NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

Purple text represents the responses from measure developers.

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

Brief Measure Information

NQF #: 0326
Corresponding Measures:
De.2. Measure Title: Advance Care Plan
Co.1.1. Measure Steward: National Committee for Quality Assurance
De.3. Brief Description of Measure: Percentage of patients aged 65 years and older who have an advance care plan or surrogate decision maker documented in the medical record or documentation in the medical record that an advance care plan was discussed but the patient did not wish or was not able to name a surrogate decision maker or provide an advance care plan.
1b.1. Developer Rationale: This measure addresses advance care planning as one facet of high quality care for older adults. The aim of advance care planning is to ensure that care near the end of life aligns with the patient's wishes (IOM, 2014). Advanced care planning is associated with improved health outcomes for older adults, including reducing hospitalizations, intensive care unit (ICU) admissions, and hospital and ICU lengths of stay (Brinkman-Stoppelenburg, 2014; Hall et al., 2011; Khandelwal et al., 2015; Martin et al., 2016). However, most older adults do not have advance care planning conversations with their clinicians even though there is consensus among diverse stakeholders that advance care planning is a key component of high quality care (NQF 2006; IOM, 2014). The intent of this measure is to promote advance care planning discussions between older adults and their providers and documentation of that discussion in the patient's record.

As people age, consideration should be given to their treatment wishes in the event that they lose the ability to manage their care. A large discrepancy exists between the wishes of dying patients and their actual end-of-life care. Advance directives (AD) are widely recommended as a strategy to improve compliance with patient wishes at the end of life, and thereby ensure appropriate use of health care resources at the end of life. A recent systematic review found only a few studies, all of which were conducted in the United States concerning advanced care planning in palliative care. Although the results were promising, more high-quality studies need to be conducted (Hall, et al., 2011).

Brinkman-Stoppelenburg, A., Rietjens, J. A., & van der Heide, A. (2014). The effects of advance care planning on end-of-life care: a systematic review. Palliative Medicine, 28(8), 1000-1025.

Hall, S., Kolliakou, A., Petkova, H., Froggatt, K., & Higginson, I. J. (2011). Interventions for improving palliative care for older people living in nursing homes. Cohrane Database of Systematic Reviews, 3.

Institute of Medicine (IOM). (2014). Dying in America: improving quality and honoring individual preferences near the end of life. Washington, DC: The National Academies Press.

Khandelwal, N., Kross, E. K., Engelberg, R. A., Coe, N. B., Long, A. C., & Curtis, J. R. (2015). Estimating the effect of palliative care interventions and advance care planning on ICU utilization: a systematic review. Critical Care Medicine, 43(5), 1102-1111.

Martin, R. S., Hayes, B., Gregorevic, K., & Lim, W. K. (2016). The effects of advance care planning interventions on nursing home residents: a systematic review. Journal of the American Medical Directors Association, 17(4), 284-293.

National Quality Forum. (2006). A National Framework and Preferred Practices for Palliative and Hospice Care Quality. Washington, DC: National Quality Forum.

S.4. Numerator Statement: Patients who have an advance care plan or surrogate decision maker documented in the medical record or documentation in the medical record that an advance care plan was discussed but patient did not wish or was not able to name a surrogate decision maker or provide an advance care plan.
S.6. Denominator Statement: All patients aged 65 years and older.
S.8. Denominator Exclusions: N/A
De.1. Measure Type: Process
S.17. Data Source: Claims, Electronic Health Data
S.20. Level of Analysis: Clinician: Group/Practice, Clinician: Individual

Preliminary Analysis: Maintenance of Endorsement

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

Criteria 1: Importance to Measure and Report

- 1a. Evidence

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

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

The developer provides the following evidence for this measure:

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

Summary of prior review in 2016

- For the 2016 review, the developer referenced a 2014 systematic review that evaluates the effect of ACP on hospitalization and length of stays. Evidence from the 21 studies showed that use of an ACP is linked to a decreased rate of hospitalizations.
    - The Committee agreed that the updated evidence is directionally the same since the last NQF endorsement evaluation, and therefore the Committee accepted the prior evaluation of this criterion without further discussion or vote.

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:**

Updates:

- In the current submission, the developer provided two additional studies to support the systematic review provided in the previous review. The developer states that the new studies provide additional support for the measure.

Questions for the Committee:

- The evidence provided by the developer is updated and directionally the same compared to that for the previous NQF review. Does the Committee agree there is no need for repeat discussion and vote on Evidence?

Guidance from the Evidence Algorithm

Intermediate outcome measure with systematic review (Box 3) → Summary of the QQC provided (Box 4) → Systematic review concludes moderate quality evidence (Box 5b).

The highest possible rating is "High" for Evidence

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

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

Maintenance measures – increased emphasis on gap and variation

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

Gap Data

- For the current submission, the developer provided the following rates:
  - 2017: Mean = 74%, Min = 0%, 10th - 90th Percentile Range = 13% - 100%, Max = 100%, IQR = 42%

Disparities

- In the previous submission, some Committee members expressed concern that there is missing disparities information and the Committee strongly encouraged the developer to collect and provide the disparities information in the future.
- In the current submission, the developer did not provide disparities data and indicated that QPP MIPS data does not include disparities results.
- The developer did summarize literature addressing disparities and advance care plans. One study found that beneficiaries who are African American are less likely to have formal documentation of their end of life wishes while another study found that African American beneficiaries along with those who are Latino, less educated, or had lower income were less likely to have an advance care plan.
- Another study found that while racial and ethnic minorities were aware of advance care plans, they were less likely to have completed one.

Questions for the Committee:

- Does the performance data provided continue to warrant a national performance measure?
- If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?

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

- **Committee Pre-evaluation Comments:**

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

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

- This is a straightforward process measure; I have no concerns.
- 0326 Advance Care Plan is a process measure. Evidence from the 21 studies showed that use of an ACP is linked to a decreased rate of hospitalizations.
- pass
- No concerns
- There is evidence to support a focus on practitioners having an advanced care planning discussion/conversation with patients aged 65 years and older
- The evidence does relate to the outcome being measured and the more recent evidence provided by the developer continues to provide support. I feel comfortable with the current evidence and so not feel a repeat discussion or vote is necessary.
- The evidence indicates that ACP decreases hospitalization.
- Pass, sufficient and appropriate confirmatory evidence for maintenance measure.
- Process measure. Notes documentation of ACP, surrogate decision maker, or that a discussion was had but patient could create ACP at that time. Evidence of the benefits of ACPs is promising, but more data needed.
- New studies provided for 2020 review.
- The developer provided updated evidence for this measure and is directionally the same as the previous review.
- This measure is still not strong in demonstrating outcomes but I can support continuation
- I am aware that the evidence for advance care planning is mixed and that some on the field question it as a result. I think the committee should discuss these concerns but am reluctant to not approve this measure as it is the only one used widely across Medicare programs for this topic. It could be improved by including diversity information but I would not hold up its use for that information.

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

- The updated performance gap data is helpful
- developer did not provide disparities data and indicated that QPP MIPS data does not include disparities results. The developer did summarize literature addressing disparities and advance care plans. One study found that beneficiaries who are African American are less likely to have formal documentation of their end of life wishes while another study found that African American beneficiaries along with those who are Latino, less educated, or had lower income were less likely to have an advance care plan. Another study found that while racial and ethnic minorities were aware of advance care plans, they were less likely to have completed one.
- gap is widely known along populations but measure should be considered for when asked or not asked
- No concerns.
- Agree that the measure demonstrates an opportunity for improvement. Also agree that there needs to be included disparities information
- The gap in regard to disparities and advance care plans is solidly supported my numerous studies and I think the performance data does continue to warrant a national performance measure. I do feel strongly that the developer should collect and provide disparities data, although I understand that QPP MIPS does not include it. Moving forward this is an important consideration.

- There is evidence of a performance gap. There are definite racial disparities in the completion of ACP.
- Moderate, variability is demonstrated in the data provided but specific performance of this measure within subgroups is not described.
- Based on 2017 (mean 74%) there seems to be room for improvement. Still no disparities data presented by developer.
- No disparity info included. Seems this would be good to include, especially for racial disparities with ACP.
- The developer was strongly encouraged to collect and provide disparities information, which they did not do in this submission. They did summarize literature addressing advanced care plans that indicated that African Americans along with those who are Latino, less educated, or had lower income were less likely to have an advance care plan.
- Criteria met to continue.
- Similar comment as above: when done comprehensively, advance care planning can reduce stress of participants and allow the medical team to better meet people's goals and preferences. Whether this measure is the best one for this process is perhaps debatable, but I would not want to reject it as we need some way of measuring and promoting this process to providers and patients.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

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

- Reliability

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

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

- Validity

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

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

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

Evaluators: NQF Staff
NQF Staff Review

Reliability

- In the previous submission, the developer did not provide updated reliability testing for this maintenance review.
    - Committee members noted that the previous testing is from a small sample of records from only four sites of care. However, the results indicated strong reliability with an overall kappa score of 0.97.

- The Committee accepted the prior evaluation of the reliability subcriterion without further discussion and accepted a motion to carry over votes from the previous evaluation on reliability.

