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: Click 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 #: 3497

Corresponding Measures:

De.2. Measure Title: Evaluation of Functional Status (Basic and Instrumental Activities of Daily Living [ADL]) for Home-Based Primary Care and Palliative Care Patients

Co.1.1. Measure Steward: American Academy of Home Care Medicine

De.3. Brief Description of Measure: Percentage of actively enrolled home-based primary care and palliative care patients who receive an ADL and IADL assessment.

*Basic ADLs must include but are not limited to: bathing, transferring, toileting, and feeding; Instrumental ADLs (IADL) must include but are not limited to: telephone use and managing own medicationsPercentage of actively enrolled home-based primary care and palliative care patients who receive an ADL and IADL assessment.

*Basic ADLs must include but are not limited to: bathing, transferring, toileting, and feeding; Instrumental ADLs (IADL) must include but are not limited to: telephone use and managing own medications

1b.1. Developer Rationale: Functional limitations, frailty, and homebound status render approximately 4 million US adults unable to easily access office-based primary care (1). Compared to the non-homebound population, homebound individuals have twice as many chronic conditions, are more likely to have depression or dementia, are nearly two to two-and-a-half times more likely to have been hospitalized in the past year, and are among the most costly and sickest patients in the US healthcare system (2-4). Further, one study of Medicare beneficiaries found that 50.7% are homebound in the last year of life, suggesting that half are unable to receive office-based care at a time when they may have the most need (5). To address these care access and cost issues, home-based primary and palliative care is a model growing with the support of providers and policymakers alike (2). Between 2000 and 2006, the number of physician house calls to Medicare beneficiaries more than doubled while the number of providers practicing home care has decreased (6). This suggests a growing emphasis on home-based primary and palliative care as a focused medical specialty, with dedicated providers seeing a greater percentage of their patient panels in the patients' homes than before (3,6). Seizing on the opportunities inherent in this emerging practice model, the Centers for Medicare and Medicaid Services (CMS) has invested millions of dollars into the Independence at Home (IAH) Demonstration, which tests a payment incentive and service delivery model for home based primary care (2). Outside of this demonstration project, however, there are no currently endorsed performance measures specific to home-based primary care, leaving a growing, high priority practice model with disease-specific measures that do not take into account the clinical complexity and multi-morbidity of the homebound population (3).

Poor functional status is a major contributor to costs for the homebound, with nearly half of patients in the IAH Demonstration reporting the need for assistance with four or more Activities of Daily Living (ADLs) (2). The high cost of care associated with multimorbidity in chronic illness is largely driven by functional limitations, not the presence of multiple chronic conditions alone. Medicare enrollees with functional limitations and chronic conditions account for one third of total Medicare spending—this is twice the average amount for Medicare beneficiaries with three or more chronic conditions but without functional limitations (7). Hospital spending is the largest source of this gap, with Medicare enrollees with both chronic conditions and functional limitations using the Emergency Department and admitted to the hospital more than enrollees with multiple chronic conditions and no functional limitations (7). Home-based primary and palliative care providers have an opportunity to play a role in preventing these costly admissions by engaging homebound patients in functional assessments. Identifying functional needs allows for both treatment and linkage with resources to prevent falls and enhance in-home safety. Although there are no dedicated home-based primary care guidelines based on systematic review, functional assessments are supported in palliative and geriatric nursing guidelines (3,8,9). Despite these recommendations, current provider performance demonstrate an opportunity for improvement. 2017-2018 registry data for this measure, based on 221 providers and 64,394 quality events indicate an average performance rate of 67%, with a range from 16% to 93%. This variability and lower rate of performance indicate that home-based primary and palliative care providers have room to improve on functional assessment rates in care of the homebound population. Achieving high performance should ultimately lead to enhanced patient-centered care in the home and improved outcomes for this complex population and is a crucial first step in achieving many high priority quality of care standards identified by The National Home-Based Primary and Palliative Care Network (3).

1. Qiu WQ, Dean M, Liu T, et al. Physical and mental health of homebound older adults: an overlooked population. J Am Geriatr Soc. 2010;58(12):2423-2428.

2. Report to Congress: Evaluation of the Independence at Home Demonstration. Centers for Medicare and Medicaid Services;2018.

3. Leff B, Carlson CM, Saliba D, Ritchie C. The invisible homebound: setting quality-of-care standards for home-based primary and palliative care. Health Aff (Millwood). 2015;34(1):21-29.

4. Ornstein KA, Leff B, Covinsky KE, et al. Epidemiology of the Homebound Population in the United States. JAMA internal medicine. 2015;175(7):1180-1186.

5. Soones T, Federman A, Leff B, Siu AL, Ornstein K. Two-Year Mortality in Homebound Older Adults: An Analysis of the National Health and Aging Trends Study. J Am Geriatr Soc. 2017;65(1):123-129.

6. Peterson LE, Landers SH, Bazemore A. Trends in physician house calls to Medicare beneficiaries. J Am Board Fam Med. 2012;25(6):862-868.

7. Komisar HL, Feder J. Transforming Care for Medicare Beneficiaries with Chronic Conditions and Long-Term Care Needs: Coordinating Care Across All Services. Georgetown University;2011.

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

9. Kresevic DM. Assessment of physical function. In: Boltz M, Capezuti E, Fulmer T, Zwicker D, editor(s). Evidence-based geriatric nursing protocols for best practice. 4th ed. New York (NY): Springer Publishing Company; 2012. p. 89-103.

S.4. Numerator Statement: Submission Criteria 1 - Newly enrolled:

Number of newly enrolled home-based primary care and palliative care patients who were assessed for basic ADL and IADL impairment at enrollment.

Submission Criteria 2 - Established patients:

Number of established home-based primary care and palliative care patients who were assessed for ADL and IADL impairment at enrollment and annually

S.6. Denominator Statement: Submission Criteria 1 - Newly enrolled:

Total number of newly enrolled (and active) home-based primary care and palliative care patients. The enrollment period includes 90 days from the first recorded new patient E&M visit code with the practice. *A patient is considered active if they have at least 2 E&M visit codes with a provider from the practice within the reporting period.

Submission Criteria 2 - Established patients:

Total number of established enrolled and active home-based primary care and palliative care patients. A patient is considered established if they have at least 2 Established Patient Encounter E&M visit codes with a provider from the practice within the reporting period.

S.8. Denominator Exclusions: Submission Criteria 1 - Newly enrolled:

Denominator Exceptions:

Most recent new patient encounter (with subsequent established patient encounter) occurs within the last 90 days of the measurement period

Submission Criteria 2 - Established patients:

There are no exceptions or exclusions for this submission criteria.

De.1. Measure Type: Process

S.17. Data Source: Registry Data

S.20. Level of Analysis: Clinician : Individual

Preliminary Analysis: New Measure

Criteria 1: Importance to Measure and Report

1a. Evidence

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

The developer provides the following evidence for this measure:

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

Evidence Summary

The developer cited three clinical practice guidelines to support this measure. These include:

- Clinical Practice Guidelines for Quality Palliative Care, 4th edition (2018)
    - Three recommendations from the guidelines support conducting an initial functional status assessment, ongoing assessments of functional status, and assessment of functional

limitations that impact activities of daily living (ADLs) and instrumental activities of daily living (IADLs).

