. The search results were narrowed by selecting studies in humans published in English. Results were confined to the following article types: Clinical Trial, Guideline, Meta-Analysis, Multicenter Study, Observational Study, Randomized Control Trial, Systematic Reviews, and Validation Studies. Where high level evidence is lacking, recommendations are based on lower-level evidence and the NCCN panels' expert opinion.
2. Integration of Palliative Care Into Standard Oncology Care: ASCO Clinical Practice Guideline. 2017.
    a. Quantity: A total of nine new RCTS, two publications reporting on one large quasiexperimental trial, and five secondary publications based on prior published RCTs. met eligibility criteria and/or were suggested by the Expert Panel and form the evidentiary basis for the guideline recommendations. The identified trials were published between 2011 and 2016.
    b. Quality: Study quality was formally assessed for nine RCTs and one quasiexperimental study identified on palliative care interventions in the cancer population. Assessment of study quality was performed for included evidence by one methodologist. Study design aspects related to individual study quality (eg, randomization method and allocation concealment) and risk of bias were evaluated. Assessment generally indicated low potential risk of bias for most identified evidence.
3. ICSI Health Care Guidelines: Palliative Care for Adults. 2020.
    a. Recommendation 1 (Palliative Care):
        i. Quantity: A total of 2 Systematic Review/Meta-Analysis; 1 Report; 1 Review; 1 Summary; 1 Consensus Report form the evidentiary basis for the guideline recommendation. Literature searches for this guideline were done during the time frame of Jan. 1, 2013, through Dec. 1, 2018.
        ii. Quality: There is broad research supporting the benefit of early, specialized palliative care for patients with advanced solid cancers, indicating that such early involvement (from near the time of diagnosis) improves patient and family quality of life and coping when compared to usual care (Temel, 2017). The American Society of Clinical Oncology urges early integration of specialty palliative care teams into the care of all patients with advanced cancers (Aslakson, 2017), and strongly recommend organizations implement policies and programs to make specialty palliative care available to all patients with advanced cancers from the time of diagnosis.
    b. Recommendation 2 (Hospice Referral):
        i. Quantity: A total of 2 Systematic Reviews; 1 Observation Study; 1 Summary; 1 Controlled Trial form the evidentiary basis for the guideline recommendation. Literature searches for this guideline were done during the time frame of Jan. 1, 2013, through Dec. 1, 2018.
        ii. Quality: Studies included in one systematic review have shown that those who die at home and those enrolled in hospice programs have improved quality of life and symptom control (Brinkman-Stoppelenburg, 2014). Early referral and admission into hospice services allows the patient and family time to get symptoms under control and time to plan with the entire IDT for the next stage of life, reduce the risk for complicated grief and improve overall quality of life. Some patients even experience improved life expectancy (Joseph, 2016).
4. Associations Among End-of-Life Discussions, Health-Care Utilization, and Costs in Persons With Advanced Cancer: A Systematic Review. 2019.
    a. Quantity: Based on review criteria, 20 studies were included. One study was a retrospective analysis of a randomized clinical trial (RCT); one study was non-randomized, intervention-based; and 18 studies were observational. The systematic review included a search of PubMed, Embase, and CINAHL databases to find studies conducted in the United States published from January 1, 2012 to January 8, 2019 that explored relationships between EOL discussions and financial costs, healthcare utilization, or place of death in adults with advanced cancer.
    b. Quality: Of the 11 studies that assessed hospice use, nine studies found significant associations with EOL discussions (ORs ranging 1.79 to 6.88). Findings were strongest among studies that defined discussions based on EOL, goal-of-care (GOC), and treatment preference conversations.
    c. Quality: A couple of studies did measure time between hospice enrollment and death.

i. Lopez-Acevedo and colleagues, for example, found early EOL discussions were associated with significantly more days of hospice care before death (median length of enrollment 53 days vs. 11 days, $p < 0.001$) and a lower likelihood of late enrollment in hospice within three days of death (OR 0.16, $p = 0.02$).
ii. Zakhour and colleagues, whose sample predominantly engaged in GOC discussions but also may have completed advance directives (ADs) or physician orders for life-sustaining treatment (POLST), found patients who had late EOL discussions were eight times as likely to either enroll in hospice within three days of death or not enroll at all (OR 8.0, 95% CI 3.3–19.2, $p < 0.0001$) than those who had an early conversation.

[Response Ends]

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

[Response Begins]
2022 Submission:

1. NCCN Clinical Guidelines in Oncology: Palliative Care. 2021.
    a. The NCCN guidelines does not provide this information.
2. Integration of Palliative Care Into Standard Oncology Care: ASCO Clinical Practice Guideline. 2017.
    a. Strength of Recommendation is strong- There is high confidence that the recommendation reflects best practice. This is based on (1) strong evidence for a true net effect (eg, benefits exceed harms); (2) consistent results, with no or minor exceptions; (3) minor or no concerns about study quality; and/or (4) the extent of Expert Panelists' agreement. Other compelling considerations (discussed in the guideline's literature review and analyses) may also warrant a strong recommendation.
3. ICSI Health Care Guidelines: Palliative Care for Adults. 2020.
    a. Recommendation 1 (Palliative Care) Benefit: Palliative care involvement in patients with serious illness has been shown to improve quality of life and resource utilization. Patients are more likely to receive care that is consistent with their wishes. No information provided on the consistency across studies.
        i. Systematic literature reviews also show that individuals who receive advance care planning or palliative care interventions are less likely to be admitted to intensive care units, and, if admitted, have shorter lengths of stay (Khandelwal, 2015).
    b. Recommendation 2 (Hospice Referral) Benefit: Earlier (hospice) referral provides a better chance of relieving symptoms, improved quality of life, and addressing patient and family needs. No information provided on the consistency across studies.
4. Associations Among End-of-Life Discussions, Health-Care Utilization, and Costs in Persons With Advanced Cancer: A Systematic Review. 2019.
    a. Again the systematic review included 20 studies- One study was a retrospective analysis of a randomized clinical trial (RCT); one study was non-randomized, intervention-based; and 18 studies were observational. Strengths of the studies included clearly-stated objectives and inclusion criteria, sample sizes adequate for meeting objectives, and well-defined outcomes and variables.
    b. There is wide variation in how studies defined EOL discussions. Most studies based EOL discussions on documentation in the medical record or patient/surrogate reports of an EOL conversation with a healthcare provider; and others defined advance care planning (ACP) in terms of documentation of medical orders such as DNR/DNI, POLST, AD, or living will.
    c. There is variation in studies in how clinicians conceptualize aggressive care and overuse of healthcare services near EOL.

[Response Ends]

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

[Response Begins]
2022 Submission:

1. NCCN Clinical Guidelines in Oncology: Palliative Care. 2021.
    a. There were no harms identified.
2. Integration of Palliative Care Into Standard Oncology Care: ASCO Clinical Practice Guideline. 2017.

    a. There were no harms identified by integrating palliative care interventions/services in the studies included. As stated above, the strength of recommendation indicates that the evidence shows a true net effect (eg, benefits exceeds harms).
3. ICSI Health Care Guidelines: Palliative Care for Adults. 2020.
    a. There were no (clinical) harms identified, although it is recognized that the patient's choice regarding hospice may cause conflict among family and providers. However, patients at any time have the opportunity to change their minds about hospice and resume life-prolonging care. Finally, it is noted that access to hospice may be limited.
4. Associations Among End-of-Life Discussions, Health-Care Utilization, and Costs in Persons With Advanced Cancer: A Systematic Review. 2019.
    a. There were no harms identified.

[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]
2022 Submission:
No relevant studies have been conducted and published since the clinical guidelines/systematic review.

[Response Ends]

Group 2 - Evidence - Systematic Reviews Table

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

[Response Begins]
2016 Submission:
Smith TJ, Temin S, Alesi ER, et al. American Society of Clinical Oncology Provisional Clinical Opinion: The Integration of Palliative Care into Standard Oncology Care. J Clin Oncol 2012;30:880-887. Available at: http://www.instituteforquality.org/asco-provisional-clinical-opinion-integration-palliative-care-standard-oncology-care.
Gomes, B., N. Calanzani, et al. (2013). "Effectiveness and cost-effectiveness of home palliative care services for adults with advanced illness and their caregivers." Cochrane Database Rev **6**: CD007760 Available at: http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD007760.pub2/pdf.

[Response Ends]

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

[Response Begins]
2016 Submission:
A 2012 American Society of Clinical Oncology (ASCO) Provisional Clinical Opinion (PCO) addresses the integration of palliative care (PC) services into standard oncology care at the time a person is diagnosed with metastatic cancer and/or high symptom burden.
A 2013 Cochrane Review, 'Effectiveness and cost-effectiveness of home palliative care services for adults with advanced illness and their caregivers', evaluated the impact of home palliative care services on outcomes for adults with advanced illness or their family caregivers, or both. The aim of the review was to quantify the effect of home palliative care services on a patients' odds of dying at home, examine the clinical effectiveness of home palliative care services on other outcomes such as symptom control, quality of life, caregiver distress and satisfaction with care, and comparing resource use and costs associated with these services.

[Response Ends]

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

[Response Begins]
2016 Submission:
2012 ASCO PCO (p. 881):
The American Society of Clinical Oncology (ASCO) has established a rigorous, evidence-based approach—the provisional clinical opinion (PCO)—to offer a rapid response to emerging data in clinical oncology. The PCO is intended to offer timely clinical direction to ASCO's oncologists after publication or presentation of potentially practice- changing data from major studies.
The PCO may serve in some cases as interim direction to the membership pending the development or updating of an ASCO clinical practice guideline. As such, the evidence is not graded in a PCO and is a result of expert consensus. A clinical guideline on palliative care integration with recommendations and the associated grading is under development.
2013 Cochrane Review (p. 12):
Two independent reviewers assessed all included studies for methodological quality using the standard criteria developed by the Cochrane EPOC Review Group for RCTs/CCTs, CBAs and ITSs. The checklist for RCTs/CCTs contains seven items qualified as done, unclear and not done for concealment of allocation, follow-up of professionals, follow up of patients or episodes of care, blinded assessment of primary outcome(s), baseline assessment, reliable primary outcome measure(s) and protection against contamination. Blinding and reliability of all outcomes were also assessed.
Each criterion was scored zero (not done), 0.5 (not clear or when scores varied across outcomes) and one (done). Total scores for RCTs/ CCTs ranged from zero to six; studies with a score of 3.5 or above were considered of high quality. Integration of the results of the quality assessment in data analysis was done in addition to meta-analyses with sensitivity analyses including only high quality studies.