- In the current submission, the developer conducted measure score level reliability testing.
  - Using 2017 MIPS data from 1,031 group/practices, the developer used a beta-binominal model to assess the signal-to-noise ratio. Using this method, the overall mean reliability score was 0.999.
  - The developer concluded the scores indicated good reliability.

Validity

- During the 2016 review, several Committee members noted that a significant reconsideration of validity was not warranted unless there is evidence that the use of CPT codes for ACP have changed substantially since testing was first conducted. The Committee accepted a motion to carry over votes from the previous evaluation on validity.
- For the 2020 submission, validity testing was performed at the measure score level through construct validity testing and face validity.
- The developer conducted Pearson correlation for construct validity against NCQA's Documentation of Current Medications in the Medical Record measure.
- Results:
  - Positive Correlation: Advance Care Plan
    - Correlation coefficient = 0.63, $p < 0.001$
  - The developer concluded that there is a moderate correlation between the Documentation of Current Medications in the Medical Record measure and the Advance Care Plan measure.
- The developer also noted that face validity was also conducted.
  - The developer convened a panel of 33 members to assess the face validity of this measure.
  - The developer referred to the 2016 face validity results for the 2020 submission in which the panel and public found the measure to be valid.

Questions for the Committee regarding reliability:

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

Questions for the Committee regarding validity:

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

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

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

- **Committee Pre-evaluation Comments:**

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

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

- updated data is very helpful

- For the 2020 submission, validity testing was performed at the measure score level through construct validity testing and face validity. Face validity conducted by 33 member panel.
- no concerns
- No concerns
- Reliability demonstrated
- I have no concerns regarding the measure being consistently implemented
- Reliability is high.
- Elements are clearly defined but I have concerns about the consistency with which the numerator detail 1124F "patient did not wish or was not able to name a SDM or provide an ACP" which allows for subjective assessment of "viewed as harmful" related to a patient's cultural and/or spiritual beliefs and permits preclusion of an ACP discussions would viewed as harmful.
- Using 2017 MIPS data from 1,031 group/practices, the developer used a beta-binominal model to assess the signal-to-noise ratio. Using this method, the overall mean reliability score was 0.999.
- No concerns
- It appears that this measure can be consistently implemented.
- It still does not relate to outcomes
- As an existing measure, I feel this has been addressed but am open to discussing it in the committee

2a2. Reliability - Testing: Do you have any concerns about the reliability of the measure?
- no
- No concerns noted.
- no concerns
- No
- Reliability demonstrated
- I do not have any concerns regarding reliability, particularly since the developer did conduct additional reliability testing and the score was solid. I feel there is no need to discuss or vote on reliability.
- No
- High reliability testing.
- No
- No
- no concerns
- no concerns
- As an existing measure, I feel this has been addressed but am open to discussing it in the committee

2b1. Validity -Testing: Do you have any concerns with the testing results?
- no
- No concerns noted.
- no concerns
- No
- Validity demonstrated
- I have not concerns regarding the validity testing, I am satisfied by the data provided by the developer.
- No
- No concerns. High validity rating.
- moderate correlation between the Documentation of Current Medications in the Medical Record measure and the Advance Care Plan measure. The developer convened a panel of 33 members to assess the face validity of this measure. - No concerns
- No
- no
- no
- As an existing measure, I feel this has been addressed but am open to discussing it in the committee

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

- no concerns
- no concern - based on 2016
- no concerns
- No concerns
- No risk adjustment was included
- No issue
- Risk adjustment was not used.
- No concerns.
- Excludes People < 65 YO
- No risk adjustment used
- Exclusions seem to be consistent, the analyses indicate acceptable results.
- adequate
- None I am aware of

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

- no concerns
- Committee is using 2016 face reliability.
- no concerns
- No concerns
- Need to include diversity information
- I see no threats to validity
- No. The data is publicly available through CMS.
- Opportunity to compare performance for those who met criteria by numerator 1123F and those by 1124F across organizations.
- Only used for > 65 YO. No missing data analysis because data publicly available (seems like the measure assesses missing information in the medical record.
- No
- There does not appear to be any missing data, statistically significant differences, or multiple data sources.
- outcome data would be helpful but difficult to obtain
- As an existing measure, I feel this has been addressed but am open to discussing it in the committee

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.

- Data elements for this measure are coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims.
- All data elements are in defined fields in a combination of electronic sources
- This measure uses the Clinical Quality Measure reporting method. Some components of this measure draw on structured fields, while others are available in narrative notes or other non-structured fields.

Questions for the Committee:

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

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

- **Committee Pre-evaluation Comments:**

Criteria 3: Feasibility

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

- yes, very much should be routinely collected and thus feasible
- The developer noted that only data from 2017 is available for this measure thus performance trends cannot be reported.
- no concern
- No concerns
- Definitely at least moderate feasibility. Only concern would be when an advanced care planning conversation occurs within a "routine" patient encounter where an advanced care planning CPT/ICD code would not be generated
- I have no concerns.
- Some data elements are available in structured fields while others are in narrative notes or non-structured fields.
- Moderate feasibility, as ease of capture of data elements (especially patients for numerator 1124F) that require chart abstraction rather than discrete fields for reporting.
- Data elements for this measure are coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims. • All data elements are in defined fields in a combination of electronic sources 5 • This measure uses the Clinical Quality Measure reporting method. Some components of this measure draw on structured fields, while others are available in narrative notes or other non- structured fields.
- No concerns
- For this measure, all data elements are in defined fields in electronic clinical data. Some of components are still available in narrative form.
- no concerns
- As an existing measure I would expect that it is feasible but am open to hearing information if that's not the case

Criterion 4: Usability and Use

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

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

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

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

Current uses of the measure

- This measure is in use in CMS Merit-based Incentive System (MIPS).

Publicly reported? ☒ **Yes** ☐ **No**

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

OR

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

Accountability program details

- The measure is currently used in the CMS Merit-Based Incentive Payment System (MIPS). Data are reported publicly via Physician Compare. The measure is also used in the CMS Quality Payment Program (QPP).

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

- CMS provides clinicians and groups who report their performance on this measure with feedback reports to inform performance improvement efforts. Additionally, all performance data is reported publicly through Physician Compare, and annual benchmarks are publicly available to enable clinicians to understand how their performance compares to national benchmarks.
- The developer notes that no feedback has been received from those being measured via CMS or NCQA portals.

***Questions for the Committee**:*

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

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

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

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

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

Improvement results

- The developer noted that only data from 2017 is available for this measure thus performance trends cannot be reported.

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 stated that no unexpected findings were identified for this measure.

Potential harms

- The developer stated that no unintended consequences were identified for this measure.

Questions for the Committee:

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

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

Rationale: Data were not provided to demonstrate improvement or trends in performance.

- **Committee Pre-evaluation Comments:**

Criteria 4: Usability and Use

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

- no concerns
- The developer noted that only data from 2017 is available for this measure thus performance trends cannot be reported.
- no concern
- No concerns
- Yes publicly reported & used in an accountability program.
- From reading the information provided, it appears that clinicians have access to the data results and also receive feedback on how to improve.
- CMS publicly reports this data. No feedback has been received from those being measured.
- Pass, public reporting exists. It is not clear if feedback is specifically requested, or just not reported.
- The measure is currently used in the CMS Merit-Based Incentive Payment System (MIPS). Data are reported publicly via Physician Compare. The measure is also used in the CMS Quality Payment Program (QPP). The developer notes that no feedback has been received from those being measured via CMS or NCQA portals.
- Measure is public reported on Physician Compare site
- This measure is publicly reported and is used in an accountability program CMS MIPS and CMS QPP.
- unknown
- As an existing measure, I feel this has been addressed but am open to discussing it in the committee

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

- see that updated data is not provided, but believe more data is not needed here

- No harms noted.
- need data such as how CPT codes and reimbursement have affected measure
- How soon would additional data be available?
- I can speak from my clinical practice personal experience that results of having advanced care plan conversations can be used to further the goal of high quality, efficient healthcare for older adults. The benefits outweigh harms
- I think sharing the results with clinicians and providing them with feedback could further the goal of high-quality, efficient healthcare. At this time, however, data is only available for 2017 and so it is difficult to determine if that is happening. I do think the benefits outweigh any potential unintended consequences.
- The benefits of improving care at the end of life outweighs any unintended negative consequence.
- Insufficient information is provided to rate.
- No unexpected findings, no unintended consequences. Only have data from 2017--noted as "insufficient"
- None
- Only data from 2017 is available and performance trends cannot be reported. No harms were identified.
- Weighing practical value
- As an existing measure, I feel this has been addressed but am open to discussing it in the committee

Criterion 5: Related and Competing Measures

Related or competing measures

- There are no related or completing measures for this measure.

Harmonization

- No harmonization is needed for this measure.

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

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

- no competing measures
- None.
- no other competing measures
- No concerns
- No competing measures identified
- I am not aware of any related and competing measures.
- There are none.
- n/a
- No related or competing measures.
- No
- No related or competing measures
- none of which I am aware
- I am not aware of related or competing measures and believe this is the only ACP measure in use among Medicare programs.