- These recommendations are not associated with graded evidence in the guidelines.
- Grades were assigned to evidence related to Key Questions (KQ) included in a systematic review, four of which the developer used in support of this measure. The developer did summarize the number of studies and quality of the evidence related to these four KQs***. NOTE, however, that these KQs do not directly address the linkage between functional status assessment and improved patient outcomes.***
  - KQ1a: What is the effect of interdisciplinary team care on patient and family/caregiver outcomes?
    - There is moderate-quality evidence supporting positive impact of interdisciplinary teams on quality of life on advance care planning decisions, death at home, and patient and family satisfaction with care, but **low-quality evidence** for a positive impact of interdisciplinary palliative care teams on patient physical symptoms and patient and family psychological outcomes.
  - KQ1b: What is the impact of palliative care interventions to improve continuity and coordination of care on patient and family/caregiver outcomes?
    - There is high-quality evidence supporting positive impact of home-based palliative care for the outcome of dying at home, but **low/very low-quality evidence** for home-based palliative care on other outcomes including physical symptoms, quality of life, and family/caregiver burden and satisfaction.
  - KQ 2: What Is the Impact of Palliative Care Interventions on Physical Symptom Screening, Assessment, and Management of Patients?
    - Much evidence graded as **low-quality**, due to inconsistent findings regarding impact on patient symptoms
  - KQ 4: Does an Assessment of Environmental or Social Needs as Part of a Comprehensive Palliative Assessment Improve Needs Identification and Access to Relevant Services?
    - Evidence regarding social work interventions result in greater access to social service is graded as **low-quality**.
    - Caregiver support services interventions had a positive, but not statistically significant or consistent increase in social support.

- Assessment of physical function. In: Evidence-based geriatric nursing protocols for best practice
  - Includes three recommendations to assess function.
  - The quality of the evidence for these recommendations is graded as Level IV (non-experimental study, n=1), Level V (care report/program evaluation/narrative literature reviews, n=1), and Level VI (expert opinion, n=2). ***NQF does not accept case reports/expert opinion as evidence***. Thus, only the one (1990) non-experimental study may be relevant in terms of NQF requirements (although this is not certain without additional information regarding the study).
- Age-related changes in health. In: Evidence-based geriatric nursing protocols for best practice
  - Includes one recommendation to assess function.
  - The quality of the evidence for this recommendation is graded as Level V (care report/program evaluation/narrative literature reviews, n=1). ***NQF does not accept case reports/expert opinion as evidence***.

The developer also described meta-analyses and systematic reviews that have assessed the value of comprehensive geriatric assessments (CGA) for older adults in a variety of care settings. The developer notes that a functional status assessment is always a part of this comprehensive geriatric assessment.

- Two of these focused specifically on community-dwelling older adults in the context of home-based care.
    - Elkan, et al., 2001. This meta-analysis of 8 studies assessed the effectiveness of home-based support for older adults. Findings varied across the studies, but the evidence suggests that home-based care of older adults is associated with reduced mortality and fewer admissions to institutional long-term care. However, this care was not associated with statistically significant reductions in hospital admission, health, or ADLs.
        - Staff note: Developers did not specifically describe the quality of studies included in this meta-analysis. A brief perusal of the Elkan, et al. article indicates that most of the studies included were fair-to-moderate quality randomized trials.
    - Huss, et al., 2009. This meta-analysis of 21 studies assessed the effectiveness of preventive home visit programs in community-dwelling older adults. The evidence suggests these programs reduce mortality in younger populations, but not in older populations. Inclusion of a clinical exam at initial assessment was associated with a reduction in functional decline, although this relationship was not seen with other interventions. The effect of various interventions on nursing home admissions was inconclusive.
        - Staff note: Developers did not specifically describe the quality of studies included in this meta-analysis. A brief perusal of the Huss, et al. article indicates that the studies included were analyses of randomized trials. While not all of the studies included adequate blinding, study authors did not find that this influenced results.
- Studies in other settings have also shown that comprehensive geriatric assessments are associated with reductions in hospital and nursing home admissions, although their benefit in other endpoints such as dependence, length of hospital stay, quality of life, and cognition is less clear.

Questions for the Committee:

- To what extent is the evidence that is provided directly applicable to functional status assessment in home-based primary care and palliative care patients?
- How strong is the evidence linking functional status assessment to improved patient outcomes such as home comfort and safety or other improvements in care (or at least to improved care planning and management)?

Guidance from the Evidence Algorithm

Measure assesses a process of care and relies on systematic review of the evidence (Box 3) → QQC mostly provided (Box 4) → Quantity: high; Quality: fair-to moderate; consistency: moderate [for long-term care admissions] (Box 5b) → MODERATE

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

RATIONALE: The majority the evidence supporting the recommended clinical practice guidelines does not meet NQF's requirements. However, the various meta-analyses/reviews focusing on comprehensive geriatric assessments suggest a link between these assessments (which include functional status assessments) and reduced admissions to institutional long-term care. Associations with other outcomes such as mortality and hospital admissions are not consistent across the studies. The numerous studies included in these meta-analyses/reviews appear to be of fair-to-moderate quality.

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

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

2017-2018 National Home-Based Primary Care & Palliative Care Registry data provided by the developer (based on 221 providers and 64,394 quality events) indicates the following:

- Mean performance rate: 67%
- Standard deviation: 15%
- Minimum: 16%
- 25th percentile: 57%
- 75th percentile: 77%
- Maximum: 93%

Disparities

The developer did not provide performance data for population subgroups.

Questions for the Committee:

- Is there a gap in care that warrants a national performance measure?
- Are you aware of evidence of disparities in the assessment of functional status among home-based primary and palliative care patients?

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

Comments:

- I agree with the staff's preliminary rating of moderate given the fair to moderate level of quality of the evidence. I appreciate the steward's combination of separate ADL/IADL measures to expedite the process.
- Three guidelines noted. Some of the evidence is a bit tangential. The most compelling to me was, "Studies in other settings have also shown that comprehensive geriatric assessments are associated with reductions in hospital and nursing home admissions..."
- I would rate the evidence as moderate. I am including a cite to an article on the CAPABLE Program: Sarah L. Szanton, Bruce Leff, Jennifer L. Wolff, Laken Roberts, and Laura N. Gitlin (2016). Home-Based Care Program Reduces Disability And Promotes Aging In Place. HEALTH AFFAIRS 35,NO. 9 (2016): 1558-1563. I am also including a cite to a proposed study on this topic. Although the study has not been conducted yet, there are studies referenced that may provide additional evidence. Reckrey et al. (2018). Rationale and design of a randomized controlled trial of home-based primary care versus usual care for high-risk homebound older adults. Contemporary Clinical Trials, 68. 90-94
- The developer provided a systematic review of evidence specific to this measure, there was high quality, quantity and consistency of the evidence and it was graded. Although the majority of the evidence does not meet NQF's standards, the various meta-analyses/reviews suggest a link between functional status assessments and reduced admissions to institutional long-term care.
- process measure; palliative care clinical practice guidelines support measurement of functional status. NQF did not accept geriatric nursing protocols yet those protocols went through the AHRQ evidence-based review to be included on the National Guideline Clearinghouse

1b. Performance Gap

Comments:

- I agree that there is important room for improvement and think that measuring functional status in this population can lead to an important focus on quality of life.
- Significant gap with a mean performance rate of 67%
- Low - the developer did not provide performance data for subgroups
- There is definitely a need for improvement with a mean performance rate of 67%, and a range of 16%-93%. Disparities in care were not demonstrated.
- gap in care indicated that warrants performance measure - no subgroup data provided

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: Staff review

Staff Evaluation Summary:

Reliability

- Reliability testing was conducted at the measure score level.
- Testing of the measure score:
  - Developers conducted a signal-to-noise analysis using the Adams beta-binomial method. Testing was conducted for the measure as specified.
  - Data used in testing were obtained from the National Home-Based Primary Care & Palliative Care Registry during the period between 11/1/2017-10/31/2018 (n=221 providers; 64,394 quality events (i.e., patients)).
  - Developers provided an average reliability estimate as well as reliability estimates by decile.
    - Signal-to-noise reliability estimate: Mean=0.95
    - Reliability estimates by decile ranged from 0.94 to 0.99.

Validity

- Validity testing was conducted at the measure score level via a face validity assessment.
- Face validity of the performance measure score was assessed by 12 experts.

    - Of these, 11 (92%) either agreed or strongly agreed that this measure can accurately distinguish good and poor quality, while one person disagreed with this statement.
    - The average rating was 4.5 (from a 5-point scale).
    - The individual who disagreed with the statement regarding the measure's ability to differentiate between good and poor quality did not offer a rationale.
- This measure is not risk-adjusted.
- Individuals excluded from this measure include patients who enroll within the last 90 days of the measurement period (the rationale is there may not be opportunity to do the assessment during the measurement time period). Analyses provided by the developer indicate that there are relatively few exceptions per provider and the number of exceptions was higher for providers who had more patients.
- In response to the question regarding whether there are meaningful differences in performance scores between providers, developers pointed to the wide variation in performance, even after outliers were excluded from the analysis.
- The dataset used in testing did not have missing values. Developers note that data may have been rejected by the registry but no information regarding how often this might occur was provided.