[Response Ends]

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

[Response Begins]
2016 Submission:
See 1a.05 for this information.

[Response Ends]

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

[Response Begins]
2016 Submission:
See 1a.05 for this information.

[Response Ends]

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

[Response Begins]
2016 Submission:
See 1a.05 for this information.

[Response Ends]

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

[Response Begins]
2016 Submission:

2012 ASCO PCO: 2004-2012. This PCO did not provide an assessment of the overall quality of evidence across the studies. This analysis will be completed during the development of the upcoming clinical guideline.
2013 Cochrane Review: 1950 – November 2012
2013 Cochrane Review : p. 3: The direction of the effect was consistent across all studies but did not reach statistical significance in 3; ORs ranged from 1. 36 (95% CI 0.80 to 2.31) to 2. 86 (95% CI 0.78 to 10.53) Sensitivity analyses showed that exclusion of the 2 CCTs (both of Swedish hospital-based services with a pooled OR 3.44, 95% CI 0.60 to 19.57) and inclusion of only high quality RCTs resulted in a reduction of the OR to 1.28 (95% CI 1.28 to 2.33) and 1.75 (95% CI 1.24 to 2. 47) respectively, with more precision and less heterogeneity.
p. 22: Pooled data from seven studies (five RCTs, three of high quality, and two CCTs with 1222 participants) showed that those receiving home palliative care had statistically significantly higher odds of dying at home than those receiving usual care (OR 2.21, 95% CI 1.31 to 3.71; Z = 2.98, P value = 0.003; Chi2 = 20.57, degrees of freedom (df) = 6, P value = 0.002; I2 = 71%). The study population control risk was of 307 home deaths per 1000 deaths; based on this ACR of 0.307, the NNTB was 5 (95% CI 3 to 14), meaning that for one additional patient to die at home five more would need to receive home palliative care as opposed to usual care. Assuming a medium cancer home death rate population ACR of 0.278 (i.e. 278 home deaths per 1000 cancer deaths), the NNTB was 6 (95% CI 3 to 15). This means that for one additional cancer patient to die at home in a population where there are 278 home deaths per 1000 cancer deaths, six more would need to receive home palliative care. NNTB estimates ranged from 9 patients (95% CI 5 to 16) when applied to a low home death rate population such as the one observed in Norway (128 home deaths per 1000 cancer deaths) to 5 patients (95% CI 3 to 13) when applied to a high home death rate population such as the one observed in the Netherlands (454 home deaths per 1000 cancer deaths).

[Response Ends]

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

[Response Begins]
2016 Submission:
2012 ASCO PCO (p. 884):
Seven published randomized trials demonstrate the feasibility of providing various components of PC alongside usual oncology care. There is, however, a dearth of data evaluating the integration of modern PC practices into standard oncology care, especially in concert with ongoing antitumor therapy. Overall, the addition of PC interventions to standard oncology care delivered via different models to patients with cancer provided evidence of benefit.
2013 Cochrane Review (p. 22):
The study population control risk was of 307 home deaths per 1000 deaths; based on this ACR of 0.307, the NNTB was 5 (95% CI 3 to 14), meaning that for one additional patient to die at home five more would need to receive home palliative care as opposed to usual care. Assuming a medium cancer home death rate population ACR of 0.278 (i.e. 278 home deaths per 1000 cancer deaths), the NNTB was 6 (95% CI 3 to 15). This means that for one additional cancer patient to die at home in a population where there are 278 home deaths per 1000 cancer deaths, six more would need to receive home palliative care. NNTB estimates ranged from 9 patients (95% CI 5 to 16) when applied to a low home death rate population such as the one observed in Norway (128 home deaths per 1000 cancer deaths) to 5 patients (95% CI 3 to 13) when applied to a high home death rate population such as the one observed in the Netherlands (454 home deaths per 1000 cancer deaths).

[Response Ends]

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

[Response Begins]
2016 Submission:
2012 ASCO PCO (p. 884-885):
No harm to any patient was observed in any trial, even with discussions of EOL planning, such as hospice and ADs. Two of five trials measuring change in symptoms, two of five studies measuring QOL, two of three studies measuring patient/caregiver satisfaction, and one of three studies measuring survival found statistically significant improvements with PC. Three of six studies measuring mood, two of five studies measuring resource use, and one of four studies measuring outcomes of advance care planning found statistically significant differences, and one outcome of borderline significance was also found in each of these three areas, Therefore, most trials showed benefits ranging from equal to

improved overall survival, reduced depression, improved caregiver and/or patient QOL, and overall lower resource use and cost because EOL hospitalizations were avoided.
2013 Cochrane Review: Discussion of harms was not addressed.

[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]
2016 Submission:
No relevant studies have been conducted and published since the systematic reviews.

[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]
2022 Submission:
1.Family Perspectives on Aggressive Cancer Care Near the End of Life. 2016.
This was a survey of family members of elderly Medicare patients with advanced-stage lung or colorectal cancer in the Cancer Care Outcomes Research and Surveillance study (a multi-regional, prospective, observational study) who died by 2011. In the survey, it provides the family member's rating of the quality of end of life care. It also examined whether patients treated with aggressive care near death received care that was congruent with their preferences.

[Response Ends]

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

[Response Begins]
2022 Submission:
1.Family Perspectives on Aggressive Cancer Care Near the End of Life. 2016.
This study is one of the first empirical validations of these aggressive end-of-life care indicators using patient- and family-centered outcomes in a population-based cohort. Of 1,146 cancer patients (median [IQR] age, 76.0 [65.0-87.0] years, 55.8% male), bereaved family members reported excellent EOL care quality of 51.3%. In this large, diverse population-based cohort, the study found that this measure of aggressive end-of-life care (no or ≤3 days of hospice service) was associated with relatively large differences in family reported end-of-life care quality and a lower likelihood that advanced-stage cancer patients received care congruent with their preferences.
In adjusted analyses, families of patients who received no hospice or ≤3 days of services before death were less likely to report excellent quality end-of-life care than those who received more [43.1% (236/547) vs. 58.8% (352/599); adjusted difference: -16.5 percentage points; 95% CI: -22.4 to -10.7].
Family of patients who enrolled in hospice >3 days before death were much more likely to report that patients died in their preferred place [72.8% (287/394) vs. 40.0% (152/380); adjusted difference: 34.4 percentage points; 95% CI: -41.7 to -27.0].
Among patients with colorectal cancers, those who enrolled in hospice ≤3 days before death were less likely to report that patients' wishes were followed "a great deal" (66.6% vs. 84.4%; adjusted difference=-17.8 percentage points; 95% CI=-28.9 to -6.7; P=0.002).

[Response Ends]

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

[Response Begins]
2022 Submission:

1.Family Perspectives on Aggressive Cancer Care Near the End of Life. 2016.
This study is one of the first empirical validations of these aggressive end-of-life care indicators using patient- and family-centered outcomes in a population-based cohort. An NQF Standing Committee member referenced this study during the 2015-2016 NQF Maintenance review of this measure.

[Response Ends]

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

[Response Begins]
2022 Submission:
1.Wright, A.A., Keating, N.L., Ayanian, J.Z., Chrischilles, E.A., Kahn, K.L., Ritchie, C.S., Weeks, J.C., Earle, C.C., Landrum, M.B. (2016). Family Perspectives on Aggressive Cancer Care Near the End of Life. *JAMA*, 315(3), 284-92. doi: 10.1001/jama.2015.18604. Retrieved from https://pubmed.ncbi.nlm.nih.gov/26784776/

[Response Ends]

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

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

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

[Response Begins]
The Institute of Medicine's report, Dying in America, advocates for measures to improve the quality and sustainability of end-of-life care, urging the federal government to "require public reporting on quality measures, outcomes, and costs regarding care near the end-of-life" (Daly et al., 2016). Hospice care is a form of palliative care for patients with a limited life expectancy of six months or less (according to Medicare hospice coverage criteria) who want to focus on quality of life and comfort rather than life-prolonging care. Most insurance plans provide coverage for hospice (ICSI Guideline, 2020). Currently the median length of stay in hospice before death for Medicare cancer patients is about 19 days (NCCN Guidelines, 2021). Approximately 28% of hospice patients died or were discharged within 7 days of admission to hospice care (NCCN Guidelines, 2021). This is despite the fact that the hospice benefit is at least six months or longer if needed. This short length of stay means that the patient, family and care team have limited time to get a plan of care in place before death and that the focus tends to be more on the care of the imminently dying patient than living life to the fullest before the final decline in function. Early referral to hospice increases the likelihood that pain and other symptoms will be managed more aggressively and therefore there will be less anxiety and distress at the end of life (ICSI Guideline, 2020).
One retrospective study of more than 64,000 patients with cancer who were admitted to hospice found that over 16% of those patients were only enrolled in the last three days of life or less (O'Connor, 2015). The rate of patients who do not have a hospice referral prior to death continues to be higher than desired with one study reporting that more than 30% of patients were not referred and of those patients, only 7% had a documented discussion on the option of palliative care (O'Connor, 2015).
Patients enrolled in hospice experience increased survival times along with a reduction in resource use such as aggressive end of life care and hospital admissions; benefits that increased the longer patients are enrolled in hospice (Lee, 2015; Langton, 2014).- Patients who use hospice, compared with those who do not use hospice, have markedly improved symptoms, less caregiver distress, reduced costs of approximately $8,700 per Medicare beneficiary, and, according to two published reports, actually live longer (ASCO Guideline, 2017).
References:
Daly, B., Hantel, A., Wroblewski, K., Balachandran, J.S., Chow, S., DeBoer, R., Fleming, G.F., Hahn, O.M., Kline, J., Liu, H., Patel, B.K., Verma, A., Witt, L.J., Fukui, M., Kumar, A., Howell, M.D., Polite, B.N. (2016). No Exit: Identifying Avoidable Terminal Oncology Intensive Care Unit Hospitalizations. J Oncol Pract,12(10), e901-e911. doi: 10.1200/JOP.2016.012823. Retrieved from https://pubmed.ncbi.nlm.nih.gov/27601514/
ICSI Health Care Guidelines: Palliative Care for Adults. (2020). Sixth edition. Retrieved from https://www.icsi.org/wp-content/uploads/2021/11/PalliativeCare_6th-Ed_2020_v2.pdf

Ferrell, B. R., Temel, J. S., Temin, S., Alesi, E. R., Balboni, T. A., Basch, E. M., Firn, J. I., Paice, J. A., Peppercorn, J. M., Phillips, T., Stovall, E. L., Zimmermann, C., & Smith, T. J. (2017). Integration of Palliative Care Into Standard Oncology Care: American Society of Clinical Oncology Clinical Practice Guideline Update. *Journal of Clinical Oncology*, *35*(1), 96–112. Retrieved from https://doi.org/10.1200/jco.2016.70.1474
O´Connor, T. L., N. Ngamphaiboon, et al. (2015). "Hospice utilization and end-of-life care in metastatic breast cancer patients at a comprehensive cancer center." J Palliat Med 18(1): 50-55.
Langton, J. M., B. Blanch, et al. (2014). "Retrospective studies of end-of-life resource utilization and costs in cancer care using health administrative data: a systematic review." Palliat Med 28(10): 1167-1196.
Lee, Y. J., J. H. Yang, et al. (2015). "Association between the duration of palliative care service and survival in terminal cancer patients." Support Care Cancer 23(4): 1057-1062.
National Comprehensive Cancer Network Clinical Practice Guidelines in Oncology. (2021). Palliative Care (version 2.2021). Retrieved from https://www.nccn.org/guidelines/guidelines-detail?category=3&id=1454.