Public and Member Comments

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

- Comment by: National Committee for Quality Assurance

  This comment addresses the Fall 2020 cycle measure #0326 Advance Care Plan.
  NCQA would like to add the following data to the 4b. Usability (4a1. Improvement; 4a2. Benefits of measure) section of the submission:
  PQRS (Data Source: Centers for Medicare & Medicaid Services (CMS), 2016 PQRS Experience Report Appendix Tables)
  EPs who Reported Continuously 2013-2016: 3,220
  Average Performance Rate in 2013: 69.6%
  Average Performance Rate in 2014: 72.9%
  Average Performance Rate in 2015: 75.3%
  Average Performance Rate in 2016: 76.6%
  Improvement Rate: 3.3%

- No Public or NQF Member comments submitted as of this date.

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 0326

Measure Title: Advance Care Plan

Type of measure:

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

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

Data Source:

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

Level of Analysis:

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

Measure is:

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

RELIABILITY: SPECIFICATIONS

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

   Submission document: "MIF_0326" document, items S.1-S.22

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

2. **Briefly summarize any concerns about the measure specifications.**
   - No concerns about the measure specifications

RELIABILITY: TESTING

Submission document: "MIF_0326" document for specifications, testing attachment questions 1.1-1.4 and section 2a2

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

   Submission document: Testing attachment, section 2a2.2
   - The developer calculated signal-to-noise using the Beta-binomial model on 2017 MIPS data from 1,031 groups/practices.
   - The developer also calculated the standard error and 95% confidence interval of the mean signal-to-noise reliability as well as the distribution of reliability estimates.
7. **Assess the results of reliability testing**

   Submission document: Testing attachment, section 2a2.3
   - Out of the sample of 1,031 groups/practices, the mean signal-to-noise reliability was 0.999
   - The developer also reported the following signal-to-noise reliability for each tercile of their sample:

Stratification	Number of Group/Practices	Number of Eligible Patients per Group/Practice (min – max)	Mean Signal-To-Noise Reliability	SE	95% CI
All groups/ practices	**1,031**	**20 – 74,453**	**0.999**	**0.000**	**(0.998, 0.999)**
Tercile 1	340	20 – 580	0.995	0.001	(0.994, 0.996)
Tercile 2	340	583 – 2,013	0.999	0.000	(0.999, 0.999)
Tercile 3	351	2,016 – 74,453	1.000	0.000	(1.000, 1.000)

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

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

   ☐ **Not applicable** (score-level testing was not performed)
9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?

   Submission document: Testing attachment, section 2a2.2

   ☐ **Yes**

   ☐ **No**

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

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

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

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

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

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

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**
    - The mean signal-to-noise reliability score of 0.999 indicates strong reliability of this measure.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.
    - This measure has no exclusions.
    - No concerns.

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

    Submission document: Testing attachment, section 2b4.
    - The developer calculated an inter-quartile range of the performance scores to examine the differences between the 25th percentile and the 75th percentile.
    - The developer stated that the 42% IQR represents 1,177 additional older adult patients having an advance care plan in high-performing practices versus low-performing practices.
    - The developer concluded that there are meaningful differences in performance among group practices.
    - No concerns.

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
    Submission document: Testing attachment, section 2b5.
    - This measure has only one data source.
    - Not applicable.

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

    Submission document: Testing attachment, section 2b6.
    - The developer states that it is unable to conduct missing data analysis because the data is publicly available through CMS.
    - No concerns.

16. **Risk Adjustment**

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

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

    ☐ Yes ☐ No ☒ Not applicable

    16c. **Social risk adjustment:**

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

16c.2 Conceptual rationale for social risk factors included? ☐ Yes ☐ No

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

16d. **Risk adjustment summary:**

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

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

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

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

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

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

- N/A

For cost/resource use measures ONLY:

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

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

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

VALIDITY: TESTING

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

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

☒ Face validity

☒ Empirical validity testing of the measure score

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

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

Submission document: Testing attachment, section 2b2.2

- The developer used a Pearson correlation test to assess the correlation between this measure and NQF 0419 Documentation of Current Medications in Medical Record.
- The developer referred to their 2016 methods for face validity testing.
- In the 2016 submission, the developer abstracted over 200 patient records to calculate inter-rater reliability of the concept through an expert panel used to assess face validity of the measure concept. Workgroups convened by PCPI and NCQA established the measure's ability to capture as designated using a process consisting of multiple stakeholders input and a review of the input received during a public comment period.
- For Face Validity, the developer convened a panel of 33 members to assess the face validity of this measure.

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

Submission document: Testing attachment, section 2b2.3

- The developer provided the following Pearson Correlation Coefficients for groups/practices for the *Advance Care Plan* Measure using 2017 Data

Measure	Documentation of Current Medications in the Medical Record
Advance Care Plan	0.63
(N=, p value =)	(436, p < 0.001)

- The developer referred to their 2016 results for face validity testing
    - The developer stated that both the panel and public found the measure to be valid.

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

    Submission document: Testing attachment, section 2b1.

    ☒ **Yes**

    ☐ **No**

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

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

    NOTE that data element validation from the literature is acceptable.

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

    ☐ **Yes**

    ☐ **No**

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

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

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

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

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

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

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

    - Potential threats assessed (box 1) → Empirical validity conducted (box 2) → Computed performance measure scores (box 5) → appropriate methods (box 6) → moderate certainty (box 7b) → Moderate rating

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

    ☐ **High**

    ☐ **Moderate**

    ☐ **Low**

☐ **Insufficient**

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

ADDITIONAL RECOMMENDATIONS

29. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**
    - No concerns

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

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

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

nqf_evidence_attachment_7.1.docx

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

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

Yes

- 1a. Evidence (subcriterion 1a)

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

Outcome

☐ Outcome:

☐ Patient-reported outcome (PRO):

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

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

☒ Process: The measure assesses the percentage of patients, aged 65 years and older, who have an advance care plan or surrogate decision maker documented in the medical record or documentation in the medical record that an advance care plan was discussed.

☐ Appropriate use measure:

☐ Structure:

☐ Composite:

1a.2 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

2020 Submission

The measure assesses whether the patient has an advance care plan or surrogate decision maker documented in the medical record, or documentation that an advance care plan was discussed.

Logic Model:

Clinician initiates advance care planning discussion with patient >>> Clinician documents patient's decision-making around end of life care in the medical record >>> Clinician provides end of life care as specified by the patient's advance care plan or surrogate decision maker >>> Patient receives end of life care as specified by his

or her advance care plan or surrogate decision maker >>> Patient experiences improved quality of life at end-of-life >>> Patient experiences improved outcomes at end-of-life.

2016 Submission

The measure assesses whether the patient has an advance care plan or surrogate decision maker documented in the medical record, or documentation that an advance care plan was discussed. Given the risks of poorly coordinated care at the end of life (e.g., potential to disrespect patient and/or surrogate decision-maker wishes, increased hospitalizations, increased intensive care unit admissions, increased health care spending), prior documentation of patient and family's decision-making is particularly important for ensuring optimal quality of life improved outcomes. The path envisioned is as follows.

Logic Model:

Clinician initiates advance care planning discussion with patient >>> Clinician documents patient's decision-making around end of life care >>> Clinician provides end of life care as specified by the patient's advance care plan >>> Patient receives end of life care as specified by his or her advance care plan >>> Patient experiences improved quality of life >>> Patient experiences improved outcomes.

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

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

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

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

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

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Evidence

Source of Systematic Review:

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

2020 Submission

The Effects of Advance Care Planning on End-of-Life Care: A Systematic Review.

Brinkman-Stoppelenburg A, Rietjens JA, van der Heide A.

September 2014

Palliative Medicine, Vol. 28 (No. 8), pages 1000-1025.

http://pmj.sagepub.com/content/28/8/1000.short

2016 Submission

The Effects of Advance Care Planning on End-of-Life Care: A Systematic Review.

Brinkman-Stoppelenburg A, Rietjens JA, van der Heide A.

September 2014

Palliative Medicine, Vol. 28 (No. 8), pages 1000-1025.

http://pmj.sagepub.com/content/28/8/1000.short

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

2020 Submission

"The effects of different types of advance care planning have been studied in various settings and populations using different outcome measures. There is evidence that advance care planning positively impacts the quality of end-of-life care" (page 1000).

2016 Submission

"The effects of different types of advance care planning have been studied in various settings and populations using different outcome measures. There is evidence that advance care planning positively impacts the quality of end-of-life care" (page 1000).

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

2020 Submission

The level of evidence for each study was graded on a scale of I, II, or III using criteria proposed by Higginson 2002. Grade I is defined as a randomized controlled trial or RCT review, grade II is defined as a prospective study with a comparison group, or a retrospective study which controls effectively for confounding variables, and grade III is defined as a retrospective, observational, or cross-sectional study.

2016 Submission

This systematic review was designed to review and evaluate evidence, but not to provide a recommendation.

Provide all other grades and definitions from the evidence grading system

2020 Submission

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

2020 Submission

This systematic review was designed to review and evaluate evidence, but not to provide a recommendation.

2016 Submission

This systematic review was designed to review and evaluate evidence, but not to provide a recommendation.