Questions for the Committee regarding reliability:

- Do you have any questions about the measure specifications?
- Do you have any concerns that the measure can be consistently implemented?
- NQF staff is satisfied with the reliability testing for the measure (both the methods and the results). Does the Committee think there is a need to vote on reliability?

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, missing data, etc.)?
- Do you have any questions regarding how the registry is populated?
- NQF staff is satisfied with the validity analyses for the measure (including testing results and analyses to examine potential threats to validity). Does the Committee think there is a need to vote on validity?

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

Measure Number: 3497

Measure Title: Evaluation of Functional Status (Basic and Instrumental Activities of Daily Living [ADL]) for Home-Based Primary Care and Palliative Care Patients

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_xxxx" 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.**

None.

RELIABILITY: TESTING

Submission document: "MIF_xxxx" 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 <u>VALIDITY</u> testing** of <u>patient-level data</u> conducted?

   ☐ **Yes** ☐ **No**

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

   Submission document: Testing attachment, section 2a2.2

Developers tested the reliability of the measure using data from the National Home-Based Primary Care & Palliative Care Registry during the period between 11/1/2017-10/31/2018 (n=221 providers; 64,394 quality events (i.e., patients).

To assess reliability of the measure score, developers conducted a signal-to-noise analysis using the Adams beta-binomial method. This is an appropriate method to assess reliability. Developers tested according to the measure specifications. Developers also provided reliability estimates by decile.

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

   Submission document: Testing attachment, section 2a2.3

Signal-to-noise reliability estimates: Mean=0.95. Average reliability by decile ranged from 0.94 to 0.99.

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.**

Testing was conducted for the measure as specified using an appropriate methodology. Results indicate reliability estimates ranging from 0.94 through 0.99.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

None. Exceptions are for newly-enrolled patients who enroll within the last 90 days of the measurement period (thus, there may not be opportunity to do the assessment within the required 90-day timeframe). This seems to be a reasonable exception. There are relatively few exceptions per provider, and as expected, the number of exceptions increased with the number of patients.

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

    Submission document: Testing attachment, section 2b4.

None. There is wide variation in performance, even after outliers were excluded from the analysis.

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.

N/A: This measure is specified for only one data source/with one set of specifications, and is not risk-adjusted.

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

Submission document: Testing attachment, section 2b6.

The data used in testing did not have missing data. Developers note that data may have been rejected by the registry but no information regarding how often this might occur was provided**.**

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**

Not applicable – this is a simple process measure.

<u>For cost/resource use measures ONLY</u>:

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

Developers conducted a face validity assessment through a systematic and transparent process by identified experts. The question posed to the experts explicitly addressee whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. Developers provided data to demonstrate the degree of consensus. They did not discuss areas of disagreement but noted that the one panel member who disagreed with the question used to assess face validity did not provide a rationale for their disagreement.

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

    Submission document: Testing attachment, section 2b2.3

    Of the 12 experts who participated in the face validity assessment, 11 either agreed or strongly agreed that the scores obtained from the measure as specified will accurately differentiate quality across providers. One of the experts disagreed with this statement but did not provide a rationale.

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.**

Good results from a systematic assessment of face validity.

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.

Committee Pre-evaluation Comments:
Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

2a1. Reliability-Specifications

Comments:

- I have no concerns.
- Reliability is reasonable with ï‚§ Signal-to-noise reliability estimate: Mean=0.95 ï‚§ Reliability estimates by decile ranged from 0.94 to 0.99.
- I have no concerns with reliability specifications
- I do not have any concerns.
- no concerns noted

2a2. Reliability – Testing

Comments:

- No concerns.
- I have no concerns
- no
- none noted

2b1. Validity –Testing

Comments:

- face validity done with a group or experts. No concerns.
- Face validity is the lowest test for validity but the experts were in agreement (except for the one expert but we do not know why they had concerns
- no
- none noted

2b4-7. Threats to Validity

Comments:

- I think it is curious that there was no missing data and no information about how often this occurred was provided.
- Developers show, The range of the performance rate is 0.77, with a minimum rate of 0.16 and a maximum rate of 0.93. There was no missing data. Developers note that data may have been rejected by the registry but no information regarding how often this might occur was provided.
- Since face validity was the only form of validity tested, no tests of significance conducted, and no missing data, I cannot comment.
- I agree with the NQF staff assessment of validity.
- No missing data - exclusion for pts enrolled in last 90 days of measurement period (to give time to assess). Comment: 90 days may be excessive as functional status should be assessed close to time of admission.

2b2-3. Other Threats to Validity

Comments:

- This is a process measure; risk adjustment is not necessary.
- Exemptions reasonable. Risk adjustment - NA
- No comment
- This measure is not risk adjusted, I agree with the rationale provided.
- not risk-adjusted, no missing data, few exceptions noted

Criterion 3. Feasibility

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

- The developer indicates that data for this measure are:
  - generated and collected by healthcare personnel during provision of health care,
  - coded by someone other than the initial individual collecting data, and
  - abstracted by someone other than the initial individual collecting data, for placement into the registry.
- Data from this measure come from defined fields included in the National Home-Based Primary Care & Palliative Care Registry.
- The developer reported no difficulties in collecting data for the measure or in implementing the measure.
- The measure is copyrighted. The developer allows use of the measure for noncommercial purposes but requires a license agreement for commercial use.

Questions for the Committee:

- Do you have any questions regarding the effort needed to collect data and report that data to the registry?
- Do you have any questions regarding costs associated with participation in the registry?
- Do you have any questions or concerns regarding the ability to implement this measure?

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. Feasibility

Comments:

- It seems that as part of a regular assessment tool, this is feasible.
- The developer reported no difficulties in collecting data for the measure or in implementing the measure.
- I do not know much about the registry but did look it up and the annual cost is $350. I would think that even a small agency could afford this yearly fee and that they would get a lot out of having access to the registry data. It also seems like functional assessment data would already be collected by those who serve elders - it is key information for this population group. I am not sure what role the copyright would play in regard to feasibility. Moderate
- I have no concerns about feasibility and the collection of data.
- defined field in National Home-Based Primary and palliative Care Registry

Criterion 4: Usability and Use

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

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

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

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

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

Accountability program details**:**

- The measure currently is used in the following:
  - MIPS QCDR (accountability program: payment)
  - American Board of Internal Medicine Approved Quality Improvement Activity (accountability program: certification/recognition)
  - National Home-based Primary care and Palliative Care Learning Collaborative (quality improvement program)
- This measure is not publicly reported at this time. However, CMS does plan to make all measures under the MIPS quality performance category available for public reporting on Physician Compare, as feasible.

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

Feedback on the measure by those being measured or others

- The developer notes that monthly reports of their measure results were shared with the 221 providers who submitted data to the registry.
- Feedback mechanisms for comments/questions include e-mail available through the National Home-based Primary Care and Palliative Care Consortium website; the MIPS QCDR portal (e.g., feasibility assessments), and the CMS annual QCDR self-nomination process.
- Feedback received to-date resulted in combining two measures of ADL/IADL assessments into the current measure and in specifying the ADLs/IADLs included in the measure.

Questions for the Committee:

- How has the measure been vetted in real-world settings by those being measured or others?
- Do you have any concerns regarding the ability to provide feedback on the measure?
- Do you have any concerns regarding the developers' use of feedback when considering modifications to the measure?

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 developers reported the following reporting and performance statistics:
    - 2016: 56.0%, with 303 providers reporting
    - 2017: 70.3%, with 271 providers reporting
    - 2018: 57.5%, with 246 providers reporting
- The developer provides some ideas about why reporting on the measure has varied during the 2016-2018 time period.