[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]
2022 Submission:
Below are performance scores provided by 3 sources where this registry measure has been in use (QOPI Registry and MIPS). The scores indicate that there continues to be a performance gap for this measure, there is variation amongst providers, and there is still an opportunity for improvement.
1.CMS provided 2017 MIPS reporting data for analysis. Based on the 215 included physicians from 2017 MIPS Measure reporting, the mean performance rate is 16.88% the median performance rate is 0%. The standard deviation is 28.14%. The range of the performance rate is 100%, with a minimum rate of 0% and a maximum rate of 100%. The interquartile range is 25% with a 25th percentile of 25% and a 75th percentile of 0%.
a. Performance Scores Quartiles (from 2017 MIPS reporting data for analysis) for NQF 0216 (QPP 457)

N	Minimum	25th Percentile	Median	75th Percentile	Maximum	(Absolute Value) IQR	(Absolute Value) Range
215	100	25	0	0	0	25	100

NQF 0216 (QPP 457) Performance Score Quartiles and Ranges

b. Performance Scores Mean and Standard Deviation (from 2017 MIPS reporting data for analysis) for NQF 0216 (QPP 457)

N	Mean	Standard Deviation	Variance	CI for Mean	Percent Outside CI
215	16.88	28.14	791.6023	(13.00, 21.00)	89%

NQF 0216 (QPP 457) Performance Score Mean and Variation

2. The CMS MIPS Benchmark Reports, published on the QPP website, provide performance rates for PY 2017-2019. CMS did not provide the number of entities measured, the standard deviation, nor the interquartile range in these reports. This is an inverse measure:
MIPS PY 2017-2020 Performance for NQF 0216 (QPP 457)

MIPS Benchmark Report Year	Mean	Standard Deviation	30th Percentile	40th Percentile	50th Percentile	60th Percentile	70th Percentile	80th Percentile	90th Percentile	100th Percentile
2022 (MIPS PY 2020 data)	8.88	Not provided	14.81 - 11.27	11.26 - 8.78	8.77- 5.39	5.38 - 4.16	4.15 - 2.54	2.53 - 1.36	1.35- 0.01	0
2021 (MIPS PY 2019 data)	8.48	Not provided	11.32 - 10.01	10.0 - 8.17	8.16 - 6.99	6.98 - 4.27	4.26 - 3.04	3.03 - 2.23	2.22 - 0.01	0
2020 (MIPS PY 2018 data)	10.83	Not provided	24.9 - 20.52	20.51 - 18.78	18.77- 14.96	14.95 - 10.01	10 - 5.51	5.5 - 3.54	3.53 - 2.89	<= 2.88
2019 (MIPS PY 2017 data)	10.7	7.6	15.38 - 15.01	15.00 - 13.52	13.51 - 11.11	11.10 - 9.65	9.64 - 6.46	6.45 - 0.01	*	0

NQF 0216 (QPP 457) Reported Performance in MIPS PY 2017-2020

*This cell intentionally left blank.

This measure was first implemented into the Merit-based Incentive Payment System (MIPS) program in 2017. It is important to note that the Merit-based Incentive Payment System (MIPS), has been and remains a voluntary reporting program. Participants are allowed to self-select measures and may choose those that will result in high performance rates. As a result, performance rates may not be nationally representative.

3. The measure was implemented in the ASCO Quality Oncology Practice Initiative (QOPI) Registry through PY 2020. Below are the performance rates in QOPI from 2017-2020. QOPI offers two data abstraction rounds per year in which practices select patient charts to report on based on chart selection criteria provided by QOPI. QOPI Round 1 generally spans the first half of the year, January - June, while QOPI Round 2 generally runs from July - December. This is an inverse measure:

ASCO QOPI Registry PY 2017-2020 Performance for NQF 0216 (QOPI EOL 44)

QOPI Measure Performance Report Year	# of Practices	# of Charts	Mean	Standard Dev.	(Absolute Value) IQR	Min	10th percentile	Lower quartile	Median	Upper quartile	90th percentile	Max
2020 Round 1	68	1298	22.84	16.78	17.95	100	43.07	33.33	25.00	15.38	9.69	0
2019 Round 1	86	1412	18.60	15.63	18.43	61.54	43.91	33.33	22.22	14.90	7.33	0
2019 Round 2	171	1797	19.10	*	*	66.67	*	*	*	*	*	0
2018 Round 1	96	1674	17.86	16.15	14.8	100	41.73	28.00	18.18	13.20	9.71	0
2018 Round 2	101	1796	19.11	15.70	14.28	66.67	49.74	28.57	19.14	14.29	10.00	0
2017 Spring	154	2630	16.85	15.15	13.93	100	33.33	26.43	18.40	12.50	8.00	0
2017 Fall	154	2419	20.29	20.87	18.49	100	49.59	31.13	20.00	12.64	8.33	0

NQF 0216 (QOPI EOL 44) Reported Performance in ASCO QOPI Registry PY 2017-2020

*This cell intentionally left blank.

Previous Submission:

This data was produced from the QOPI® registry and data was abstracted for a sample of patients seen with the data collection period. Data is reported at the practice level.
In 2013, 178 practices reported on 4951 charts.
In 2014, 170 practices reported on 5021 charts.
In 2015, 222 practices reported on 7239 charts.

_____	2013	2014	2015
Total Patient Population (%)	10.91	16.79	16.95
Mean	16.63	18.22	17.86
Minimum	0	0	0
Maximum	100	100	100
Standard Deviation	16.46	17.60	14.50
Percentiles			
10	0	0	0
25	4.76	7.14	7.14
50	12.97	14.64	15.38
75	25.00	23.81	25.81
90	36.36	38.28	33.33
95	50	50.00	46.67

[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]
2022 Submission:
The literature noted below supports that there is still an existing performance gap for this measure.

- One retrospective study of more than 64,000 patients with cancer who were admitted to hospice found that over 16% of those patients were only enrolled in the last three days of life or less (O'Connor, 2015). The rate of patients who do not have a hospice referral prior to death continues to be higher than desired with one study reporting that more than 30% of patients were not referred and of those patients, only 7% had a documented discussion on the option of palliative care (O'Connor, 2015).

Reference:
O´Connor, T. L., N. Ngamphaiboon, et al. (2015). "Hospice utilization and end-of-life care in metastatic breast cancer patients at a comprehensive cancer center." *J Palliat Med,* 18(1), 50-55. Retrieved from https://www.researchgate.net/publication/267744949_Hospice_Utilization_and_End-of-Life_Care_in_Metastatic_Breast_Cancer_Patients_at_a_Comprehensive_Cancer_Center

[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]
2022 Submission:
While this measure is included in the MIPS federal program, CMS has not yet made sociodemographic data available for their quality measures to developers to analyze for disparities.

To advance health equity, the ASCO's measure development team plans to look more closely at disparities in care among oncology patient populations. This includes bringing transparency to where gaps exist (or do not exist). We are currently evaluating the ASCO programs that use our measures and are assessing the feasibility of adding the required collection of race and ethnicity. To ensure clinicians were not overly burdened, only the data elements needed to calculate the measure have been required, thus the non-required data elements (race and ethnicity) have only been sparsely populated.

Previous Submission:

This data was produced from the QOPI® registry and data was abstracted from a sample of patients seen during the data collection period. Data is calculated at the chart level as demographics are not currently calculated at the practice level.
In 2013, 178 practices reported on 4951 charts.
In 2014, 170 practices reported on 5021 charts.
In 2015, 222 practices reported on 7239 charts.
Total Measure
Population 10.91 16.79 16.95
Female 10.56 14.86 16.38
Male 11.24 18.57 17.48
Hispanic 10.91 25.84 19.83
White 12.02 16.18 17.08
Black 7.89 19.66 21.33
Other 10.61 13.79 13.87

[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]

2022 Submission:

The literature noted below supports that there are potentially disparities for this measure.

- One example of a disparity was found in a 2010 prospective longitudinal cohort study of black and white patients with advanced cancer conducted by Mack et al, who examined the differences in how patients and physicians communicate about EOL care by race. The investigation was based on the observation that some black patients receive more life-prolonging EOL care than some white patients. (ASCO Guideline, 2017). Mack and colleagues found that compared to White patients, Black/African American patients (OR 0.37, 95% CI 0.19 to 0.73) and Hispanic patients (OR 0.29, 95% CI 0.12 to 0.73) were less likely to experience EOL discussions (p = 0.005) (Starr et al, 2019)
- Another study showed dissemination of information about palliative and EOL care may not reach some black audiences in the United States at all. Fishman et al performed a content analysis of cancer news intended for black audiences or nonspecific audiences to find discussions of palliative care–related topics, including adverse events, EOL care, or palliative or hospice care. The authors found statistically significantly lower reporting on adverse events, treatment failure, and death and dying in black American media. The authors found zero articles on palliative care or hospice in the media directed at black audiences. (ASCO Guideline, 2017)
- Studies have shown that patients covered by Medicaid in the United States have not received guideline- or quality-adherent palliative care (eg, receipt of chemotherapy at EOL; v those with Medicare). (ASCO Guideline, 2017)
- A 2020 retrospective cohort study analyzed the degree of aggressive end-of-life care among 349 adult Medicaid beneficiaries diagnosed between 2011 to 2015 with stage IV breast (30%) and colorectal (70%) cancer, and who died by January 2016. Data was obtained from a New Jersey State cancer Registry-Medicaid claims linked data set. The study found racial and ethnic disparities in aggressive end of life care, after adjusting for sociodemographic and clinical factors. Aggressive end of life care measured in the study include: >1 hospitalization; >1 ED visit; were admitted to the ICU in the last 30 days of life; or received chemotherapy (34%) in the last 14 days of life (Yang et al., 2020).
    - The study data support the hypothesis that a higher proportion of patients with cancer enrolled in Medicaid receive aggressive EOL care compared with Medicare or privately insured populations.