Provide all other grades and definitions from the recommendation grading system

2020 Submission

Body of evidence:

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

2020 Submission

- Quantity – 113 studies were included in the systematic review
- Quality – Most studies included in the review were observational (95%), while 5 were experimental (5%). The level of evidence for each study was graded on a scale of I, II, or III using criteria proposed by Higginson 2002. 5% of studies included in drawing the conclusions received Grade I (defined as a randomized controlled trial or RCT review). 59% of the studies included received Grade II (defined as a prospective study with a comparison group, or a retrospective study which controls effectively for confounding variables). 36% of included studies received Grade III (defined as a retrospective, observational, or cross-sectional study).

2016 Submission

- Quantity – 113 studies were included in the systematic review
- Quality – Most studies included in the review were observational (95%), while 5 were experimental (5%). The level of evidence for each study was graded on a scale of I, II, or III using criteria proposed by Higginson 2002. 5% of studies included in drawing the conclusions received Grade I (defined as a randomized controlled trial or RCT review). 59% of the studies included received Grade II (defined as a prospective study with a comparison group, or a retrospective study which controls effectively for confounding variables). 36% of included studies received Grade III (defined as a retrospective, observational, or cross-sectional study).

Estimates of benefit and consistency across studies

2020 Submission

Out of 26 studies that evaluated the effect of advance care planning (ACP) on hospitalization or length of stay, 21 studies found that ACP was associated with a decreased rate of hospitalization or length of stay, while 5 studies found that ACP was associated with an increased rate of hospitalization or length of stay. Of 13 studies that evaluated the effect of ACP on patients' and families' symptoms, 5 studies found that ACP decreased symptoms, and no studies found that ACP increased symptoms. (the remaining studies found neither an increase nor a decrease).

2016 Submission

Out of 26 studies that evaluated the effect of advance care planning (ACP) on hospitalization or length of stay, 21 studies found that ACP was associated with a decreased rate of hospitalization or length of stay, while 5 studies found that ACP was associated with an increased rate of hospitalization or length of stay. Of 13 studies that evaluated the effect of ACP on patients' and families' symptoms, 5 studies found that ACP decreased symptoms, and no studies found that ACP increased symptoms. (the remaining studies found neither an increase nor a decrease).

What harms were identified?

2020 Submission

No harms were identified.

2016 Submission

No harms were identified.

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

2020 Submission

Lum, H. D., Sudore, R. L., & Bekelman, D. B. (2015). Advance care planning in the elderly. *Medical Clinics*, *99*(2), 391-403.

https://escholarship.org/uc/item/8m40n1p0

Weathers, E., O'Caoimh, R., Cornally, N., Fitzgerald, C., Kearns, T., Coffey, A., Daly, E., O'Sullivan, R., McGlade, C., & Molloy, D. W. (2016). Advance care planning: A systematic review of randomised controlled trials conducted with older adults. *Maturitas*, *91*, 101–109. https://doi.org/10.1016/j.maturitas.2016.06.016

The new evidence provides additional support that advance care plans are a key component of high-quality patient care.

2016 Submission

No new studies have been conducted that contradict the conclusion that advance care plans are a key component of high quality patient care.

1a.4 OTHER SOURCE OF EVIDENCE

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

2020 Submission

2016 Submission

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

2020 Submission

2016 Submission

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

2020 Submission

2016 Submission

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

2020 Submission

2016 Submission

- 1b. Performance Gap

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

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

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

This measure addresses advance care planning (ACP) as one facet of high quality care for older adults. ACP is intended to engage patients in proactive conversations and documentation about their care preferences

should there be an event in which they cannot independently express their wishes and decisions (Advance Care Plan Decisions, 2019).

It is widely agreed that ACP is a critical part of patient care, as it can lead to improved end of life care, increased trust in providers, decreased psychological distress, improved quality of life and can facilitate hope (Rosenberg et al., 2020). Patients with advance care plans have been found to experience improved quality of care at end of life (Bischoff et al., 2013). A systematic review found that among nursing home respondents, ACP can reduce hospitalizations between 9% and 26%, decrease costs and increase compliance with patients wishes (Martin et al., 2016).

However, many individuals are not having conversations regarding their care preferences in the event they are unable to make decisions. When asked, 70% of providers indicated that they only have ACP conversations with their patients experiencing advanced illness (Bires et al., 2017). Additionally, the benefits of ACP may only be realized if advance care plan documentation is created and the care team has access to, and follows, the patient's advance care plan. Evidence indicates that only between 35% – 38% of individuals have some form of an advance care plan (Yadav et al, 2017; Lendon et al., 2018). Additionally, one study found that while 70% of patients were familiar with advance directives, only 35% had completed one.

The intent of this measure is to promote advance care planning discussions between older adults and their providers and documentation of that discussion in the patient's record.

Advance Care Plan Decisions. (2019) Why Advance Care Planning is a Crucial Part of Population Health Strategy. Retrieved July 23, 2020, from https://acpdecisions.org/why-advance-care-planning-is-a-crucial-part-of-population-health-strategy/

Bires, J. L., Franklin, E. F., Nichols, H. M., & Cagle, J. G. (2018). Advance Care Planning Communication: Oncology Patients and Providers Voice their Perspectives. Journal of Cancer Education, 33(5), 1140–1147. https://doi.org/10.1007/s13187-017-1225-4

Bischoff, K. E., Sudore, R., Miao, Y., Boscardin, W. J., & Smith, A. K. (2013). Advance Care Planning and the Quality of End-of-Life Care in Older Adults. Journal of the American Geriatrics Society, 61(2), 209–214. https://doi.org/10.1111/jgs.12105

Lendon, J.P, Caffrey, C. & Lau, D. (2018). Advance directive documentation among adult day services centers and use among participants, by region and center characteristics?: National Study of Long-Term Care Providers, 2016. National Health Statistics Reports, 117. https://stacks.cdc.gov/view/cdc/58975

Martin, R. S., Hayes, B., Gregorevic, K., & Lim, W. K. (2016). The Effects of Advance Care Planning Interventions on Nursing Home Residents: A Systematic Review. Journal of the American Medical Directors Association, 17(4), 284–293. https://doi.org/10.1016/j.jamda.2015.12.017

Rosenberg, A. R., Popp, B., Dizon, D. S., El-Jawahri, A., & Spence, R. (2020). Now, More Than Ever, Is the Time for Early and Frequent Advance Care Planning. Journal of Clinical Oncology, JCO.20.01080. https://doi.org/10.1200/JCO.20.01080

Yadav, K. N., Gabler, N. B., Cooney, E., Kent, S., Kim, J., Herbst, N., Mante, A., Halpern, S. D., & Courtright, K. R. (2017). Approximately One In Three US Adults Completes Any Type Of Advance Directive For End-Of-Life Care. Health Affairs, 36(7), 1244–1251. https://doi.org/10.1377/hlthaff.2017.0175

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(**This is required for maintenance of endorsement**. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

Variation in Performance for groups/practices for the Advance Care Plan Measure, Calendar Year 2017 Data

Reporting level	N	Mean eligible population	Mean	Min	10th	25th	50th	75th	90th	Max	IQR	p-value

Group/ practice	1,031	2,803	0.74	0	0.13	0.58	0.90	1	1	1	0.42 <0.001

N: Number of groups/practices reporting

IQR: Interquartile Range

p-value: p-value of independent samples t-test comparing groups/practices at the 25th percentile to groups/practices at the 75th percentile.

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

N/A

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

0326 is reported via the QPP MIPs program and does not include disparities results.

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

In a study conducted by Kulkarni et al. (2011) among 369 multiethnic, multilingual participants, fewer than half (41%) reported that they discussed an advance care plan with their physician during a hospitalization. ACP rates are lower among older Blacks and Latinos, when compared to Whites (Carr, 2011). Estimates vary across samples, but most research studies found that when compared to their White counterparts, African Americans were less likely to participate in ACP and are more likely to informally discuss end of life care then formally document their wishes (Sanders et al, 2016). A study conducted by Harrison et al. in 2016 found that 2012 Medicare beneficiaries who were Latino, African American, were less educated, or lower income were less likely to have participated in ACP. In a study of beliefs about ACP in cancer patients at an urban, multispecialty cancer center, researchers found differences among both providers and patients in terms of their knowledge, preferences, and practices related to ACP. While 70% of patients were familiar with advance directives (100% of White patients and 45.5% of Black patients), only 35% of them reported having completed one (55.6% of White patients and 18.2% of Black patients, although not statistically significant) (Bires et al, 2017).