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

Unexpected findings (positive or negative) during implementation

- The developer noted no unexpected findings resulting from implementation of this measure.

Potential harms

- The developer did not indicate any potential harms or unexpected benefits from this measure.

Questions for the Committee:

- Do you have any concerns regarding the usability of the measure?
- Do the benefits of the measure outweigh any potential unintended consequences?

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

Committee Pre-evaluation Comments:
Criteria 4: Usability and Use

4a1. Use - Accountability and Transparency

Comments:

- The developers incorporated feedback and combined ADL/IADL measures into one to enhance use.
- This measure is not publicly reported at this time. However, CMS does plan to make all measures under the MIPS quality performance category available for public reporting on Physician Compare, as feasible. Providers get feedback on data. Measure changed to combination of ADL/IADL based on feedback.
- Pass
- The measure is not reported publicly but CMS plans to do make it available. Monthly reports are shared with providers who submitted data. Feedback can be submitted through various mechanism and has resulted in the combining of the ADL/IADL measures.
- feedback resulted in combining ADL/IADL assessments into measure

4b1. Usability – Improvement

Comments:

- I see benefits but no harms.
- Improvement was variable. Variability attributed to practice management changes. No reported benefits or harms.

- I wish the developer had a better idea regarding the variability of measurement reporting but since I cannot see any unintended consequences I would rate usability as moderate.
- The improvement results vary and not consistent during the 2016-2018 time frame. The benefits outweigh the harms with no unexpected findings noted.
- rationale provided, no unexpected findings or harms noted

Criterion 5: Related and Competing Measures

Related measures:

- **2524e**: Rheumatoid Arthritis: Patient-Reported Functional Status Assessment [*clinician-level measure used in outpatient setting; target population: adults with rheumatoid arthritis*]
- **2624**: Functional Outcome Assessment [*clinician-level measure (individual and group) used in outpatient setting; target population: adults with outpatient visit*]
- **2631**: Percent of Long-Term Care Hospital (LTCH) Patients With an Admission and Discharge Functional Assessment and a Care Plan That Addresses Function [*facility-level measure used in outpatient setting; target population: long-term care hospital patients*]

Harmonization

- NQF may ask the Committee to make recommendations for harmonizing measures.

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

5. Related and Competing Measures

Comments:

- I don't think that the three measures listed can be harmonized because each focus on a very situation-specific population --as does 3497--with distinct needs.
- Three other measures were identified but focus on different populations.
- I look forward to a discussion of how to harmonize the two measures. I would need to give thought to how this could be done.
- There are 3 related measures but none are competing.
- measure 2524e (Rheumatoid arthritis: patient reported functional status), measure 2624 (functional outcome assessment - measure of clinician outcomes for adult outpatient visits). Could consider harmonization for 2524e

Public and Member Comments

No NQF members have submitted support/non-support choices as of June 11, 2019.
No comments have been submitted as of June 11, 2019.

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

Functional_evidence_attachment_for_resubmission_FINAL.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.

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:**

Measure Title: Evaluation of Functional Status (Basic and Instrumental Activities of Daily Living [ADL]) for Home-Based Primary Care and Palliative Care Patients

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here:

Date of Submission: 4/16/2019

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*):

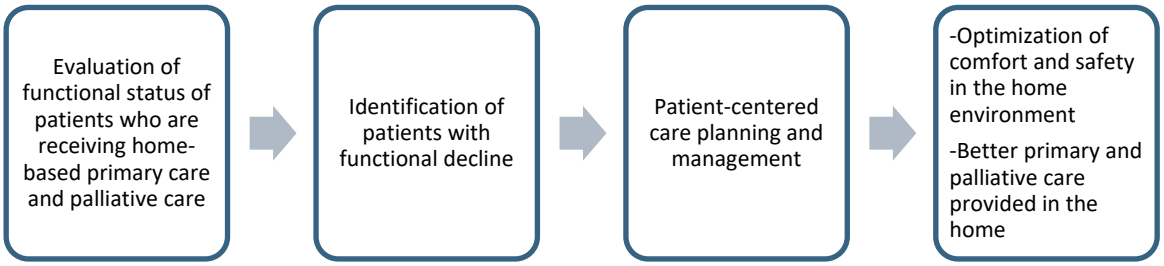
☒ Process: Rate of evaluation of functional status for home-based primary care and palliative care patients

☐ 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.



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.)

N/A

****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

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	• **Clinical Practice Guidelines for Quality Palliative Care, 4th edition** • **National Consensus Project for Quality Palliative Care** • **2018** • **National Consensus Project for Quality Palliative Care. Clinical Practice Guidelines for Quality Palliative Care, 4th edition. Richmond, VA: National Coalition for Hospice and Palliative Care; 2018. Page 1-31.** • https://www.nationalcoalitionhpc.org/wp-content/uploads/2018/10/NCHPC-NCPGuidelines_4thED_web_FINAL.pdf **Citation for systematic review:** • **Ahluwalia SC, Chen C, Raaen L, et al. A systematic review in support of the National Consensus Project Clinical Practice Guidelines for Quality Palliative Care, Fourth Edition. J Pain Symptom Manage. 2018;56(6):831-870. doi: 10.1016/j.jpainsymman.2018.09.008.**

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.	**Clinical Practice Guidelines for Quality Palliative Care, 4th edition** • **1.2.4** The initial assessment is conducted in person by one or more IDT (interdisciplinary team) members, depending on the needs and concerns of the patient, is documented, and includes: o **c. A physical examination including identification of current symptoms and functional status** • **2.2** The IDT assesses physical symptoms and their impact on well-being, quality of life, and functional status. o **2.2.3** The IDT utilizes validated symptom and functional assessment tools, treatment policies, standards, and guidelines appropriate to the care of neonates, children, adolescents, and adults with serious illnesses. o **2.2.4** The IDT conducts and regularly documents ongoing assessments of pain, other physical symptoms, **functional status**, symptom distress, and quality of life. After treatment is initiated, the IDT performs a timely reassessment to ascertain the effectiveness of the treatment. • **4.2.3** The social assessment includes: o **e. Functional limitations that impact activities of daily living (ADLs), instrumental activities of daily living (IADLs), and cognition**
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	The guidelines themselves do not reference graded evidence. However, a systematic review was performed which examined the quality of evidence in support of Key Questions (KQs). The systematic review authors assessed the quality of evidence using the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) framework. The GRADE framework allows for a transparent overview using internationally accepted criteria to differentiate high, moderate, low, and very low quality of evidence to describe confidence in the findings among studies. They downgraded for study limitations (e.g., no randomized controlled trials contributing to the evidence), inconsistency in results across studies or lack of replication, imprecision (e.g., due to lack of reported effect estimates or imprecise estimates). They used the assessment of the systematic reviews evaluating the evidence base regarding indirectness, publication bias, or other criteria, where applicable. (Ahluwalia et al, 2018) Quality of evidence was examined in support of Key Questions (KQs). (Ahluwalia et al, 2018) These KQs provided a conceptual framework which, along with expert opinion, informed guideline statements. The guideline statements themselves do not include an associated grade of evidence, but the KQs in support of them are supported by evidence graded as the following: • **H:** High quality of evidence • **M**: Moderate quality of evidence • **L:** Low quality of evidence • **V:** Very low quality of evidence
Provide all other grades and definitions from the evidence grading system	N/A
Grade assigned to the **recommendation** with definition of the grade	Grading system for recommendations not documented