- Compared with Non-Hispanic White patients, Black patients had higher odds of receiving any aggressive EOL care (odds ratio [OR], 1.87; 95% CI, 1.07 to 3.26).

References:

Ferrell, B. R., Temel, J. S., Temin, S., Alesi, E. R., Balboni, T. A., Basch, E. M., Firn, J. I., Paice, J. A., Peppercorn, J. M., Phillips, T., Stovall, E. L., Zimmermann, C., & Smith, T. J. (2017). Integration of Palliative Care Into Standard Oncology Care: American Society of Clinical Oncology Clinical Practice Guideline Update. *Journal of Clinical Oncology*, *35*(1), 96–112. Retrieved from https://doi.org/10.1200/jco.2016.70.1474

Starr, L. T., Ulrich, C. M., Corey, K. L., & Meghani, S. H. (2019). Associations Among End-of-Life Discussions, Health-Care Utilization, and Costs in Persons With Advanced Cancer: A Systematic Review. The American journal of hospice & palliative care, 36(10), 913–926. https://doi.org/10.1177/1049909119848148

Yang, A., Goldin, D., Nova, J., Malhotra, J., Cantor, J. C., Tsui, J. (2020). Racial Disparities in Health Care Utilization at the End of Life Among New Jersey Medicaid Beneficiaries With Advanced Cancer. *JCO Oncology Practice*, 16(6), e538–e548. Retrieved from https://doi.org/10.1200/jop.19.00767

[Response Ends]

Criteria 2: Scientific Acceptability of Measure Properties

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

---

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

[Response Begins]
Yes
[Yes Please Explain]
Minor update to the measure title and description had been made. “Proportion” was replaced by “Percentage”, although the measure intent and the measure calculation remains the same.

[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]
Minor update to the measure title and description had been made. “Proportion” was replaced by “Percentage”, although the measure intent and the measure calculation remains the same.
[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]
Percentage of patients who died from cancer admitted to hospice for less than 3 days
[Response Ends]

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

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

[Response Begins]
Percentage of patients who died from cancer, and admitted to hospice and spent less than 3 days there
[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
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: Transitions of Care
Other (specify)
[Other (specify) Please Explain]
Effective Clinical Care

[Response Ends]

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

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

- *Populations at Risk: Populations at Risk*

[Response Begins]
Adults (Age >= 18)
Elderly (Age >= 65)
[Response Ends]

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

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

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

[Response Begins]
Clinician: Group/Practice
Clinician: Individual
[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]
Ambulatory Care
Outpatient Services
[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]
https://qpp-cm-prod-content.s3.amazonaws.com/uploads/1690/2022+Clinical+Quality+Measure+Specifications+and+Supporting+Documents.zip
[Response Ends]

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

Attach an excel or csv file; if this poses an issue, contact staff. Provide descriptors for any codes. Use one file with multiple worksheets, if needed.
[Response Begins]
Available in attached Excel or csv file
[Response Ends]

sp.12. State the numerator.

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

[Response Begins]
Patients who died from cancer and spent fewer than three days in hospice.
[Response Ends]

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

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

[Response Begins]
Individuals in the denominator who died from cancer and spent fewer than 3 days in hospice will be counted in the numerator. This measure is to be submitted a minimum of once per performance period for patients who died of cancer during the measurement year (January 1-December 31).
Numerator Instructions: INVERSE MEASURE- A lower calculated performance rate for this measure indicates better clinical care or control. The "Performance Not Met" numerator option for this measure is the representation of the better clinical quality or control. Submitting that numerator option will produce a performance rate that trends closer to 0%, as quality increases. For inverse measures, a rate of 100% means all of the denominator eligible patients did not receive the appropriate care or were not in proper control.
Numerator Options:
Performance Met: Patient spent less than three days in hospice care (**G9860**)
OR
Performance Not Met: Patient spent greater than or equal to three days in hospice care (**G9861**)

[Response Ends]

sp.14. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]
Patients who died from cancer who were admitted to hospice
[Response Ends]

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

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

[Response Begins]
This measure is to be submitted a minimum of once per performance period for patients who died of cancer who were admitted to hospice during the measurement year (January 1- December 31). It is anticipated that eligible clinicians who provide services for patients with the diagnosis of cancer will submit this measure.
Denominator Criteria (Eligible Cases):
Diagnosis for cancer (ICD-10-CM): Refer to attached data dictionary (Excel file).
AND
At least two patient encounters during performance period (CPT): 99202, 99203, 99204, 99205, 99212, 99213, 99214, 99215
WITHOUT
Telehealth Modifier: GQ, GT, 95, POS 02
AND
Patients enrolled in hospice: G9858
AND
Patients who died from cancer: G9852

[Response Ends]

sp.16. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]
None
[Response Ends]

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

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

[Response Begins]
Not applicable
[Response Ends]

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

Include the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate. Note: lists of

individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format in the Data Dictionary field.

[Response Begins]
Not applicable
[Response Ends]

sp.19. Select the risk adjustment type.

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

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

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

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

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

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

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

[Response Begins]
Performance is calculated as:
1. Identify those patients that meet the denominator criteria defined in the measure.
2. Subtract those patients with a denominator exclusion from the denominator. Note: this measure does not have any denominator exclusions
3. From the patients who qualify for the denominator (after any exclusions are removed), identify those who meet the numerator criteria.
4. Calculation: Numerator/Denominator-Denominator Exclusions
[Response Ends]

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

[Response Begins]
[Response Ends]

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

[Response Begins]
Registry Data

[Response Ends]

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

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

[Response Begins]
Not applicable
[Response Ends]

sp.30. Provide the data collection instrument.

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

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

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

[Response Begins]
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]
Yes
[Response Ends]

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

[Response Begins]
No
[Response Ends]

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

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

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

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

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

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

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

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

OR

- rationale/data support no risk adjustment/ stratification.

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

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

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

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

2a. Reliability

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

[Response Begins]
Registry Data
[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]
2022 Submission:
The 2021 submission includes reliability testing conducted using 2017 Physician Quality Reporting System (PQRS) registry performance data. The measure has now transitioned into the Merit-based Incentive Payment System (MIPS).
Previous Submission:
ASCO engaged the Virginia Quality Health Center to conduct an inter-rater reliability study of the Quality Oncology Practice Initiative (QOPI) case report form and measures in 2008. Dataset produced during this initiative was used for reliability testing.

[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]
2022 Submission:
01/01/2017 – 12/31/2017
Previous Submission:
01/01/2008 – 12/31/2008

[Response Ends]

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

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

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

[Response Begins]
Clinician: Group/Practice
[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]
2022 Submission:
Reliability testing was conducted using 2017 PQRS registry performance data provided from CMS. Providers were identified by NPIs, and the 2017 dataset provided performance information on 215 NPIs. Additional descriptive characteristics of the measured providers, such as size and location type, are unknown. Entities submitted data for inclusion in this data set according to the eligibility and reporting requirements for PQRS during the 2017 program year. We were unable to determine from our rolled-up data sample the number of clinicians who reported to PQRS as an

individual or a group; therefore, this measure should be considered for endorsement at the group/practice level, with a potential group size as n of 1 or group of 1, so the measure could be utilized at the individual level as well.
Previous Submission:
Ninety-six (96) practices reported this measure. Data from 786 patient records were submitted for this measure. QOPI measure analytics at the practice level were generated. Practices with fewer than 5 records were not included in the calculations.

[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]
2022 Submission:
Reliability testing using 2017 PQRS registry performance data provided by CMS was conducted on 1200 denominator-eligible patients. Additional descriptive characteristics of the measured patients are unknown since they were not provided in the CMS created dataset. Eligible patients were included in this dataset according to the reporting requirements for the 2017 PQRS program year.
Previous Submission:
Ninety-six (96) practices reported this measure. Data from 786 patient records were submitted for this measure. QOPI measure analytics at the practice level were generated. Practices with fewer than 5 records were not included in the calculations.

[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]
2022 Submission:
Reliability Testing Data/Sample:
Reliability testing was conducted using 2017 PQRS registry performance data provided from CMS. Providers were identified by NPIs, and the 2017 dataset provided performance information on 215 NPIs.
Validity Testing Data/Sample:
A subset of 12 providers from the 2017 PQRS registry performance dataset of 215 providers was utilized for measure validity analysis. These 12 providers reported on both Measure 457 - Proportion Admitted to Hospice for less than 3 days (NQF 216) and Measure 453 - Proportion Receiving Chemotherapy in the Last 14 Days of Life (NQF 210), which was correlated with Measure 457.
Previous Submission:
Different testing approaches as well as data sets were used for reliability and validity analyses.

[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]
2022 Submission:
Patient data for social risk factors was not available to perform an analysis.
Previous Submission:

Analysis of social factors and their impact on measure performance was not required.