Bires, J. L., Franklin, E. F., Nichols, H. M., & Cagle, J. G. (2018). Advance Care Planning Communication: Oncology Patients and Providers Voice their Perspectives. Journal of Cancer Education, 33(5), 1140–1147. https://doi.org/10.1007/s13187-017-1225-4

Carr, D. (2011). Racial Differences in End-Of-Life Planning: Why Don't Blacks and Latinos Prepare for the Inevitable? Omega, 63, 1–20. https://doi.org/10.2190/OM.63.1.a

Harrison, K. L., Adrion, E. R., Ritchie, C. S., Sudore, R. L., & Smith, A. K. (2016). Low Completion and Disparities in Advance Care Planning Activities Among Older Medicare Beneficiaries. JAMA Internal Medicine, 176(12), 1872–1875. https://doi.org/10.1001/jamainternmed.2016.6751

Kulkarni, S.P., Karliner, L.S., Auerbach, A.D. & Perez-Stable, E.J. (2011). Physician Use of Advance Care Planning Discussions in a Diverse Hospitalized Population. J Immigrant Minority Health 13, 620–624. https://doi.org/10.1007/s10903-010-9361-5

Sanders, J.J., Robinson, M.T. & Block, S.D. (2016). Factors Impacting Advance Care Planning among African Americans: Results of a Systematic Integrated Review. Journal of Palliative Medicine, 19(2): p. 202 – 227.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

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

Care Coordination

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

Elderly, Populations at Risk : Dual eligible beneficiaries

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

https://qpp.cms.gov/docs/QPP_quality_measure_specifications/Claims-Registry-Measures/2019_Measure_047_MedicarePartBClaims.pdf

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

This is not an eMeasure **Attachment:**

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

No data dictionary **Attachment:**

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

No, this is not an instrument-based measure **Attachment:**

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

Not an instrument-based measure

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

No

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

N/A

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

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

Patients who have an advance care plan or surrogate decision maker documented in the medical record or documentation in the medical record that an advance care plan was discussed but patient did not wish or was not able to name a surrogate decision maker or provide an advance care plan.

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

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

Report the CPT Category II codes designated for this numerator:

- 1123F: Advance care planning discussed and documented; advance care plan or surrogate decision maker documented in the medical record

- 1124F: Advance care planning discussed and documented in the medical record; patient did not wish or was not able to name a surrogate decision maker or provide an advance care plan

Documentation that patient did not wish or was not able to name a surrogate decision maker or provide an advance care plan may also include, as appropriate, the following: That the patient's cultural and/or spiritual beliefs preclude a discussion of advance care planning, as it would be viewed as harmful to the patient´s beliefs and thus harmful to the physician-patient relationship.

NUMERATOR NOTE: The CPT Category II codes used for this measure indicate: Advance Care Planning was discussed and documented. The act of using the Category II codes on a claim indicates the provider confirmed that the Advance Care Plan was in the medical record (that is, at the point in time the code was assigned, the Advance Care Plan in the medical record was valid) or that advance care planning was discussed. The codes are required annually to ensure that the provider either confirms annually that the plan in the medical record is still appropriate or starts a new discussion.

The provider does not need to review the Advance Care Plan annually with the patient to meet the numerator criteria, documentation of a previously developed advanced care plan that is still valid in the medical record meets numerator criteria.

Services typically provided under CPT codes 99497 and 99498 satisfy the requirement of Advance Care Planning discussed and documented minutes. If a patient received these types of services, submit CPT II 1123F or 1124F.

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

All patients aged 65 years and older.

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

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

Denominator Criteria (Eligible Cases):

Patients aged > 65 years on date of encounter

AND

Patient encounter during the reporting period (CPT or HCPCS): 99201, 99202, 99203, 99204, 99205, 99212, 99213, 99214, 99215, 99218, 99219, 99220, 99221, 99222, 99223, 99231, 99232, 99233, 99234, 99235, 99236, 99291*, 99304, 99305, 99306, 99307, 99308, 99309, 99310, 99324, 99325, 99326, 99327, 99328, 99334, 99335, 99336, 99337, 99341, 99342, 99343, 99344, 99345, 99347, 99348, 99349, 99350, G0402, G0438, G0439

*Clinicians indicating the place of service as the emergency department will not be included in this measure.

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

N/A

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

N/A

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

N/A

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

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Higher score

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

Step 1: Determine the eligible population. The eligible population is all patients aged 65 years and older.

Step 2: Determine number of patients meeting the in Question S.7. above.

Step 3: Determine the number of patients who meet the numerator criteria as specified in Question S.5. above. The numerator includes all patients who have an advance care plan or surrogate decision maker documented in the medical record or documentation in the medical record that an advance care plan was discussed but patient did not wish or was not able to name a surrogate decision maker or provide an advance care plan.

Step 4: Calculate the rate by dividing the total from Step 3 by the total from Step 2.

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

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

N/A

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

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

N/A

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

If other, please describe in S.18.

Claims

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

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

None

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

No data collection instrument provided

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

Clinician : Group/Practice

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

Outpatient Services

If other:

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

N/A

2. Validity – See attached Measure Testing Submission Form

nqf_testing_attachment_7.1-637417211619924357.docx

2.1 For maintenance of endorsement

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

Yes

2.2 For maintenance of endorsement

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

Yes

2.3 For maintenance of endorsement

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

No - This measure is not risk-adjusted

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

Measure Number (*if previously endorsed*)**:** #0326
Measure Title: Advance Care Plan
Date of Submission: **8/3/2020**

Type of Measure:

Measure	Measure (continued)
☐ Outcome (*including PRO-PM*)	☐ Composite – *STOP – use composite testing form*
☐ Intermediate Clinical Outcome	☐ Cost/resource
☒ Process *(including Appropriate Use)*	☐ Efficiency
☐ Structure	*

*cell intentionally left blank

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

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

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

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.17***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ claims	☒ claims
☐ registry	☐ registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other:	☐ other:

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

2020 Submission

Testing was performed using data reported to the Centers for Medicare & Medicaid Services (CMS) Merit-Based Incentive (MIPS) Program.

2016 Submission

1.3. What are the dates of the data used in testing?

2020 Submission

Testing was performed using 2017 data.

2016 Submission

January 1, 2009- December 31, 2009

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

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.20*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☒ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other:	☐ other:

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

2020 Submission

Group/Practice: 1,031

2016 Submission

Four practice sites representing various types, locations and sizes were identified to participate in testing the measures. One practice with paper medical records and three practices with EHR participated in this testing project

- The number of geriatricians per site ranged from 1-16 in number
- The sites were located in four different regions of the United States
- Patient visit volume per site ranged from 500 – 1,000 geriatric patients per month
- Site 1 (Paper): 2,500 patients
- Site 2 (EHR): 1,800 patients
- Site 3 (EHR): 3,700 outpatients/2,000 LTC patients
- Site 4 (EHR): 2,500 patients

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

2020 Submission

Below are mean, median and ranges of denominator sizes (i.e., patients) included in the measure results.

Table 1. Advance Care Plan: Denominator Size

Reporting level	N	Mean	Min	10th	25th	50th	75th	90th	Max	Interquartile Range
Group/ practice	1,031	2803	20	148	387	1,178	2,876	6,220	74,453	2,489

2016 Submission

A random sample of 70 geriatric patient charts were identified per site; resulting in approximately 220 patient records for purposes of this study.

Sample limited to Medicare patient office visits with dates of service between January 1, 2009- December 31, 2009.

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

2020 Submission

N/A

2016 Submission

N/A

1.8 What were 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.

2020 Submission

We did not assess data by social risk factors. Social risk factor data were not available in reported results. This measure is specified for older adults, 65 years and older. NCQA is actively engaged with partners including the CMS Office of Minority Health in identifying feasible methods to further integrate social risk factors into health plan quality measures, with a focus on stratification, which will inform clinician level measures. This is aligned with recent recommendations from MedPAC and ASPE on optimal methods for addressing social risk in quality measurement and programs.[1,2] This is an NCQA wide initiative. Our intent is to implement methods to bridge data concerns in the future.

1. Medicare Payment Advisory Commission. (2020). The Medicare Advantage program: Status report. In Report to the Congress: Medicare Payment Policy (p. 397). http://medpac.gov/docs/default-source/reports/mar20_medpac_ch13_sec.pdf
2. Office of the Assistant Secretary for Planning and Evaluation, & U.S. Department of Health & Human Services. (2020). Second Report to Congress on Social Risk and Medicare's Value-Based Purchasing Programs. https://aspe.hhs.gov/social-risk-factors-and-medicares-value-basedpurchasing-programs

2016 Submission

N/A

2a2. RELIABILITY TESTING

Note: *If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)
☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)
☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level 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*)

2020 Submission

We utilized the methodology described by John Adams (Adams, J.L. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California: RAND Corporation. TR-653-NCQA, 2009) to calculate signal-to-noise reliability. This methodology uses the Beta-binomial model to assess how well one can confidently distinguish the performance of one reporting entity from another. Conceptually, the Beta-binomial model is the ratio of signal to noise. The signal is the proportion of the variability in measured performance that can be explained by real differences across reporting entities (plans, physicians, etc.) in performance. The Beta-binomial model is an appropriate model when estimating the reliability of simple pass/fail rate measures, such as the *Advance Care Plan* measure. Reliability scores range from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (i.e., noise), whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance across reporting entities.

For the *Advance Care Plan* measure, groups and practices are the reporting entity. For the formulas and explanations below, we use groups/practices as the reporting entity.

The formula for signal-to-noise reliability is:

$$\text{Signal-to-noise reliability} = \sigma^2_{\text{group/practice-to-group/practice}} / (\sigma^2_{\text{group/practice-to-group/practice}} + \sigma^2_{\text{error}})$$

Therefore, we need to estimate two variances: 1) variance between groups/practices ($\sigma^2_{\text{group/practice-to-group/practice}}$); 2) variance within groups/practices (σ^2_{error}).