Provide all other grades and definitions from the recommendation grading system	Grading system for recommendations not documented
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	The systematic review (Ahluwalia et al, 2018) referenced in these guidelines synthesizes best current evidence across eight domains. Ten key questions (KQs) were selected to guide the review in support of this guideline framework. Guideline statements used in this measure were developed based on the results of four KQs as well as expert opinion. **KQ1a)** ***What is the effect of interdisciplinary team care on patient and family/caregiver outcomes?*** Thirteen reviews met inclusion criteria for evaluating the effect of an interdisciplinary team on patient and family outcomes. There was moderate quality evidence for the impact of interdisciplinary teams on quality of life. Seven reviews reported on quality of life outcomes and consistently concluded that interdisciplinary care teams improved quality of life for patients with advanced illness. A 2017 review noted growing support for the utilization of palliative care teams for improving quality of life in advanced illness. Another review found an association between the number of disciplines included on the interdisciplinary care team and improved quality of life. Two reviews focused on cancer and heart failure patients also showed positive impacts. There was moderate-quality evidence for a positive impact of interdisciplinary teams on advance care planning (ACP) decisions, death at home, and patient and family satisfaction with care but low-quality evidence for a positive impact of interdisciplinary palliative care teams on patient physical symptoms and patient and family psychological outcomes. The quality of the evidence for the impact of teams on these outcomes was downgraded because of inconsistent conclusions and lack of pooled effect estimates. **KQ1b)** ***What is the impact of palliative care interventions to improve continuity and coordination of care on patient and family/caregiver outcomes?*** Eighteen reviews met inclusion criteria for evaluating the impact of palliative care interventions to improve continuity and coordination of care on patient and family/caregiver outcomes. Interventions included telehealth, early/integrated palliative care, home-based palliative care, case management, and care coordinators. Home-based care: There was high-quality evidence from two reviews for the impact of home-based palliative care on the likelihood of a patient dying at home. One review found that patients in home-based palliative care were more likely to die at home compared to usual care (odds ratio [OR] 2.21; CI 1.31-3.71; five RCTs and two controlled clinical trials). Another review found a similar significant association (RR 1.33; CI 1.14-1.55; three RCTs). However, there was low- to very low-quality evidence for home-based palliative care on other outcomes including physical symptoms, quality of life, and family/caregiver burden and satisfaction. ***KQ 2: What Is the Impact of Palliative Care Interventions on Physical Symptom Screening, Assessment, and Management of Patients?*** Forty-eight reviews met inclusion criteria

	for evaluating the impact of palliative care interventions on physical symptom screening, assessment, and management of patients. Much of the evidence in this domain is low quality largely due to inconsistent findings regarding the impact of interventions on symptoms. ***KQ 4: Does an Assessment of Environmental or Social Needs as Part of a Comprehensive Palliative Assessment Improve Needs Identification and Access to Relevant Services?*** Two reviews met inclusion criteria for evaluating an assessment of environmental or social needs as part of a comprehensive palliative assessment. There is low-quality evidence that social work interventions lead to the identification of and access to social services. One review found that various caregiver support services helped to identify caregivers' direct support service needs, decreased caregiver role overload, and led to a nonsignificant increase in social support and benefits finding; however, the review reported no effect estimate across studies.
Estimates of benefit and consistency across studies	Much of the evidence in the palliative care domain is low quality largely due to inconsistent findings regarding the impact of interventions on symptoms.
What harms were identified?	The guidelines do not describe any potential harms related to the guideline recommendations used to support this measure.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	No new studies have been identified that relate to functional assessment of the home-based palliative care patient.

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	• **Assessment of physical function in: Evidence-based geriatric nursing protocols for best practice** • **Hartford Institute for Geriatric Nursing** • **2003 (revised 2012)** • **Kresevic DM. Assessment of physical function. In: Boltz M, Capezuti E, Fulmer T, Zwicker D, editor(s). Evidence-based geriatric nursing protocols for best practice. 4th ed. New York (NY): Springer Publishing Company; 2012. p. 89-103.** • https://www.guidelinecentral.com/summaries/assessment-of-physical-function-in-evidence-based-geriatric-nursing-protocols-for-best-practice/#section-society

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.	• The clinician should document baseline functional status and recent or progressive declines in function (Graf, 2006 [Level V]). • Function should be assessed over time to validate capacity, decline, or progress (Applegate, Blass, & Franklin, 1990 [Level IV]; Callahan et al., 2002 [Level VI]; Kane & Kane, 2000 [Level VI]). • Standard instruments selected to assess function should be efficient to administer and easy to interpret. They should provide useful practical information for clinicians and should be incorporated into routine history taking and daily assessments (Kane & Kane, 2000 [Level VI]; Kresevic et al., 1998 [Level VI]) (see the "Availability of Companion Documents" field for tools).
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	• Level IV: Non-experimental studies • Level V: Care report/program evaluation/narrative literature reviews • Level VI: Opinions of respected authorities/consensus panels
Provide all other grades and definitions from the evidence grading system	• Level I: Systematic reviews (integrative/meta-analyses/clinical practice guidelines based on systematic reviews) • Level II: Single experimental study (randomized controlled trials [RCTs]) • Level III: Quasi-experimental studies
Grade assigned to the **recommendation** with definition of the grade	Grading system for recommendations not documented
Provide all other grades and definitions from the recommendation grading system	Grading system for recommendations not documented
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	The evidence cited in support of the guideline recommendations is comprised of 5 studies. 2 studies in support of these guideline recommendations are classified as opinions of respected authorities and consensus panels. 1 studies in support of these guideline recommendations is classified as a non-experimental study, and 1 study in support of these guideline recommendations is classified as a care report/program evaluation/narrative literature review. Taken together, these guideline statements and their supporting evidence support the practice of evaluating functional status for home-based primary and palliative care patients to improve care provided in the home.

Estimates of benefit and consistency across studies	The guideline recommendations list several potential benefits of functional status evaluations for home-based primary and palliative care patients, including attainment of highest quality of life despite functional level, the provision of necessary adaptations to maintain safety and independence in the home, and reduced incidence and prevalence of functional decline. These benefits manifest at patient, provider, and health care system levels. Estimates of benefit and consistency across studies supporting the guideline statements in support of functional assessments were not directly addressed, though relevant studies were noted to include consensus statements, narrative literature reviews, and non-experimental studies.
What harms were identified?	No potential or realized harms were listed in these guidelines.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	No new studies have been identified that relate to functional assessment of the home-based palliative care patient.

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	• **Age-related changes in health In: Evidence-based geriatric nursing protocols for best practice** • **Hartford Institute for Geriatric Nursing** • **2008 (revised 2012)** • **Smith CM and Cotter VT. Nursing Standard of Practice Protocol: Age-related Changes in Health. In: Boltz M, Capezuti E, Fulmer T, Zwicker D, editor(s). Evidence-based geriatric nursing protocols for best practice. 4th ed. New York (NY): Springer Publishing Company; 2012. p. 23-47.** https://www.guidelinecentral.com/share/summary/52d56124ed39c#section-society
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.	• Assess, with periodic reassessment, baseline functional status (Craft, Cholerton, & Reger, 2009 [Level V])
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	• Level V: Care report/program evaluation/narrative literature reviews
Provide all other grades and definitions from the evidence grading system	• Level I: Systematic reviews (integrative/meta-analyses/clinical practice guidelines based on systematic reviews) • Level II: Single experimental study (randomized controlled trials [RCTs]) • Level III: Quasi-experimental studies • Level IV: Non-experimental studies • Level VI: Opinions of respected authorities/consensus panels
Grade assigned to the **recommendation** with definition of the grade	Grading system for recommendations not documented

Provide all other grades and definitions from the recommendation grading system	Grading system for recommendations not documented
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	The evidence cited in support of the guideline recommendations is comprised of 1 study. 1 study in support of these guideline recommendations is classified as a care report/program evaluation/narrative literature review.
Estimates of benefit and consistency across studies	The guideline recommendations list several potential benefits of functional status evaluations for home-based primary and palliative care patients, including successful aging of older adult through appropriate lifestyle practices and health care, the identification of normative changes in aging and differentiation from pathological processes, and the development of interventions to correct for adverse effects. These benefits manifest at patient, provider, and health care system levels.
What harms were identified?	No potential or realized harms were listed in these guidelines.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	No new studies have been identified that relate to functional assessment of the home-based primary care or palliative care patient.

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.