[Response Ends]

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

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

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

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

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

[Response Begins]
2022 Submission:
Reliability of the computed measure scores was determined using the signal to noise analysis. The signal is the proportion of the variability in measured performance scores that can be explained by real differences in provider performance and the noise is the total variability in measured performance. Reliability is calculated as the ratio of the provider-to-provider variance divided by the sum of the provider-to-provider variance plus the error variance specific to a provider. A reliability approaching zero implies that all the variability in performance scores is attributable to measurement error. A reliability equal to one implies that all the variability is attributable to real differences in provider performance. A reliability of 0.70 – 0.80 is generally considered the acceptable threshold for reliability, 0.80 – 0.90 is considered high reliability, and 0.90 – 1.0 is considered very high.[1]
To assess signal-to-noise, we employed the beta-binomial model as described by JL Adams[1]. Two hundred fifteen (215) unique providers reported numerators and denominators in accordance with the measure specifications. Through the estimation of the beta-binomial parameters (often referred to as alpha and beta) as described by Adams[1], we estimated the provider-to-provider variance and the within provider variance.
1.) Adams JL, Mehrotra A, McGlynn EA, Estimating Reliability and Misclassification in Physician Profiling, Santa Monica, CA: RAND Corporation, 2010. www.rand.org/pubs/technical_reports/TR863
Previous Submission:
2008 IRR study: ASCO engaged the Virginia Quality Health Center to conduct an inter-rater reliability study of the QOPI case report form and measures. Trained, independent nurse abstractors served as the 'gold standard' against which practice abstractions were compared for accuracy. Sampling is described above. The 264 sampled records allowed for reliability analysis at a 95% confidence level with a +/- 3.88% marking of error.
Kappa statistics were used to analyze the reliability of the audit data set compared to the submitted data. Kappa statistics are the commonly accepted standard for determining inter-rater reliability in the healthcare setting1 (Allison, Calhoun, et al, 2000). The Kappa statistic is conceptually similar to the rate of agreement between two reviewers, but it imposes a more stringent standard than simple agreement and mismatch rates. The following standards were used 2,3 (Sim and Wright, 2005; Cohen, 1960):

Kappa > .0.75 denotes excellent reliability,
Kappa between 0.40 and 0.75 denotes good reliability, and
Kappa less than 0.40 denote marginal reliability.

1.) Allison, J. J., Wall, T. C., Spettell, C. M., Calhoun, J., Fargason, C. A., Kobylinski, R. W., Farmer, R., & Kiefe, C. (2000). The art and science of chart review. The Joint Commission Journal on Quality Improvement, 26(3), 115–136. https://doi.org/10.1016/s1070-3241(00)26009-4
2.) Sim, J., & Wright, C. C. (2005). The Kappa statistic in Reliability Studies: Use, interpretation, and sample size requirements. Physical Therapy, 85(3), 257–268. https://doi.org/10.1093/ptj/85.3.257

3.) Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37–46. https://doi.org/10.1177/001316446002000104

[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]
2022 Submission:
Score-Level Reliability

N	Mean	Standard Deviation	Minimum	25th Percentile	Median	75th Percentile	Maximum
215	0.7921	0.3082	0.1099	0.5302	1	1	1

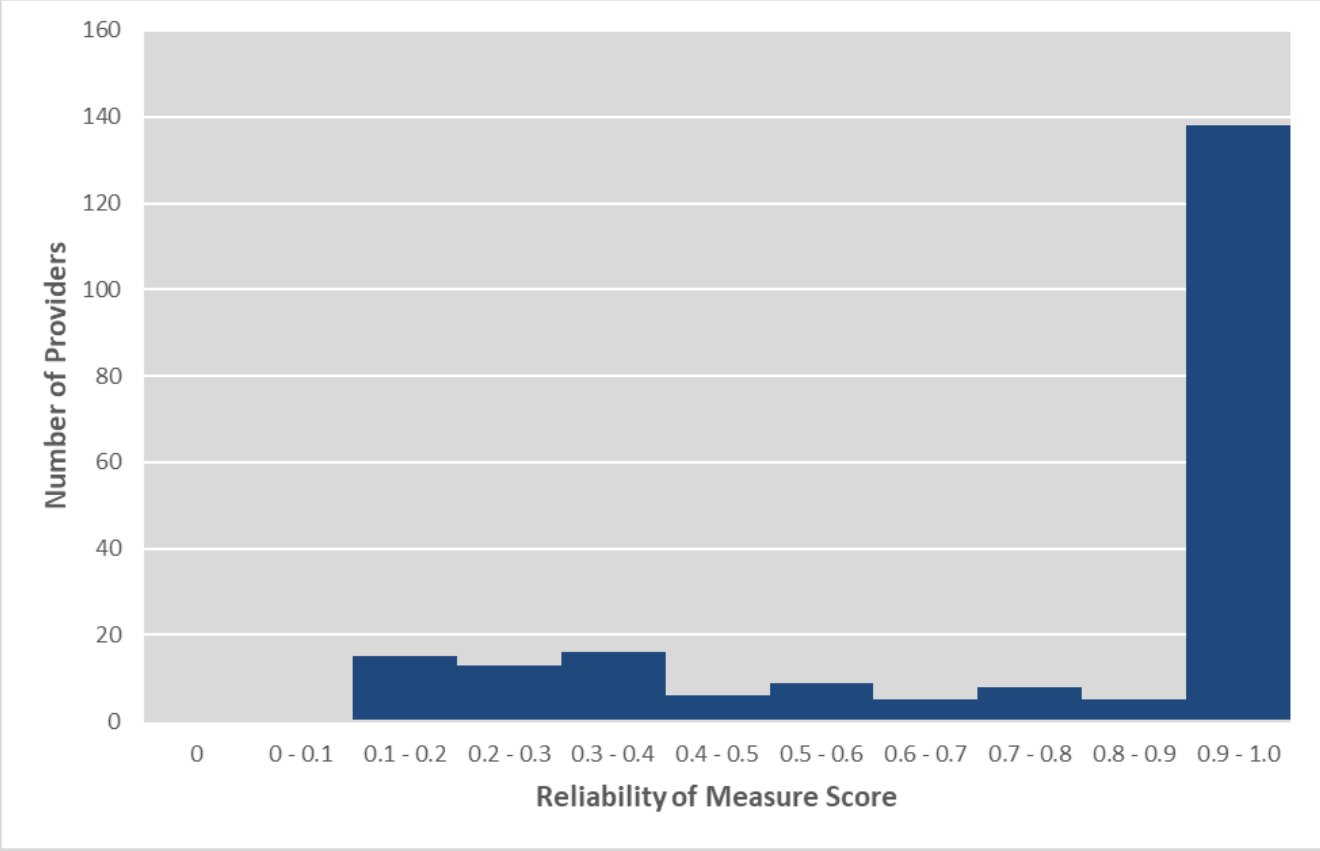
Score-Level Reliability Statistics for NQF 216 (QPP 457)

Beta-Binomial Model Parameters

Mu	Alpha	Beta
0.1590	1.2186	6.4476

Parameters of the Beta-Binomial Model

Distribution of Score-Level Reliability among Providers



Distribution of Reliability for All Providers

Previous Submission:
2008 IRR study: A sample of 300 charts was planned for re-abstraction in four geographic locations: Midwest, Northeast, South, and West. Within each region, the practices are within proximity of one another. Forty-four (44) practices submitted data, resulting in 264 charts in the audit data set (6 charts per practice). The table lists practice counts and the number of charts by region.

Region	Number of Practices	Number of Charts
Midwest	10	60
Northeast	7	42
South	16	96
West	11	66

Number of Practices and Charts by Region

The audit data set sample was randomly selected from the population of 13,561 charts in the spring 2007 submission. The 264 sampled charts allowed VHQC to perform its reliability analysis at a 95% confidence level within a +/- 3.88% margin of error. That ensures that if, for example, a calculated statistical rate is 78%, there is 95% confidence that the true value lies within the range of +/- 3.88%.

VHQC computed Kappa statistics for the 52 measures and for the 97 questions with discrete values. If the measure had no value because the charts sampled did not address the measure, it was assigned a value of "." indicating "Not Applicable". Note that k statistics values range from 1 to –1.

For the ***percentage of patients who died from cancer admitted to hospice for less than 3 days*** measure, the overall sample of 264 encounters showed 55.13% agreement for all data elements and data element combinations assessed. Kappa values ranged from a low of 0.51846 for the numerator to a high of 0.58414 for the denominator. The numerator sensitivity reflects the inconsistent documentation of a patient's enrollment into hospice. The denominator sensitivity reflects the inconsistent documentation of a patient's death in the EHR.

[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]
2022 Submission:
Overall measure reliability is acceptable and very close to the threshold for high reliability. The mean reliability of 215 providers reporting on the measure is 79%. Half of providers reporting on the measure have reliability of 100%.
Previous Submission:
2008 IRR study: The Kappa in the inter-rater reliability study of the QOPI program was 0.551. This Kappa value indicates good reliability.

[Response Ends]

2b. Validity

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

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

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

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

[Response Begins]
2022 Submission:
To demonstrate concurrent validity, a bivariate correlation analysis was performed to evaluate the strength of the association between two measures. The value of the correlation coefficient calculated through this analysis varies between +1 and -1. A value of +1 or -1 indicates a perfect degree of association between the two measures. As the correlation coefficient value nears 0, the relationship between the two measures weakens. The direction of the

relationship is indicated by the sign of the coefficient. A plus (+) sign indicates a positive relationship and a minus (-) sign indicates a negative relationship.
ASCO hypothesized that a positive association exists between Measure 453 - Proportion Receiving Chemotherapy in the Last 14 Days of Life (NQF 210) and Measure 457 - Proportion Admitted to Hospice for less than 3 days (NQF 216) due to the similarities in both the domain of the quality action and patient populations. ASCO performed a Pearson correlation analysis using 2017 PQRS datasets for measures 453 and 457. A Pearson correlation coefficient was calculated to evaluate the association between performance scores of NPIs who had scores on both measures.
The following criteria were used to evaluate the strength of the correlation[1]:

Correlation Coefficient	Interpretation
±1	Perfect
±0.9, ±0.8, ±0.7	Strong
±0.6, ±0.5, ±0.4	Moderate
±0.3, ±0.2, ±0.1	Weak
0	Zero

Correlation Coefficient Interpretation Criteria

1. Akoglu H. (2018). User's guide to correlation coefficients. Turkish journal of emergency medicine, 18(3), 91–93. https://doi.org/10.1016/j.tjem.2018.08.001