1. Variance between groups/practices = $\sigma^2_{\text{group/practice-to-group/practice}} = (\alpha\,\beta) / (\alpha + \beta + 1)(\alpha + \beta)^2$

   α and β are two shape parameters of the Beta-Binomial distribution, $\alpha > 0$, $\beta > 0$

2. Variance within groups/practices: $\sigma^2_{\text{error}} = \hat{p}(1- \hat{p})/n$

   $\hat{p}$ = observed rate for the group/practice

   n = group/practice-specific denominator for the observed rate (the number of eligible patients per group/practice)

Using Adams' 2009 methodology, we estimated the reliability for each reporting entity, then averaged these reliability estimates across all reporting entities to produce a point estimate of signal-to-noise reliability. We label this point estimate "mean signal-to-noise reliability". The mean signal-to-noise reliability measures how well, on average, the measure can differentiate between reporting entity performance on the measure.

Along with the point estimate of mean signal-to-noise reliability, we are also providing:

The standard error (SE) and 95% confidence interval (95% CI) of the mean signal-to-noise reliability for all groups/practices and stratified by the denominator size (number of eligible patients per group/practice). The SE and 95% CI of the mean signal-to-noise reliability provides information about the

stability of reliability. The 95% CI is the mean signal-to-noise reliability ± (1.96*SE). The narrower the confidence interval, the less the mean signal-to-noise reliability estimate will change due to idiosyncratic features of specific groups/practices. We also stratified the results by the denominator size using terciles of the distribution to provide additional information about the stability of reliability.

The distribution (minimum, 10th, 25th, 50th, 75th, 90th, maximum) of the group/practice-level signal-to-noise reliability estimates. Each group/practice's reliability estimate is a ratio of signal to noise, as described above [$\sigma^2_{\text{group/practice-to-group/practice}} / (\sigma^2_{\text{group/practice-to-group/practice}} + \sigma^2_{\text{error}})$]. Variability between groups/practices ($\sigma^2_{\text{group/practice-to-group/practice}}$) is the same for each group/practice, while the specific group/practice error (σ^2_{error}) varies. Reliability for each group/practice is an ordinal measure of how well one can determine where a group/practice lies in the distribution across groups/practices, with higher estimates indicating better reliability. We also stratified the results by the denominator size using terciles of the distribution to provide additional information about the distribution of group/practice-level signal-to-noise reliability estimates. The number of groups/practices in each stratum and the per-group/practice denominators of the performance rates are displayed in the summary tables.

This methodology allows us to estimate the reliability for each group/practice and summarize the distribution of these estimates.

2016 Submission

Data abstracted from randomly sampled patient records were used from the AMA-PCPI Testing Project to calculate inter-rater reliability for the measure.

Data analysis included:

- **Percent agreement**
- **Kappa statistic of reliability**

Kappa: Strength of Agreement

0.00: Poor
0.01 – 0.20: Slight
0.21 – 0.40: Fair
0.41 – 0.60: Moderate
0.61 – 0.80: Substantial
0.81 – 0.99: Almost perfect

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

2020 Submission

The point estimate of mean signal-to-noise reliability using above methodology at the group/practice level is 0.999

Table 2 provides the point estimate of mean signal-to-noise reliability, its standard error, and the 95% CI for the *Advance Care Plan* measure for groups/practices overall and stratified by the denominator size (distribution of the number of eligible patients per group/practice).

Table 2. Mean Signal-To-Noise Reliability, Standard Error (SE) and 95% Confidence Interval (95% CI) for the *Advance Care Plan* Measure by Terciles of the Denominator Size and for All Submissions, Calendar Year 2017 Data

Stratification	Number of Group/Practices	Number of Eligible Patients per Group/Practice (min - max)	Mean Signal-To-Noise Reliability	SE	95% CI
All groups/ practices	**1,031**	**20 – 74,453**	**0.999**	**0.000**	**(0.998, 0.999)**
Tercile 1	340	20 – 580	0.995	0.001	(0.994, 0.996)
Tercile 2	340	583 – 2,013	0.999	0.000	(0.999, 0.999)
Tercile 3	351	2,016 – 74,453	1.000	0.000	(1.000, 1.000)

SE: Standard Error of the mean.

95% CI: 95% confidence interval.

Table 3 summarizes the distribution of group/practice-level signal-to-noise reliability estimates for the *Advance Care Plan* measure. This table also includes the distribution of group/practice-level signal-to-noise reliability estimates stratified by denominator size. Reliability estimates are higher for groups/practices with a larger denominator.

Table 3. Distribution of Group/Practice-Level Signal-To-Noise Reliability for the *Advance Care Plan* Measure by Terciles of the Denominator Size and for All Submissions, Calendar Year 2017 Data

Stratification	Number of Groups/Practices	Distribution of Group/Practice Estimates of Signal-to-Noise Reliability: Min	Distribution of Group/Practice Estimates of Signal-to-Noise Reliability: P10	Distribution of Group/Practice Estimates of Signal-to-Noise Reliability: P25	Distribution of Group/Practice Estimates of Signal-to-Noise Reliability: P50	Distribution of Group/Practice Estimates of Signal-to-Noise Reliability: P75	Distribution of Group/Practice Estimates of Signal-to-Noise Reliability: P90	Distribution of Group/Practice Estimates of Signal-to-Noise Reliability: Max
Overall	**1,031**	**0.942**	**0.997**	**0.999**	**1.000**	**1.000**	**1.000**	**1.000**
Tercile 1	340	0.928	0.988	0.995	0.999	1.000	1.000	1.000
Tercile 2	340	0.997	0.998	0.999	1.000	1.000	1.000	1.000
Tercile 3	351	0.999	1.000	1.000	1.000	1.000	1.000	1.000

2016 Submission

Advance Care Plan:
[N, % Agreement, Kappa (95% Confidence Interval)]
Denominator: 116, 99.15%, Kappa is non-calculable*
Numerator: 116, 98.28%, 0.95 (0.87 to 1.00)
Overall: 116, 98.29%, 0.95 (0.87 to 1.00)

***This is an example of the limitation of the Kappa statistic. While the agreement can be 90% or greater, if one classification category dominates, kappa can be significantly reduced.**
(http://www.ajronline.org/cgi/content/full/184/5/1391)

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

2020 Submission

The value for the signal-to-noise reliability estimate is greater than 0.9, indicating the measure has very good reliability.

Stratified analyses show that reliability increases as the number of eligible patients per group/practice increases and maintains above 0.9. Results from the stratified analyses show that reliability exceeds 0.9 for all terciles.

2016 Submission

Overall, this measure is highly reliable.

2b1. VALIDITY TESTING

2b1.1. What level of validity testing was conducted? (*may be one or both levels*)

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

☒ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*) **NOTE**: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

2b1.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

2020 Submission

Empiric Validity Testing of Performance Measure Score

We tested for construct validity by exploring whether the *Advance Care Plan* measure was correlated to the *Documentation of Current Medications in the Medical Record* measure. Both measures center on assessment which is particularly important for older adults. We hypothesized that reporting entities that perform well on the *Advance Care Plan* measure should perform well on the *Documentation of Current Medications in the Medical Record* measure.

To test these correlations, we used a Pearson correlation test. This test estimates the strength of the linear association between two continuous variables. The magnitude of correlation ranges from -1 to +1. A value of 1 indicates a strong positive linear association: an increase in values of one variable is associated with increase in value of another variable. A value of 0 indicates no linear association. A value of -1 indicates a strong negative relationship in which an increase in values of the first variable is associated with a decrease in values of the second variable. The significance of a correlation coefficient is evaluated by testing the hypothesis that an observed coefficient calculated for the sample is different from zero. The sample size for the correlation analysis is the number of groups/practices that reported both measures. The resulting p-value indicates the probability of obtaining a difference at least as large as the one observed due to chance alone. We adjusted our p-values to account for testing multiple correlations and used a threshold of 0.05 to evaluate the test results. P-values less than this threshold imply that it is unlikely that a non-zero coefficient was observed due to chance alone.

Systematic Assessment of Face Validity of Performance Measure Score

Validity refers to whether the measure represents the concept being evaluated. During development, our team reviewed the specifications and field test results with our advisory panels, which included individuals well positioned to speak to a measure's face validity. We convened a 33 member multi-stakeholder advisory panel with representation from a wide range of stakeholders, whose specialties included internal medicine, geriatrics, anesthesia, orthopedic surgery, physical medicine & rehabilitation, neurology, palliative medicine, urology, geriatric psychiatry, emergency medicine, nephrology, radiation oncology, ophthalmology, medical epidemiology, methodology, hospital medicine, family medicine, and bioethics. During measure development, the NCQA and PCPI-convened expert work groups assessed the face and content validity the measure. The group established the measure's ability to capture what it is designed to capture using a consensus process that consisting of multi-stakeholder input, including practicing physicians and experts with technical measure expertise, as well as a review of additional input received through a public comment period.

2016 Submission

The measure focuses on advance care planning in the elderly population. The evidence is consistent with the focus and scope of this measure.

As described in section 1.6, a total of 220 patient records were abstracted to complete inter-rater reliability resting of the measure concept.

An expert panel was used to assess face validity of the measure, based on the data sample. This panel consists of 33 members, whose specialties include internal medicine, geriatrics, anesthesia, orthopedic surgery, physical medicine & rehabilitation, neurology, palliative medicine, urology, geriatric psychiatry, emergency medicine, nephrology, radiation oncology, ophthalmology, medical epidemiology, methodology, hospital medicine, family medicine, and bioethics.