There are no systematic reviews of assessment of functional status or cognitive status for people receiving home-based primary care or palliative care, per se. However, there is a robust literature and evidence base (randomized clinical trials, meta-analyses, and systematic reviews) regarding the value of comprehensive geriatric assessment (CGA) for older adults in varied settings and clinical situations. **CGA always includes assessment of functional status and cognitive assessment.**

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

Two meta-analyses focused specifically on community-dwelling older adults in the context of home-based care found reductions in mortality and improvement in functional status. [1,2] In addition, there have been multiple meta-analyses and systematic reviews of CGA in other settings where patients receiving home-based primary care and palliative care commonly receive care, including hospital, emergency department, long-term care, post-hospital discharge, outpatient consultation, and inpatient CGA consultation.[3,4,5,6,7]

Given the clinical, social, and health service utilization characteristics of homebound older adults and those homebound older adults who receive home-based primary care, this entire evidence base is relevant to our proposed measures.

We note the following systematic reviews and meta-analyses of CGA, the setting, and the relevant outcomes in this table.

Lead Author, Year Published	Paper Title	# of Studies Included	Main Results
Elkan 2001	Effectiveness of home based support for older people: systematic review and meta-analysis.	8	The pooled odds ratio for eight studies that assessed mortality in members of the general elderly population was 0.76 (95% confidence interval 0.64 to 0.89). Five studies of home visiting to frail older people who were at risk of adverse outcomes also showed a significant reduction in mortality (0.72; 0.54 to 0.97). Home visiting was associated with a significant reduction in admissions to long term institutional care in members of the general elderly population (0.65; 0.46 to 0.91). For three studies of home visiting to frail, "at risk" older people, the pooled odds ratio was 0.55 (0.35 to 0.88). Meta-analysis of six studies of home visiting to members of the general elderly population showed no significant reduction in admissions to hospital (odds ratio 0.95; 0.80 to 1.09). Three studies showed no significant effect on health (standardised effect size 0.06; -0.07 to 0.18). Four studies showed no effect on activities of daily living (0.05; -0.07 to 0.17).
Huss 2009	Multidimensional preventive home visit programs for community-dwelling older adults: a systematic review and meta-analysis of randomized controlled trials.	21	A beneficial effect on mortality was seen in younger study populations (OR 0.74, 95% CI, 0.58-0.94) but not in older populations (OR 1.14, 95% CI, 0.90-1.43). Functional decline was reduced in programs including a clinical examination in the initial assessment (OR 0.64, 95% CI, 0.48-0.87) but not in other trials (OR 1.00, 95% CI, 0.88-1.14). There was no single factor explaining the heterogenous effects of trials on nursing home admissions.
Panza, 2018	An Old Challenge with New Promises: A Systematic Review on Comprehensive Geriatric Assessment in Long-Term Care Facilities.	15	the present reviewed evidence suggested that most complex older subjects may benefit from a CGA in terms of improved quality of care and reduced hospitalization events
Jay 2017	Can consultant geriatrician led comprehensive geriatric assessment in the emergency department reduce hospital admission rates? A systematic review.	5	All of the studies reported statistically significant reductions in admission rates (ranging between 2.6 and 19.7%).

Lead Author, Year Published	Paper Title	# of Studies Included	Main Results
Ellis G, 2017	Comprehensive geriatric assessment for older adults admitted to hospital.	29	CGA increases the likelihood that patients will be alive and in their own homes at 3 to 12 months' follow-up (risk ratio (RR) 1.06, 95% confidence interval (CI) 1.01 to 1.10; 16 trials, 6799 participants; high-certainty evidence), results in little or no difference in mortality at 3 to 12 months' follow-up (RR 1.00, 95% CI 0.93 to 1.07; 21 trials, 10,023 participants; high-certainty evidence), decreases the likelihood that patients will be admitted to a nursing home at 3 to 12 months follow-up (RR 0.80, 95% CI 0.72 to 0.89; 14 trials, 6285 participants; high-certainty evidence) and results in little or no difference in dependence (RR 0.97, 95% CI 0.89 to 1.04; 14 trials, 6551 participants; high-certainty evidence). CGA may make little or no difference to cognitive function (SMD ranged from -0.22 to 0.35 (5 trials, 3534 participants; low-certainty evidence)). Mean length of stay ranged from 1.63 days to 40.7 days in the intervention group, and ranged from 1.8 days to 42.8 days in the comparison group. Healthcare costs per participant in the CGA group were on average GBP 234 (95% CI GBP - 144 to GBP 605) higher than in the usual care group (17 trials, 5303 participants; low-certainty evidence). CGA may lead to a slight increase in QALYs of 0.012 (95% CI -0.024 to 0.048) at GBP 19,802 per QALY gained (3 trials; low-certainty evidence), a slight increase in LYs of 0.037 (95% CI 0.001 to 0.073), at GBP 6305 per LY gained (4 trials; low-certainty evidence), and a slight increase in LYLAH of 0.019 (95% CI -0.019 to 0.155) at GBP 12,568 per LYLAH gained (2 trials; low-certainty evidence). The probability that CGA would be cost-effective at a GBP 20,000 ceiling ratio for QALY, LY, and LYLAH was 0.50, 0.89, and 0.47, respectively (17 trials, 5303 participants; low-certainty evidence).
Conroy, 2011	A systematic review of comprehensive geriatric assessment to improve outcomes for frail older people being rapidly discharged from acute hospital: 'interface geriatrics'	5	There was no clear evidence of benefit for CGA interventions in this population in terms of mortality [RR 0.92 (95% CI 0.55–1.52)] or readmissions [RR 0.95 (95% CI 0.83–1.08)] or for subsequent institutionalisation, functional ability, quality-of-life or cognition.

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

PubMed search of literature relevant to Comprehensive Geriatric Assessment.

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

1 Elkan R, Kendrick D, Dewey M, Hewitt M, Robinson J, Blair M, Williams D, Brummell K. Effectiveness of home based support for older people: systematic review and meta-analysis. BMJ. 2001 Sep 29;323(7315):719-25. Review. PubMed PMID:11576978; PubMed Central PMCID: PMC56889.

2 Huss A, Stuck AE, Rubenstein LZ, Egger M, Clough-Gorr KM. Multidimensional preventive home visit programs for community-dwelling older adults: a systematic review and meta-analysis of randomized controlled trials. J Gerontol A Biol Sci Med Sci. 2008 Mar;63(3):298-307. Review. Erratum in: J Gerontol A Biol Sci Med Sci. 2009 Feb;64(2):318. PubMed PMID: 18375879.

3 Pilotto A, Cella A, Pilotto A, Daragjati J, Veronese N, Musacchio C, Mello AM, Logroscino G, Padovani A, Prete C, Panza F. Three Decades of Comprehensive Geriatric Assessment: Evidence Coming From Different Healthcare Settings and Specific Clinical Conditions. J Am Med Dir Assoc. 2017 Feb 1;18(2):192.e1-192.e11. doi: 10.1016/j.jamda.2016.11.004. Epub 2016 Dec 31. Review. PubMed PMID: 28049616.

4 Panza F, Solfrizzi V, Lozupone M, Barulli MR, D'Urso F, Stallone R, Dibello V, Noia A, Di Dio C, Daniele A, Bellomo A, Seripa D, Greco A, Logroscino G. An Old Challenge with New Promises: A Systematic Review on Comprehensive Geriatric Assessment in Long-Term Care Facilities. Rejuvenation Res. 2018 Feb;21(1):3-14. doi: 10.1089/rej.2017.1964. Epub 2017 Jul 28. Review. PubMed PMID: 28635539.

5 Jay S, Whittaker P, Mcintosh J, Hadden N. Can consultant geriatrician led comprehensive geriatric assessment in the emergency department reduce hospital admission rates? A systematic review. Age Ageing. 2017 May 1;46(3):366-372. doi: 10.1093/ageing/afw231. Review. PubMed PMID: 27940568.

6 Conroy SP, Stevens T, Parker SG, Gladman JR. A systematic review of comprehensive geriatric assessment to improve outcomes for frail older people being rapidly discharged from acute hospital: 'interface geriatrics'. Age Ageing. 2011 Jul;40(4):436-43. doi: 10.1093/ageing/afr060. Epub 2011 May 26. Review.PubMed PMID: 21616954.