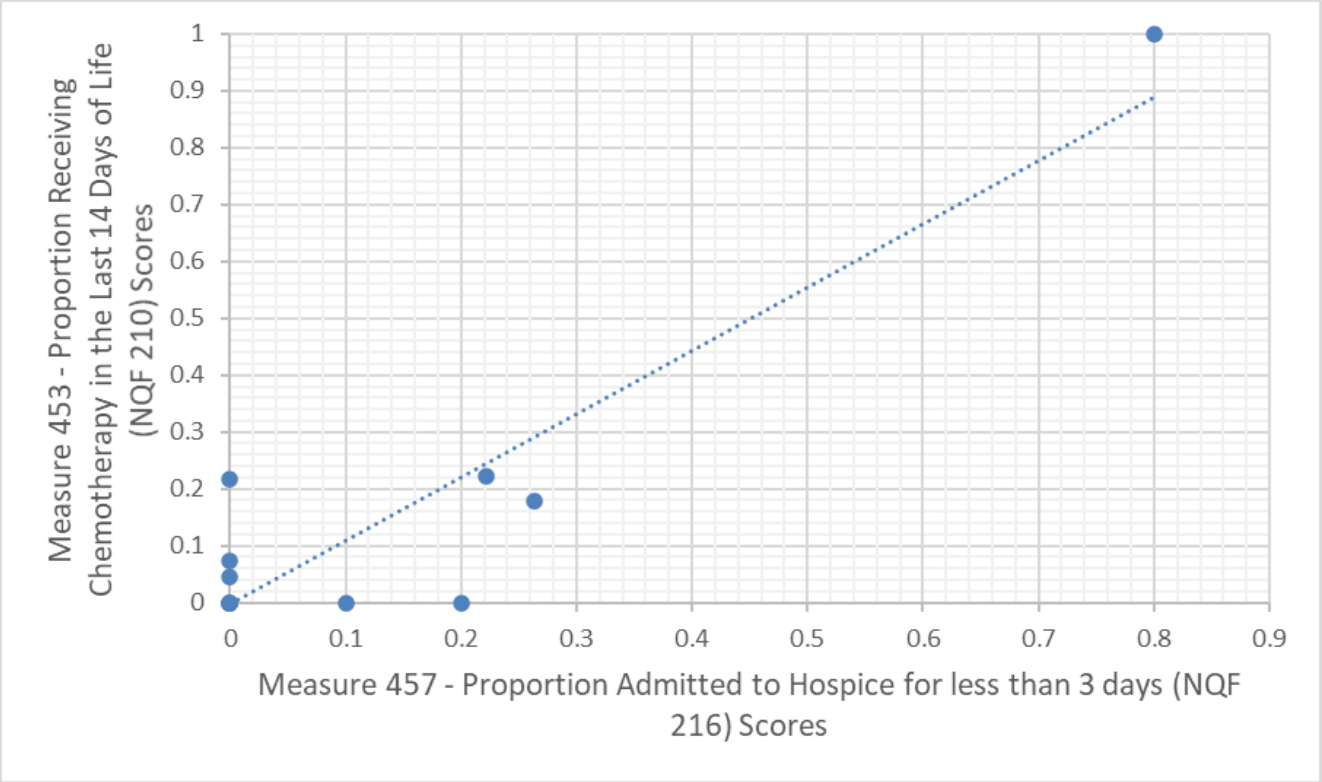
Previous Submission:
Face validity testing was accomplished by first seeking input from patients and their caregivers and then consulting a panel of clinical experts to put patients' stated preferences for care into action. ASCO-led focus groups and structured interviews with end-of-life cancer patients and bereaved caregivers collected patients' feedback regarding their care. Over 40 patients and their caregivers were surveyed and asked questions about their desire to avoid overly aggressive treatment, when to stop treatment, when to start hospice, and desire to avoid emergency room/hospital visits as much as possible. Patient responses were structured into quality actions, which were then vetted by an expert panel of 12 cancer subject matter experts (5 oncologists, 2 researchers, a hospice physician, a hospitalist, a nurse, a social worker, and a patient representative). The panel survey explicitly asked whether the scores obtained from the measures as specified will provide an accurate reflection of quality and can be used to distinguish good and poor quality. The idea that cancer patients with terminal illness do not benefit from heroic attempts at life prolongation, and that such attempts often reflect a lack of honest discussion with patients, is summed up by the statement of one of the oncologists on the expert panel who said, 'for most of our patients, a trip to the ICU is a kind of failure.'

[Response Ends]

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

Examples may include correlations or t-test results.

[Response Begins]
2022 Submission:



Correlation between Measure 457 (NQF 216) and Measure 453 (NQF 210)

Correlation Coefficient R	0.9158
R Square	0.8389
P-value	2.9×10^{-5}
Number of Providers	12

Correlation Statistics

Previous Submission:

The face validity survey asked 12 subject matter experts about the appropriateness of the measure components (denominator and numerator), given the intent of this measure. For each measure component, respondents indicated the extent to which they agreed with the stated specifications of the measure component on a 5-point Likert scale (5 = Strongly agree; 4 = Agree; 3 = Neutral; 2 = Disagree; 1 = Strongly disagree).

Agreement on the validity of the measure was high. One hundred (100) percent of subject matter experts 'strongly agreed' or 'agreed' that the measure specifications are appropriate and align with current evidence. Ninety-two (92) percent of subject matter experts 'strongly agreed' or 'agreed' that the performance score for the measure is meaningful, understandable, and useful for public reporting. Finally, the face validity survey results revealed that 100% of respondents believe that the scores obtained from the measure as specified can be used to distinguish good from poor quality. The agreement with the measure was calculated by averaging the answers to these three questions for a general agreement of 97%.

[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]

2022 Submission:

The results of the bivariate correlation indicate a strong positive relationship between Measure 457 - Proportion Admitted to Hospice for less than 3 days (NQF 216) and Measure 453 - Proportion Receiving Chemotherapy in the Last 14 Days of Life (NQF 210). This strong correlation demonstrates the criterion validity of the measure.

Previous Submission:

Face validity survey results revealed that 97% of respondents ‘strongly agree’ or ‘agree’ that this measure provides an accurate reflection of quality and can be used to distinguish good and poor quality.

[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]
2022 Submission:
The analysis of meaningful differences in performance was analyzed using calculations of several descriptive statistics, including the minimum, maximum, 25th and 75th percentile, median, IQR, and range. Additionally, we calculated the standard deviation, standard error of the mean performance, and a 95% confidence interval for the mean performance. Finally, we calculated the percent of facilities whose performance was statistically significantly different from the overall performance mean.
Previous Submission:
Benchmarks were established to identify the outlying 10th decile of practice: The proportion of patients experiencing each process of care in each Health Care Service Area (HCSA) was computed and ranked from best (least aggressive) to worst. A new cohort was created by sequentially adding HCSAs in order starting with the least aggressive until they contained at least 10% of the original cohort and the proportion experiencing each process of care was then recalculated to arrive at the 'Achievable Benchmark of Care). More detail on this, as well as a reference for the Achievable Benchmark of Care method can be found in our publication: Earle CC, Neville BA, Landrum ME, Souza JE, Weeks JC, Block SD, Grunfeld E, Ayanian JZ. Evaluating claims-based indicators of the intensity of end-of-life cancer care. Int J Qual Health Care. 2005;17(6):505-9.

[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.

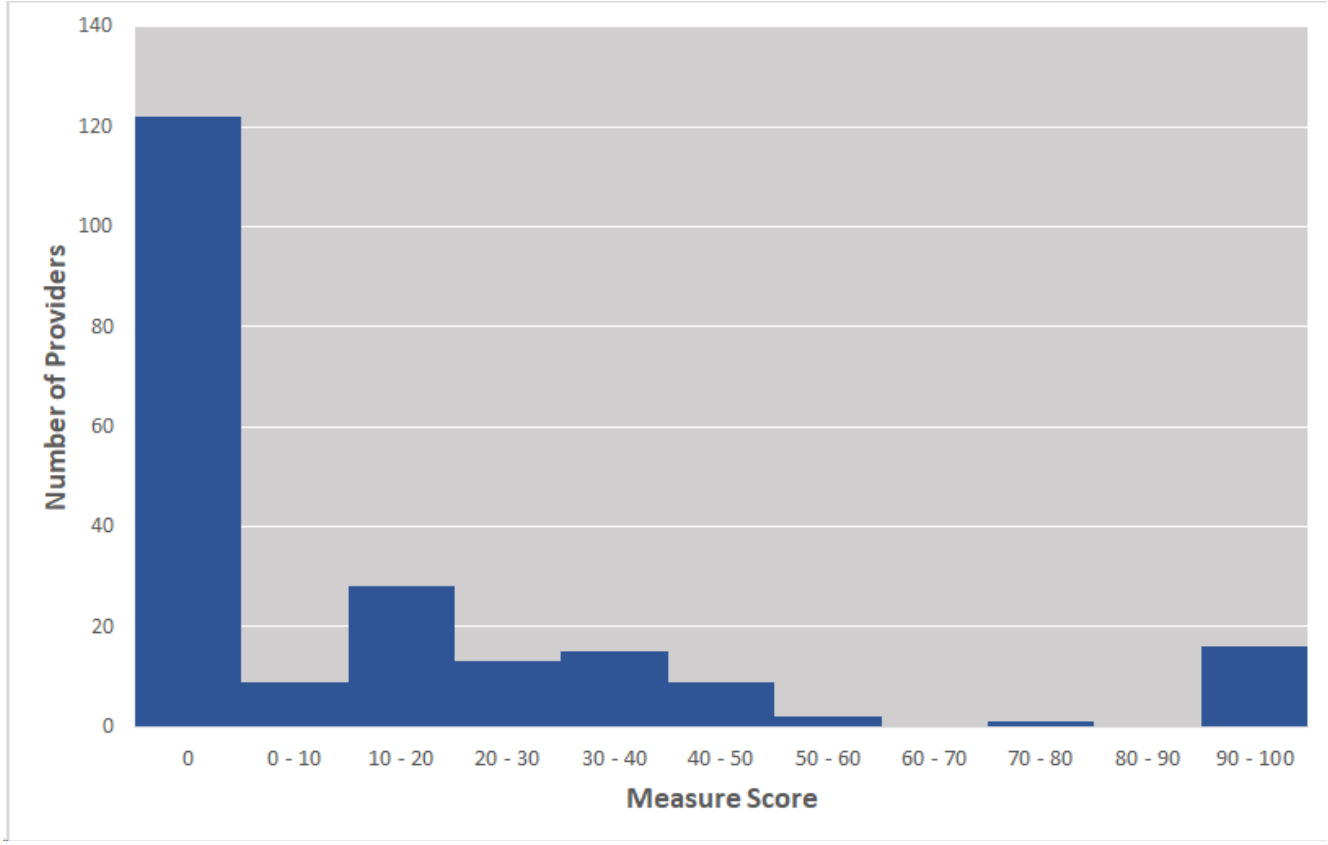
[Response Begins]
2022 Submission:
Performance Scores Quartiles

N	Minimum	25th Percentile	Median	75th Percentile	Maximum	(Absolute Value)* IQR	(Absolute Value)* Range
215	100	25	0	0	0	25	100

NQF 216 (QPP 457) Performance Score Quartiles and Ranges
* The absolute value was calculated since this measure is an inverse measure (lower score is better).
Performance Scores Mean and Standard Deviation

N	Mean	Standard Deviation	Variance	CI for Mean	Percent Outside CI
215	16.88	28.14	791.86	(13, 21)	89%

NQF 216 (QPP 457) Performance Score Mean and Variation
Distribution of Performance among Providers



Distribution of Performance for All Providers

The distribution of performance scores from 215 providers is highly skewed and leptokurtic with the largest number of providers reporting a perfect score of 0%.