The full list of panel members is provided under the section Additional Information, Ad.1. Workgroup/Expert Panel Involved in Measure Development.

During measure development, the NCQA and PCPI-convened expert work groups assess the face and content validity of each measure. The groups establish the measure's ability to capture what it is designed to capture using a consensus process that consists of input from multiple stakeholders, including practicing physicians and experts with technical measure expertise, as well as a review of additional input received through a public comment period.

2b1.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)
2020 Submission

Empirical Validity Testing of Performance Measure Score

The Pearson correlation coefficient for the *Advance Care Plan* and *Documentation of Current Medications in the Medical Record* measures was as follows:

Group/Practice level: 0.63 ($p < 0.001$)

Table 4. Results of the Pearson Correlation Coefficient for groups/practices for the *Advance Care Plan* Measure, Calendar Year 2017 Data

Measure	Documentation of Current Medications in the Medical Record
Advance Care Plan	0.63
(N=, p value =)	(436, $p < 0.001$)

For face validity results, see 2016 submission information below.

2016 Submission

As described in section 2a2.1, a total of 220 patient records were abstracted to complete inter-rater reliability resting of the measure concept.

An expert panel was used to assess face validity of the measure, based on the data sample. This panel consists of 33 members, whose specialties include internal medicine, geriatrics, anesthesia, orthopedic surgery, physical medicine & rehabilitation, neurology, palliative medicine, urology, geriatric psychiatry, emergency medicine, nephrology, radiation oncology, ophthalmology, medical epidemiology, methodology, hospital medicine, family medicine, and bioethics.

The full list of panel members is provided under the section Additional Information, Ad.1. Workgroup/Expert Panel Involved in Measure Development.

During measure development, the NCQA and PCPI-convened expert work groups assess the face and content validity of each measure. The groups establish the measure's ability to capture what it is designed to capture using a consensus process that consists of input from multiple stakeholders, including practicing physicians and experts with technical measure expertise, as well as a review of additional input received through a public comment period.

This measure was deemed valid by the expert panel.
The aforementioned panel was asked to rate their agreement with the following statement:
"The scores obtained from the measure as specified will accurately differentiate quality across providers."

Scale 1-5, where 1=Strongly Disagree; 3=Neither Disagree nor Agree; 5=Strongly Agree

The results of the expert panel rating of the validity statement were as follows:

N = 23 Mean rating = 4.35

Frequency Distribution of Ratings

(1) - Strongly Disagree - 0 panel members

(2) - Disagree - 0 panel members

(3) - Neither Disagree nor Agree - 4 panel members

(4) - Agree - 7 panel members

(5) - Strongly Agree - 12 panel members

2b1.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

2020 Submission

For the purposes of this analysis and the intended use of this measure to evaluate the quality of documented care across practices, correlation is considered high (strong) if the correlation coefficient is 0.75 to 1, moderate if 0.25 to 0.75, and low (weak) if 0 to 0.25.

The correlation value of 0.63 is moderate, suggesting that reporting entities that performed well on the *Advance Care Plan* measure are moderately likely to perform well on the *Documentation of Current Medications in the Medical Record* measure.

2016 Submission

These results indicate that the multiple experts and stakeholders concluded with good agreement that the measure as specified accurately captures quality. Our interpretation of these results is that this measure meets the test for face validity.

_____

2b2. EXCLUSIONS ANALYSIS

NA ☒ no exclusions — *skip to section 2b3*

2b2.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

2b2.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis.* **Note***: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

_____

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.

2b3.1. What method of controlling for differences in case mix is used?

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with risk factors**

☐ **Stratification by risk categories**

☐ **Other,**

2b3.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

2b3.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:

☐ **Published literature**

☐ **Internal data analysis**

☐ **Other (please describe)**

2b3.4a. What were the statistical results of the analyses used to select risk factors?

2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors *(e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b3.9

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b3.9. Results of Risk Stratification Analysis:

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b3.11. Optional Additional Testing for Risk Adjustment (**not required**, *but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b4.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

2020 Submission

To demonstrate meaningful differences in performance, we calculated an inter-quartile range (IQR) for the indicator. The IQR provides a measure of the dispersion of performance. The IQR can be interpreted as the difference between the 25th and 75th percentile on a measure.

To determine if this difference is statistically significant, we calculated an independent sample t-test of the performance difference between two randomly selected group practices at the 25th and 75th percentile. The t-

test method calculates a testing statistic based on the sample size, performance rate, and standard error of each practice. The test statistic is then compared against a t- distribution, which is similar to a normal distribution. If the p value of the test statistic is less than 0.05, then the two practices' performance is significantly different from each other.

2016 Submission

This measure is used in the CMS Physician Quality Reporting Initiative/System (PQRI/S) in the claims (2007-2016) and registry (2009-2016) options. We are using CMS data on average performance rates by providers to determine if statistically significant and clinically meaningful differences in performance measure scores can be identified.

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

2020 Submission

Table 5 summarizes the distribution of group/practice-level performance for the *Advance Care Plan* measure.

Table 5. Variation in Performance for groups/practices for the *Advance Care Plan* Measure, Calendar Year 2017 Data

Reporting level	N	Mean eligible population	Mean	Min	10th	25th	50th	75th	90th	Max	IQR	p-value
Group/ practice	1,031	2,803	0.74	0	0.13	0.58	0.90	1	1	1	0.42	<0.001

N: Number of groups/practices reporting

IQR: Interquartile Range

p-value: p-value of independent samples t-test comparing groups/practices at the 25th percentile to groups/practices at the 75th percentile.

2016 Submission

The most recent data available is from 2014. We are including data from the PQRS claims and registry options from 2012-2014 to demonstrate overall performance and reporting trends. CMS is unable to provide us with more detailed data (such as performance at benchmark percentiles, as provided in the past). Nevertheless, the new data shows the continued gap in care.

TRENDS IN INDIVIDUAL MEASURE PERFORMANCE RATE, FOR ELIGIBLE PROFESSIONALS (EPS) WHO SUBMITTED THE MEASURE CONTINUOUSLY FROM 2012 TO 2014:

- EPs who Reported Continuously 2012-2014: 3,309
- Average Performance Rate in 2012: 62.3%
- Average Performance Rate in 2013: 63.7%
- Average Performance Rate in 2014: 67.2%
- Growth Rate: 3.9%

SUBMITTING EPS WITH AT LEAST A 90% PERFORMANCE RATE BY INDIVIDUAL MEASURE (2014)

- Percent of EPs with At Least 90% Performance Rate: 43.0%

AVERAGE PERFORMANCE RATE FOR GROUP PRACTICES REPORTING VIA REGISTRY (2014)

- Small Group Practice Reporting Option: 64.5%
- Medium Group Practice Reporting Option: 52.0%
- Large Group Practice Reporting Option: 35.6%

2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

2020 Submission

For groups/practices, the IQR was 42 percentage points. This gap represents an average of 1,177 additional older adult patients having an advance care plan in high-performing practices versus low-performing practices.

2016 Submission

Our interpretation of the results is that performance among reporting EPs has improved between 2012 and 2014, but there is still more room for improvement. In 2014, almost one-third of patients of reporting EPs did not have evidence of an advance care plan documented in the medical record or evidence that such a plan was discussed.

In addition, we see meaningful differences in performance rates for group practices reporting via a registry based on practice size, with a trend toward better performance (64.5%) for the small group practice reporting option in comparison to the large group practice reporting option (35.6%).

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

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

Note*: This item is directed to measures that are risk-adjusted (with or without 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 eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without 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.***

2b5.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)
2020 Submission

We are not able to conduct missing data analysis due to limitations of publicly available data from CMS.

2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse;* **if no empirical sensitivity analysis**, *identify the approaches for handling missing data that were considered and pros and cons of each*)

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted;* ***if no empirical analysis****, provide rationale for the selected approach for missing data*)

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)

If other:

3b. Electronic Sources

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

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

ALL data elements are in defined fields in a combination of electronic sources

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

This measure uses the Clinical Quality Measure reporting method. Some components of this measure draw on structured fields, while others are available in narrative notes or other non-structured fields.

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

Attachment:

3c. Data Collection Strategy

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

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

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

This measure has been in use for 12 years (2008 to present) in the PQRI, PQRS and QPP/MIPS quality reporting program with no feedback received from the field citing difficulties reporting the measure.

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

Broad public use and dissemination of this measure is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, "commercial use" refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
*	Public Reporting CMS Merit-Based Incentive Payment System (data reported publicly via Physician Compare) https://data.medicare.gov/data/physician-compare Payment Program CMS Merit-based Incentive Payment System (MIPS) https://qpp.cms.gov/mips/overview

*cell intentionally left blank

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

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

CMS QUALITY PAYMENT PROGRAM: This measure is used in the Quality Payment Program (QPP) which is a quality and cost incentive program that uses payment adjustments to promote high quality and high value care delivery by eligible clinicians (EC). QPP provides performance-based payment adjustments to ECs, both negative and positive, for services furnished to Medicare Part B beneficiaries. EC performance is graded on quality measure performance, cost of care, engagement in clinical practice improvement activities, and use of Certified EHR Technology (CEHRT). Performance can be reported at the individual (clinician) or group (practice) level. In 2018, 874,515 ECs participated in MIPS, representing 98% of all eligible clinicians across the 50 states. 53% participated as a part of a group, 6% as individual clinicians, and 41% as a part of an Advanced Payment Model.