7 Ellis G, Gardner M, Tsiachristas A, Langhorne P, Burke O, Harwood RH, Conroy SP, Kircher T, Somme D, Saltvedt I, Wald H, O'Neill D, Robinson D, Shepperd S. Comprehensive geriatric assessment for older adults admitted to hospital. Cochrane Database Syst Rev. 2017 Sep 12;9:CD006211. doi:10.1002/14651858.CD006211.pub3. Review. PubMed PMID: 28898390; PubMed Central PMCID: PMC6484374.

1b. Performance Gap

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

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

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.

Functional limitations, frailty, and homebound status render approximately 4 million US adults unable to easily access office-based primary care (1). Compared to the non-homebound population, homebound individuals have twice as many chronic conditions, are more likely to have depression or dementia, are nearly two to two-and-a-half times more likely to have been hospitalized in the past year, and are among the most costly and sickest patients in the US healthcare system (2-4). Further, one study of Medicare beneficiaries found that 50.7% are homebound in the last year of life, suggesting that half are unable to receive office-based care at a time when they may have the most need (5). To address these care access and cost issues, home-based primary and palliative care is a model growing with the support of providers and policymakers alike (2). Between 2000 and 2006, the number of physician house calls to Medicare beneficiaries more than doubled while the number of providers practicing home care has decreased (6). This suggests a growing emphasis on home-based primary and palliative care as a focused medical specialty, with dedicated providers seeing a greater percentage of their patient panels in the patients' homes than before (3,6). Seizing on the opportunities inherent in this emerging practice model, the Centers for Medicare and Medicaid Services (CMS)

has invested millions of dollars into the Independence at Home (IAH) Demonstration, which tests a payment incentive and service delivery model for home based primary care (2). Outside of this demonstration project, however, there are no currently endorsed performance measures specific to home-based primary care, leaving a growing, high priority practice model with disease-specific measures that do not take into account the clinical complexity and multi-morbidity of the homebound population (3).

Poor functional status is a major contributor to costs for the homebound, with nearly half of patients in the IAH Demonstration reporting the need for assistance with four or more Activities of Daily Living (ADLs) (2). The high cost of care associated with multimorbidity in chronic illness is largely driven by functional limitations, not the presence of multiple chronic conditions alone. Medicare enrollees with functional limitations and chronic conditions account for one third of total Medicare spending—this is twice the average amount for Medicare beneficiaries with three or more chronic conditions but without functional limitations (7). Hospital spending is the largest source of this gap, with Medicare enrollees with both chronic conditions and functional limitations using the Emergency Department and admitted to the hospital more than enrollees with multiple chronic conditions and no functional limitations (7). Home-based primary and palliative care providers have an opportunity to play a role in preventing these costly admissions by engaging homebound patients in functional assessments. Identifying functional needs allows for both treatment and linkage with resources to prevent falls and enhance in-home safety. Although there are no dedicated home-based primary care guidelines based on systematic review, functional assessments are supported in palliative and geriatric nursing guidelines (3,8,9). Despite these recommendations, current provider performance demonstrate an opportunity for improvement. 2017-2018 registry data for this measure, based on 221 providers and 64,394 quality events indicate an average performance rate of 67%, with a range from 16% to 93%. This variability and lower rate of performance indicate that home-based primary and palliative care providers have room to improve on functional assessment rates in care of the homebound population. Achieving high performance should ultimately lead to enhanced patient-centered care in the home and improved outcomes for this complex population and is a crucial first step in achieving many high priority quality of care standards identified by The National Home-Based Primary and Palliative Care Network (3).

1. Qiu WQ, Dean M, Liu T, et al. Physical and mental health of homebound older adults: an overlooked population. J Am Geriatr Soc. 2010;58(12):2423-2428.

2. Report to Congress: Evaluation of the Independence at Home Demonstration. Centers for Medicare and Medicaid Services;2018.

3. Leff B, Carlson CM, Saliba D, Ritchie C. The invisible homebound: setting quality-of-care standards for home-based primary and palliative care. Health Aff (Millwood). 2015;34(1):21-29.

4. Ornstein KA, Leff B, Covinsky KE, et al. Epidemiology of the Homebound Population in the United States. JAMA internal medicine. 2015;175(7):1180-1186.

5. Soones T, Federman A, Leff B, Siu AL, Ornstein K. Two-Year Mortality in Homebound Older Adults: An Analysis of the National Health and Aging Trends Study. J Am Geriatr Soc. 2017;65(1):123-129.

6. Peterson LE, Landers SH, Bazemore A. Trends in physician house calls to Medicare beneficiaries. J Am Board Fam Med. 2012;25(6):862-868.

7. Komisar HL, Feder J. Transforming Care for Medicare Beneficiaries with Chronic Conditions and Long-Term Care Needs: Coordinating Care Across All Services. Georgetown University;2011.

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

9. Kresevic DM. Assessment of physical function. In: Boltz M, Capezuti E, Fulmer T, Zwicker D, editor(s). Evidence-based geriatric nursing protocols for best practice. 4th ed. New York (NY): Springer Publishing Company; 2012. p. 89-103.

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.*

Mean: 0.67

Standard deviation: 0.15

Minimum: 0.16

Maximum: 0.93

Interquartile range: 0.20 (0.77-0.57)

Data source: 2017-2018 Registry data from the National Home-Based Primary Care & Palliative Care Registry

Number of measured entities: 221 providers; 64,394 quality events (patients)

Dates of data: 11/1/2017-10/31/2018

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.

There was not literature identified that specifically addressed the rate of completion of a functional evaluation for home based primary care and palliative care patients. However, the testing data for this measure do indicate an opportunity for improvement.

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.*

Disparities in numerator performance are not available for the measure as specified.

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

It is well established that racial and ethnic minority status is associated with higher risk for adverse health and functional outcomes.1 Patients of racial and ethnic minorities often comprise a significant proportion of those receiving home-based primary and palliative care.2,3 This research supports disparities in functional outcomes and supports the need for functional assessment of patients receiving home-based primary and palliative care. However, we are not aware of any evidence specifically related to disparities in the assessment of functional status among patients receiving home-based primary and palliative care.

1Ng JH, Bierman AS, Elliott MN, Wilson RL, Xia C, Scholle SH. Beyond black and white: Race/ethnicity and health status among older adults. Am J Manag Care. 2014; 20(3):239-248.

2Report to Congress: Evaluation of the Independence at Home Demonstration. Centers for Medicare and Medicaid Services;2018.

3Klein S, Hostetter M, McCarthy D. An overview of home-based primary care: Learning from the field. Issue Brief (Commonw Fund.) 2017; 15:1-20.

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):*

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

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.)*

The measure specifications are included in this submission. Additional measure details may be found at https://www.medconcert.com/content/medconcert/NHBCPCR/2019_National_Home-Based_Primary_Care_&_Palliative_Care_Measure_Specific.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.

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).

Submission Criteria 1 - Newly enrolled:

Number of newly enrolled home-based primary care and palliative care patients who were assessed for basic ADL and IADL impairment at enrollment.

Submission Criteria 2 - Established patients:

Number of established home-based primary care and palliative care patients who were assessed for ADL and IADL impairment at enrollment and annually

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).

Time Period for Data Collection: At least once during the measurement period

GUIDANCE:

Basic ADLs must include but are not limited to: bathing, transferring, toileting, and feeding; IADL must include but are not limited to: telephone use and managing own medications.

Submission Criteria 1 - Newly enrolled:

Report NHBPC15.NUMER.1.YES - Basic ADL and IADL assessment performed and documented within 90 days of New Patient Encounter

Submission Criteria 2 - Established patients:

Report NHBPC15.NUMER.3.YES - ADL and IADL assessment performed and documented within performance period

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

Submission Criteria 1 - Newly enrolled:

Total number of newly enrolled (and active) home-based primary care and palliative care patients. The enrollment period includes 90 days from the first recorded new patient E&M visit code with the practice. *A patient is considered active if they have at least 2 E&M visit codes with a provider from the practice within the reporting period.