Previous Submission:

Performance Scores Quartiles and Percentiles for Years 2013-2015

Year	Mean	Standard Deviation	Minimum	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile	Maximum
2013	11.47	11.87	100	24.26	15.81	9.88	3.35	0	0
2014	12.92	12.58	100	21.88	17.07	11.45	5.80	0	0
2015	13.16	11.50	100	23.08	16.67	11.95	6.45	0	0

NQF 216 (QPP 457) Performance Score Quartiles and Percentiles for Years 2013-2015

A benchmark target of < 8% of patients enrolled in hospice within only the last 3 days of life corresponds to that achieved by the highest performing regions in the country.

[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]

Submission 2022:

While slightly more than half of providers perform perfectly on the measure, the relatively wide variability among the rest of the providers (with 7% failing the measure) shows that there still exists ample room for quality improvement. The range of performance suggests there's clinically meaningful variation in physicians' performance.

Previous Submission:

The range of performance suggests there's clinically meaningful variation in physicians' performance.

[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]
2022 Submission:
The PQRS dataset provided to us from the 2017 program year did not contain missing data, so missing data testing was not performed. Due to data completeness requirements, we suspect that missing data would have been rejected when submitted to CMS, in which case those values would not be counted towards measure performance. While data that may have been missing prior to submission to CMS is unknown and therefore precluded any analysis, there is no indication that this missing data was systematic, thus their omission would lead to unbiased performance results.

[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]
2022 Submission:
Missing data testing was not performed for this measure as there was no missing data.

[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]
2022 Submission:
Missing data testing was not performed for this measure as there was no missing data.

[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]
Not applicable
[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]
Not applicable
[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]
Not applicable
[Response Ends]

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

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

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

[Response Begins]
[Response Ends]

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

[Response Begins]
2022 Submission:
The measures are used for comparison among similar providers and there is no expectation that performance will be 0%, and comorbidity risks will, if anything, decrease the likelihood of experiencing this process of care. Moreover, ASCO prefers that measures are not risk adjusted for patient factors that could possibly obscure disparities (namely age, sex, and socioeconomic status).
ASCO will continue to explore whether risk adjustment is appropriate, and if so, how it could be accomplished for our outcome measures. ASCO's palliative care measures would be included in that assessment. However, patient data availability remains a challenge for registry measures, and collecting this data is placing an additional burden on practices reporting on these measures.

[Response Ends]

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

[Response Begins]
[Response Ends]

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

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

[Response Begins]
[Response Ends]

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

[Response Begins]
[Response Ends]

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

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

[Response Begins]
[Response Ends]

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

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

[Response Begins]
[Response Ends]

2b.27. Provide risk model discrimination statistics.

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

[Response Begins]
[Response Ends]

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

[Response Begins]
Not applicable.
[Response Ends]

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

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

[Response Begins]
[Response Ends]

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

[Response Begins]
[Response Ends]

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

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

[Response Begins]
[Response Ends]

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

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

[Response Begins]
[Response Ends]

Criteria 3: Feasibility

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

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

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

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

In other words, indicate whether data elements that are needed to compute the performance measure score are in defined, computer-readable fields.
[Response Begins]
ALL data elements are in defined fields in electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)
[Response Ends]

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

[Response Begins]
All data elements needed for this measure are collected through electronic data or using keyword searches.
[Response Ends]

3.04. Describe any efforts to develop an eCQM.

[Response Begins]
ASCO is in the process of assessing the feasibility of developing an electronic clinical quality measure.
[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]
Apart from the lack of availability of disparities data for analysis, we have not identified any areas of concern or made any modifications as a result of testing and operational use of this measure in relation to data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, or other feasibility issues unless otherwise noted.
[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]
ASCO requests interested parties seek a licensing agreement prior to commercial use of this measure.
[Response Ends]

Criteria 4: Use and Usability

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

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making.
NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement, in addition to demonstrating performance improvement.

4a. Use

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

Name of program and sponsor

URL

Purpose

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

Level of measurement and setting

[Response Begins]
Public Reporting
[Public Reporting Please Explain]

1. **Name of program and sponsor:** PPS -Exempt Cancer Hospital Quality Reporting (PCHQR) Program Measures, CMS (*Note-NQF# 0216 is implemented in this program for FY 2022, 2023, and 2024 as PCH-33)
    a. **URL:** https://qualitynet.cms.gov/pch/pchqr/measures
    b. **Purpose:** PCHQR is intended to encourage hospitals and clinicians to improve the quality of inpatient care that is provided to Medicare beneficiaries. A major part of the program supports improvement by ensuring that providers are aware of and reporting on best practices for their respective facilities and type of care.
    c. **Geographic area and number and percentage of accountable entities and patients included:** Eligible hospitals are referred to as PPS-Exempt Cancer Hospitals (PCHs). These hospitals are excluded from payment under the inpatient prospective payment system (IPPS). Eleven hospitals (hospital list link) have been granted the PCH designation by CMS.
    d. **Level of measurement and setting:** Facility Level; Claims Data Source; Inpatient/Hospital Care Setting

Payment Program
[Payment Program Please Explain]

1. **Name of program and sponsor:** Merit-based Incentive Payment System (MIPS) reporting program, Center for Medicare and Medicaid Services (CMS) (*Note-NQF# 0216 is implemented in this program as of PY 2017 as QPP# 457)
    a. **URL:** https://qpp.cms.gov/mips/explore-measures
    b. **Purpose:** MIPS takes a comprehensive approach to payment by basing consideration of quality on a set of evidence-based measures that were primarily developed by clinicians, thus encouraging improvement in clinical practice and supporting advances in technology that allow for easy exchange of information.
    c. **Geographic area and number and percentage of accountable entities and patients included:** MIPS, eligible providers may earn performance-based payment adjustments for the services provided to Medicare patients in the USA.

i. Eligible providers include: Physicians (including doctors of medicine, osteopathy, dental surgery, dental medicine, podiatric medicine, and optometry), Osteopathic practitioners, Chiropractors, Physician assistants, Nurse practitioners, Clinical nurse specialists, Certified registered nurse anesthetists, Physical therapists, Occupational therapists, Clinical psychologists, Qualified speech-language pathologists, Qualified audiologists, Registered dietitians or nutrition professionals.

d. **Level of measurement and setting:** Clinician/Group Level; Registry Data Source; Outpatient Services/Ambulatory Care Setting

2. **Name of program and sponsor:** Polaris, FIGmd (NQF# 0216 is implemented in this registry for 2022 reporting as QPP# 457)
    a. **URL:** https://polaris.figmd.com/measures1/
    b. **Purpose:** Polaris, a FIGmd hosted registry, offers a simplified and innovative approach for MIPS reporting. Polaris is a CMS-approved Qualified Clinical Data Registry (QCDR) in collaboration with the American Society of Clinical Oncology (ASCO) and a Qualified Registry (QR).
    c. **Geographic area and number and percentage of accountable entities and patients included:** FIGmd is a leading Registry operator in the United States and handles submission and reporting requirements for over 77,000 providers.
    d. **Level of measurement and setting:** Clinician/Group Level; Registry Data Source; Outpatient Services/Ambulatory Care Setting

Quality Improvement with Benchmarking (external benchmarking to multiple organizations)

[Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Please Explain]

1. **Name of program and sponsor:** Quality Oncology Practice Initiative (QOPI®), American Society of Clinical Oncology (*Note, Retired from QOPI in PY 2021)
    a. **URL:** https://practice.asco.org/quality-improvement/quality-programs/quality-oncology-practice-initiative
    b. **Purpose:** QOPI® is an oncologist-led, practice-based quality assessment and improvement program. QOPI provides a standard methodology, robust library of quality metrics for oncology, and a collection tool to reliably and routinely assess care, inform quality improvement activities, and demonstrate quality to patients and external stakeholders. Collection rounds are offered twice per year, in spring and fall, for an eight-week period.
    c. **Geographic area and number and percentage of accountable entities and patients included:** QOPI® is available to all oncology practices with a least one active ASCO member located in the US, US territories, and several countries outside of the US, including Argentina, Australia, Brazil, all countries in the European Union, India, Malaysia, Mexico, New Zealand, Pakistan, Philippines, and Saudi Arabia.
    d. **Level of measurement and setting:** Clinician/Group Level; Registry Data Source; Outpatient Services/Ambulatory Care Setting

Other (specify)

[Other (specify) Please Explain]

1. **Name of program and sponsor:** Core Quality Measures Collaborative (CQMC) 2020 Medical Oncology Core Set, AHIP, CMS, and NQF (*Note, this is not a public reporting or payment program, just recommended core measure set by specialty. NQF# 0216 is included in the Medical Oncology Core Set.)
    a. **URL:** https://www.qualityforum.org/CQMC_Core_Sets.aspx
    b. **Purpose:** The CQMC defines a core measure set as a parsimonious group of scientifically sound measures that efficiently promote a patient-centered assessment of quality and should be prioritized for adoption in value-based purchasing and APMs.
    c. **Geographic area and number and percentage of accountable entities and patients included:** N/A-this is not a public reporting or payment program
    d. **Level of measurement and setting:** N/A-this is not a public reporting or payment program

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]
Measure Currently in Use
[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]
As described above, this measure is included in the CMS PQRS program. CMS is planning to publicly report QCDR data. This measure

Additionally, although the measure is currently in use, we will continue to seek opportunities to advocate for expanded use of this measure in government or other programs.
[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]
The measure was recently considered for CMS MIPS program by the NQF Measures Application Partnership MAP and received a recommendation of support. With the MAP recommendation, it can now be considered for inclusion in the next proposed rule for this program with the earliest implementation of 2017.