References:

Centers for Medicare and Medicaid Services (CMS). (2020). 2018 Quality Payment Program Reporting Experience. Baltimore, MD: Centers for Medicare and Medicaid Services. Retrieved from: https://qpp.cms.gov/about/resource-library

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

N/A - this measure is publicly reported through the Merit-Based Incentive Payment System (MIPS).

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

N/A

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

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

When clinicians and groups report their performance on this measure for MIPS, CMS provides them with feedback reports to inform performance improvement efforts. All individual and group performance data is reported publicly through Physician Compare, and annual benchmarks are publicly available to enable clinicians to understand how their performance compares to national benchmarks.

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

Feedback reports are made available to clinicians in July following the measurement year (i.e. feedback reports for 2017 were available in July of 2018). Reports include performance rates as well as the associated payment adjustment to Medicare Part B payments. The full performance data set became available on Physician Compare in 2019. The Physician Compare data set includes measure performance scores for all individual clinicians and groups that reported measures to MIPS.

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

Describe how feedback was obtained.

NCQA assesses the measure annually as part of the CMS MIPS program, culminating in the Measures Finalization Meeting convened by CMS and its contractors. We have not received feedback from measured entities specific to this measure during that process.

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

We have not received feedback from those being measured via CMS or NCQA portals.

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

We have not received any additional feedback on this measure.

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

No feedback has been received that indicate the need for modification.

Improvement

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

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

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

2017 is the only year that publicly reported MIPS data currently is available on Physician Compare for this measure. Thus, we are unable to describe demonstrated performance improvement year over year.

4b2. Unintended Consequences

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

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

No unintended consequences were identified for this measure.

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

Unexpected benefits could include better more transparent, provider-patient communication about advance care planning. Advance care planning is a critical part of patient care, as it can lead to improved end of life

care, increased trust in providers, decreased psychological distress, improved quality of life and can facilitate hope (Rosenberg et al., 2020).

Rosenberg, A. R., Popp, B., Dizon, D. S., El-Jawahri, A., & Spence, R. (2020). Now, More Than Ever, Is the Time for Early and Frequent Advance Care Planning. Journal of Clinical Oncology, JCO.20.01080. https://doi.org/10.1200/JCO.20.01080

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

No

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

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

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

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

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

5b. Competing Measures

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

OR

Multiple measures are justified.

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

N/A

Appendix

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

No appendix Attachment:

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): National Committee for Quality Assurance

Co.2 Point of Contact: Bob, Rehm, nqf@ncqa.org, 202-955-1728-

Co.3 Measure Developer if different from Measure Steward: National Committee for Quality Assurance

Co.4 Point of Contact: Brittany, Wade, wade@ncqa.org, 202-530-0463-

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

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

An expert panel was used to assess face validity of the measure. The panel consists of 33 members, whose specialties include internal medicine, geriatrics, anesthesia, orthopedic surgery, physical medicine & rehabilitation, neurology, palliative medicine, urology, geriatric psychiatry, emergency medicine, nephrology, radiation oncology, ophthalmology, medical epidemiology, methodology, hospital medicine, family medicine, and bioethics.

Caroline Blaum, MD (Work Group Co-Chair) (Geriatrics/Internal Medicine) Associate Professor of Internal Medicine, University of Michigan, Ann Arbor, MI

Carol M. Mangione, MD (Work Group Co-Chair) (Internal Medicine) Professor of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA

Chris Alexander, III, MD, FACP (Methodology) Social Security Administration, Office of Hearings and Appeals, Earlysville, VA

Patricia P. Barry, MD, MPH (Internal Medicine) American College of Physicians, Gloucester Point, VA

Frederick W. Burgess, MD, PhD (Anesthesia) Rhode Island Hospital, Department of Anesthesia, Providence, RI

Gary S. Clark, MD, MMM, CPE (Physical Medicine & Rehabilitation) Professor and Chair, MetroHealth Medical Center, Dept. of PM&R, Cleveland, OH

Eric Coleman, MD, MPH (Geriatrics) Associate Professor, Division of Health Care Policy and Research, University of Colorado Health Services Center, Aurora, CO

Stephen R. Connor, PhD Vice President, Research and International Development, National Hospice and Palliative Care Organization, Alexandria, VA

Gail A. Cooney, MD (Neurology, Palliative Medicine) Hospice of Palm Beach County, West Palm Beach, FL

Roger Dmochowski, MD (Urology) Department of Urologic Surgery, Vanderbilt University, Nashville, TN

Catherine DuBeau, MD (Geriatrics) Associate Professor of Medicine, University of Chicago, Chicago, IL

Joyce Dubow Associate Director, AARP Policy Institute, Washington, DC

Mary Fermazin, MD, MPA (Internal Medicine) Vice President, Health Policy & Quality Measurement, Health Services Advisory Group, Inc., Phoenix, AZ

Sanford I. Finkel, MD (Geriatric Psychiatry) Professor of Clinical Psychiatry, University of Chicago Medical School, Wilmette, IL

Terry Fulmer, PhD Dean, NYU College of Nursing, New York, NY

Peter Hollmann, MD (Internal Medicine/Geriatrics) Blue Cross Blue Shield, Cranston, RI

David P. John, MD (Emergency Medicine) Chair Geriatric Section, ACEP, North Haven, CT

Peter Johnstone, MD, FACR (Radiation Oncology) Professor and Chair of Radiation Oncology, Indiana University School of Medicine, Department of Radiation Oncology, Indianapolis, IN

Flora Lum, MD American Academy of Ophthalmology, Director, Quality of Care & Knowledge Base Development, San Francisco, CA

Diane E. Meier, MD Professor, Director: Hertzberg Palliative Care Institute, Director: Center to Advance Palliative Care, Mount Sinai School of Medicine, Department of Geriatrics, New York, NY

Alvin "Woody" H. Moss, MD (Nephrology and Palliative Care) Professor of Medicine & Director, Center for Health Ethics & Law, Section of Nephrology, West Virginia University, Morgantown, WV

Jaya Rao, MD, MHS Associate Professor, Pharmaceutical Outcomes and Policy, UNC Eshelman School of Pharmacy, Chapel Hill NC

Sam J. W. Romeo, MD, MBA General Partner, Tower Health & Wellness Center, LP, Turlock, CA

David J. Satin, MD (Family Medicine/Bioethics) Assistant Professor, University of Minnesota, Minneapolis, MN

Gregory B. Seymann, MD (Internal Medicine/Hospital Medicine) Associate Professor, Division of Hospital Medicine, UCSD School of Medicine, San Diego, CA

Knight Steel, MD (Internal Medicine/Geriatrics) Chief, Geriatrics, Internist, Professor of Medicine Emeritus, Hackensack University Medical Center, Hackensack, NJ

Eric Tangalos, MD (Internal Medicine/Geriatrics) Co-Director, Program on Aging, Mayo Clinic, Rochester, MN

Joan M. Teno, MD, MS (Geriatrics/Palliative Care) Professor of Community Health and Medicine, Brown Medical School, Providence, RI

David J. Thurman, MD, MPH CDC, Atlanta, GA

Mary Tinetti, MD (Internal Medicine/Geriatrics) Gladys Phillips Crofoot Professor of Medicine, Epidemiology and Public Health, Yale University School of Medicine, Section of Geriatrics, New Haven, CT

Laura Tosi, MD (Orthopaedic Surgery) American Academy of Orthopaedic Surgery, Director, Bone Health Program, Washington, DC

Gregg Warshaw, MD Director, Office of Geriatric Medicine, University of Cincinnati College of Medicine, Cincinnati, OH

Neil S. Wenger, MD (Internal Medicine/Geriatrics) Professor of Medicine, UCLA, Los Angeles, CA

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2008

Ad.3 Month and Year of most recent revision: 02, 2008

Ad.4 What is your frequency for review/update of this measure? As needed based on revisions or updates to applicable clinical guidelines.

Ad.5 When is the next scheduled review/update for this measure? 12, 2021

Ad.7 Disclaimers: These performance measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications.

THE MEASURES AND SPECIFICATIONS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND.

Ad.8 Additional Information/Comments: NCQA Notice of Use. Broad public use and dissemination of these measures is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, "commercial use" refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

These performance measures were developed and are owned by NCQA. They are not clinical guidelines and do not establish a standard of medical care. NCQA makes no representations, warranties or endorsement about the quality of any organization or physician that uses or reports performance measures, and NCQA has no liability to anyone who relies on such measures. NCQA holds a copyright in these measures and can rescind or alter these measures at any time. Users of the measures shall not have the right to alter, enhance or otherwise modify the measures, and shall not disassemble, recompile or reverse engineer the source code or object code relating to the measures. Anyone desiring to use or reproduce the measures without modification for a noncommercial purpose may do so without obtaining approval from NCQA. All commercial uses must be approved by NCQA and are subject to a license at the discretion of NCQA.