This measure has also been included in Americas Health Insurance Plans Medical Oncology Core Measure Set. The purpose of this program is to reduce variability in measure selection, specifications and implementation. The measures will be implemented nationally by private health plans using a phased-in approach. Contracts between physicians and private payers are individually negotiated and therefore come up for renewal at different points in time depending on the duration of the contract. It is anticipated that private payers will implement these core sets of measures as and when contracts come up for renewal or if existing contracts allow modification of the performance measure set. CMS is also working to align measures across public programs. They intend to include, for broad input, the agreed upon draft measure sets in the Physician Fee Schedule and other proposed rules. For measures that are not currently in CMS programs, CMS would go through the annual pre-rulemaking and rulemaking processes to solicit stakeholder and public input. Depending on public response, these measures will be included in a timeframe determined by the Agency.
[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]
CMS publicly reports performance rates and benchmarks annually for the MIPS program to help eligible providers understand how they perform relative to others. CMS offers a range of support to help eligible providers actively participate in MIPS, including webinars and the QPP Service Center.
CMS has quality of care data publicly displayed on a rolling quarter basis for the PCHQR program. The PCHQR program also offers the QualityNet Service Center for assistance, including interpretation of the measure.

ASCO's measure development team are available to receive comments and questions from measure implementers and clinicians reporting ASCO measures. If comments or questions require expert input, these are shared with ASCO's Technical Expert Panel to determine if measure modifications may be warranted in the annual maintenance of the measure. Additionally, for ASCO measures included in CMS' federal reporting programs, there is a system that has been established to elicit timely feedback and responses from ASCO measure development team, as needed.

[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]
CMS publicly reports performance rates and benchmarks annually for the MIPS program to help eligible providers understand how they perform relative to others. CMS offers a range of support to help eligible providers actively participate in MIPS, including webinars and the QPP Service Center.
CMS has quality of care data publicly displayed on a rolling quarter basis for the PCHQR program. The PCHQR program also offers the QualityNet Service Center for assistance, including interpretation of the measure.

[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]
ASCO's measure development team allows for feedback and measure inquiries from implementers and reporters via email. In addition, CMS' MIPS and PCHQR programs both offer helpdesks for questions on measures. At this time, no specific feedback has been received by ASCO on this measure through those avenues.
[Response Ends]

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

[Response Begins]
No additional feedback has been received by ASCO on this measure. However, we will continue to solicit feedback as we perform maintenance on this measure.
[Response Ends]

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

[Response Begins]
In 2020, NQF# 0216 has been deemed a priority measure by the Core Quality Measures Collaborative (CQMC) (sponsored by NQF, CMS, and AHIP). CQMC included this measure in their Medical Oncology Core Set. The CQMC defines a core measure set as a parsimonious group of scientifically sound measures that efficiently promote a patient-centered assessment of quality and should be prioritized for adoption in value-based purchasing and APMs.
In 2015-2016, the NQF MAP reviewed this measure and gave a "Support" recommendation. This NQF-endorsed measure addresses an important gap area identified by MAP in end-of-life care for reporting by oncologists, is fully-specified and tested, reflects patient-centered care, and addresses the important areas of care coordination and appropriate use. The MAP felt patients and families value this type of information for public reporting.

[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]

Thus far, ASCO has not received specific feedback on the measure specification; therefore, ASCO's TEP did not consider external feedback from those being measured during revision of measure specifications or implementation.
[Response Ends]

4b. Usability

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

[Response Begins]
In evaluating the performance data (from QOPI), the average performance rate on this measure between 2013-2015 is 17.57% and the average performance rate between 2017-2020 is %19.24 (Note-this is an inverse measure). There continues to be some fluctuations in performance. However, there is an increase in the number of practices reporting on this measure since 2013. Available performance data also indicates continued performance at lower levels, with the lowest performance score between 61.54-100%.
In evaluating the performance data (from MIPS), the mean performance rate on this measure between 2017-2020 varies between 8.48 to 10.83 (Note-this is an inverse measure). There continues to be some fluctuations in performance. However, again MIPS remains a voluntary reporting program. Participants are allowed to self-select measures and may choose those that will result in high performance rates. As a result, performance rates may not be nationally representative.
The overall performance of this measure should not be 0, to account for individual cases where there could be patients early in their treatment course that unexpectedly died and also to account for patient and family treatment preferences. However, high overall rates should be examined for clinical appropriateness.
In addition, given recent studies in the literature (refer to question 1b.05) indicating disparities for the Medicaid population and African American population, it would be helpful in the future to have available data to follow the trends for these populations.

[Response Ends]

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

[Response Begins]
At this time, we are not aware of any unintended consequences related to this measure. We take unintended consequences very seriously and therefore continuously monitor to identify actions that can be taken to mitigate them.
[Response Ends]

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

[Response Begins]
We have not observed any unexpected benefits associated with implementation of this measure.
[Response Ends]

Criteria 5: Related and Competing Measures

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

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

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

(Can search and select measures.)
[Response Begins]
0213: Percentage of patients who died from cancer admitted to the Intensive Care Unit (ICU) in the last 30 days of life
3235: Hospice and Palliative Care Composite Process Measure—Comprehensive Assessment at Admission
0210: Percentage of patients who died from cancer receiving chemotherapy in the last 14 days of life
2651: CAHPS® Hospice Survey (experience with care)
[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]
PCH-34 Proportion of Patients Who Died from Cancer Not Admitted to Hospice
Steward: American Society of Clinicial Oncology

[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]
There are no competing measures. NQF 0210, NQF 0213, PCH-34, NQF 2651, and NQF 3235 are related measures and their differences with NQF 0216 are summarized below:

- NQF 0210, NQF 0213, and PCH-34 are also stewarded by ASCO and are harmonized to the extent possible in the measure specifications with NQF 0216. All 4 measures address the same target population, patients who died of cancer. However each measure addresses 4 different measure focuses, per the numerator, specifically

undesirable events prior to death (i.e. ICU admissions within 30 days before death, hospice enrollment less than 3 days before death, not enrolled in hospice prior to death, and chemotherapy within 14 days before death).

- NQF 2651 CAHPS® Hospice Survey (experience with care) is stewarded by CMS. NQF 2651 is a PRO-PM measure in the hospice facility setting using an instrument based data source. However, NQF 0216 is an outpatient setting measure using a registry data source. NQF 2651 addresses a target population of adult primary caregivers of hospice decedents to complete the survey. NQF 0216 addresses a target population of patients who died of cancer. Although NQF 2651 and NQF 0216 both address hospice, the measure focuses differ significantly. NQF 2651 is focusing on a survey regarding the care experiences of hospice patients and their primary caregivers. Whereas NQF 0216 is focusing on if there is still extremely late enrollment to hospice for cancer patients, in this case hospice enrollment less than 3 days before their death.
- NQF 3235 Hospice and Palliative Care Composite Process Measure—Comprehensive Assessment at Admission is stewarded by CMS. NQF 3235 is a composite measure in the hospice facility setting using a standardized, patient-level data collection instrument data source. However, NQF 0216 is an outpatient setting measure using a registry data source. NQF 3235 addresses a target population of all hospice patient stays enrolled in hospice. NQF 0216 addresses a target population of patients who died of cancer. Although NQF 3235 and NQF 0216 both address hospice, the measure focuses differ significantly. NQF 3235 is focusing on the percentage of hospice stays in which patients received a comprehensive patient assessment at hospice admission. Whereas NQF 0216 is focusing on if there is still extremely late enrollment to hospice for cancer patients, in this case hospice enrollment less than 3 days before their death.

[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]
There are no competing measures.
[Response Ends]

Appendix

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

Contact Information

Measure Steward (Intellectual Property Owner): American Society of Clinical Oncology
Measure Steward Point of Contact: Drumheller, Caitlin, caitlin.drumheller@asco.org
Durakovic, Lela, lela.durakovic@asco.org

Measure Developer if different from Measure Steward: American Society of Clinical Oncology
Measure Developer Point(s) of Contact: Drumheller, Caitlin, caitlin.drumheller@asco.org
Durakovic, Lela, lela.durakovic@asco.org

Additional Information

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

[Response Begins]
No appendix
[Response Ends]

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

Describe the members' role in measure development.

[Response Begins]
ASCO Palliative Measure Development Panel
The panel is responsible for reviewing evidence and maintaining measures

Tracey Evans, MD (Chair)
University of Pennsylvania

Craig Earle, MD, FASCO (Co-Chair)
Institute for Clinical Evaluative Science

Katherine Ast, MSW, LCSW
American Academy of Hospice and Palliative Medicine

Amy Berman
The John A. Hartford Foundation

Kathleen Bickel, MD, MPhil
White River Junction VA Medical Center

Eduardo Bruera, MD
The University of Texas MD Anderson Cancer Center

Sydney Dy, MD
Johns Hopkins

Esme Finlay, MD
University of New Mexico Cancer Research and Treatment Center

Arif Kamal, MD, MHS, FAAHPM
Duke University

Kristen McNiff, MPH
Dana-Farber Cancer Institute

Michael Neuss, MD, FASCO
Vanderbilt Ingram Cancer Center

John Sprandio, MD
Consultant in Med Onc and Hem Inc

Holley Stallings, RN
Norton Cancer Institute

Jamie Von Roenn, MD, FASCO
American Society of Clinical Oncology
[Response Ends]

3. Indicate the year the measure was first released.

[Response Begins]
2007
[Response Ends]

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

[Response Begins]
January 2022
[Response Ends]

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

[Response Begins]
Annually
[Response Ends]

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

[Response Begins]
Spring 2022
[Response Ends]

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

[Response Begins]
THE MEASURES AND SPECIFICATIONS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND.

[Response Ends]

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

[Response Begins]
The Measure is not clinical guideline, does not establish a standard of medical care, and has not been tested for all potential applications.
The Measure, while copyrighted, can be reproduced and distributed, without modification, for noncommercial purposes, e.g., use by health care providers in connection with their practices. Commercial use is defined as the sale, license, or

distribution of the Measure for commercial gain, or incorporation of the Measure into a product or service that is sold, licensed or distributed for commercial gain.
Commercial uses of the Measure require a license agreement between the user and American Society of Clinical Oncology (ASCO) and prior written approval of ASCO. Contact measurement@asco.org for licensing this measure. Neither ASCO, nor its members shall be responsible for any use of the Measure.
ASCO encourages use of the Measure by other health care professionals, where appropriate.

[Response Ends]

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

[Response Begins]
N/A
[Response Ends]