Submission Criteria 2 - Established patients:

Total number of established enrolled and active home-based primary care and palliative care patients. A patient is considered established if they have at least 2 Established Patient Encounter E&M visit codes with a provider from the practice within the reporting period.

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).

Time Period for Data Collection: 12 consecutive months

Submission Criteria 1 - Newly enrolled:

New/Established Patient Encounter during the performance period (CPT): 99324, 99325, 99326, 99327, 99328, 99341, 99342, 99343, 99344, 99345

AND

At least one subsequent Established Patient Encounter during the performance period (CPT): 99334, 99335, 99336, 99337, 99347, 99348, 99349, 99350, 99497

Submission Criteria 2 - Established patients:

At least two instances of Established Patient Encounter (CPT): 99334, 99335, 99336, 99337, 99347, 99348, 99349, 99350, 99497

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

Submission Criteria 1 - Newly enrolled:

Denominator Exceptions:

Most recent new patient encounter (with subsequent established patient encounter) occurs within the last 90 days of the measurement period

Submission Criteria 2 - Established patients:

There are no exceptions or exclusions for this submission criteria.

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.)*

Time Period for Data Collection: During the measurement period.

Submission Criteria 1 - Newly enrolled:

Exceptions are used to remove a patient from the denominator of a performance measure when the patient does not receive a therapy or service AND that therapy or service would not be appropriate due to patient-specific reasons. The patient would otherwise meet the denominator criteria. Exceptions are not absolute, and are based on clinical judgment, individual patient characteristics, or patient preferences. This measure has been developed using the PCPI exception methodology, which uses three categories of reasons for which a patient may be removed from the denominator of an individual measure. These measure exception categories are not uniformly relevant across all measures; for each measure, there must be a clear rationale to permit an exception for a medical, patient, or system reason. For measure Evaluation of Functional Status (Basic and Instrumental Activities of Daily Living [ADL]) for Home-Based Primary Care and Palliative Care Patients, exceptions may include most recent new patient encounter (with subsequent established patient encounter) occurs within the last 90 days of the measurement period. Although this methodology does not require the external reporting of more detailed exception data, it is recommended that physicians document the specific reasons for exception in patients' medical records for purposes of optimal patient management and audit-readiness. The UCSF, JHU School of Medicine, and the PCPI also advocate the systematic review and analysis of each physician's exceptions data to identify practice patterns and opportunities for quality improvement.

Exception is determined by date(s) of encounter(s).

Submission Criteria 2 - Established patients:

Not applicable.

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.)*

Consistent with CMS' Measures Management System Blueprint and national recommendations put forth by the IOM (now the National Academies) and NQF, the University of California San Francisco and Johns Hopkins University School of Medicine encourage collection of race and ethnicity data as well as the results of this measure to be stratified by race, ethnicity, administrative sex, and payer.

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.)*

This measure is comprised of two populations but is intended to result in one reporting rate. The reporting rate is the aggregate of Submission Criteria 1 (Newly enrolled) and Submission Criteria 2 (Established patients), resulting in a single performance rate. For the purposes of this measure, the single performance rate can be calculated as follows:

Performance Rate = (Numerator 1 + Numerator 2)/ [(Denominator 1 – Denominator Exceptions 1) + (Denominator 2)]

To calculate performance rates for Submission Criteria 1 - Newly enrolled:

1. Find the patients who meet the initial population (ie, the general group of patients that a set of performance measures is designed to address).

2. From the patients within the initial population criteria, find the patients who qualify for the denominator (ie, the specific group of patients for inclusion in a specific performance measure based on defined criteria). Note: in some cases the initial population and denominator are identical.

3. From the patients within the denominator, find the patients who meet the numerator criteria (ie, the group of patients in the denominator for whom a process or outcome of care occurs). Validate that the number of patients in the numerator is less than or equal to the number of patients in the denominator.

4. From the patients who did not meet the numerator criteria, determine if the provider has documented that the patient meets any criteria for exception when denominator exceptions have been specified [for this measure: most recent new patient encounter (with subsequent established patient encounter) occurs within the last 90 days of the measurement period]. If the patient meets any exception criteria, they should be removed from the denominator for performance calculation. --Although the exception cases are removed from the denominator population for the performance calculation, the exception rate (ie, percentage with valid exceptions) should be calculated and reported along with performance rates to track variations in care and highlight possible areas of focus for QI.

If the patient does not meet the numerator and a valid exception is not present, this case represents a quality failure.

To calculate performance rates for Submission Criteria 2 - Established patients:

1. Find the patients who meet the initial population (ie, the general group of patients that a set of performance measures is designed to address).

2. From the patients within the initial population criteria, find the patients who qualify for the denominator (ie, the specific group of patients for inclusion in a specific performance measure based on defined criteria). Note: in some cases the initial population and denominator are identical.

3. From the patients within the denominator, find the patients who meet the numerator criteria (ie, the group of patients in the denominator for whom a process or outcome of care occurs). Validate that the number of patients in the numerator is less than or equal to the number of patients in the denominator.

If the patient does not meet the numerator, this case represents a quality failure.

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.

Not applicable. The measure is not based on a sample.

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.

Not applicable. The measure is not based on a survey.

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

If other, please describe in S.18.

Registry Data

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.

The data source is the National Home-Based Primary Care & Palliative Care Registry.

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 : Individual

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

Home Care, Other

If other: Home-based primary care and home-based palliative care; Settings include: Home, Boarding home, Domiciliary, Assisted Living Facilities, Rest Home or Custodial Care Services

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.)*

Not applicable. The measure is not a composite.

2. Validity – See attached Measure Testing Submission Form

v5_Evaluation_of_Functional_Status_nqf_testing_attachment_7.1.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.

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.

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.

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

Measure Number (*if previously endorsed*)**:** N/A
Measure Title: Evaluation of Functional Status (Basic and Instrumental Activities of Daily Living [ADL]) for Home-Based Primary Care and Palliative Care Patients
Date of Submission: 5/6/2019
Type of Measure:

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

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*).

The data source is 2017-2018 Registry data from the National Home-Based Primary Care & Palliative Care Registry.

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

The data are for the time period November 1st, 2017 through October 31st, 2018.

1.4. What levels of analysis were tested? (*testing must be provided for <u>all</u> 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 <u>measured entities</u> 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*)

We received data from 221 providers reporting on this measure through the registry for the National Home-Based Primary Care & Palliative Care Registry during the period between 11/1/2017-10/31/2018. This data set reflects individual provider data and our analysis of the data as a whole is reflected throughout this submission. Of those, 221 providers had all the required data elements and at least one quality reporting event for a total of 64,394 quality events. A quality reporting event is an event that qualifies for the denominator and is not an exclusion or an exception. In this case, a quality reporting event is a patient. For this measure, 100 percent of providers are included in the analysis, and the average number of quality reporting events are 291 for the 221 providers. The range of quality reporting events for 221 providers included is from 3 to 549.

1.6. How many and which <u>patients</u> 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*)

There were 64,394 quality events included in this reliability testing and analysis. These were the quality events that were associated with providers who had all the required data elements and at least one quality reporting event.

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.

The same data samples were used for reliability testing and exceptions analysis.

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.

In the tables within this section the average performance percentage for each category is provided:

Demographic Data Current (1 year)	Age							
	0 - 17	18 - 24	25 - 34	35 - 44	45 - 54	55 - 64	65 - 74	75+
Overall Summary	-	40.15%	55.41%	64.17%	66.12%	65.76%	71.62%	68.38%

Demographic Data Current (1 year)	Ethnicity		
	Hispanic or Latino	Not Hispanic or Latino	Unknown
Overall Summary	62.27%	68.57%	65.92%

Demographic Data Current (1 year)	Race							
	Black or African American	White	Asian	Native Hawaiian or Other Pacific Islander	American Indian or Alaska Native	Declined	Unknown	Not Provided
Overall Summary	68.17%	68.09%	68.71%	66.94%	66.43%	63.57%	65.34%	60.34%

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*)

Reliability of the computed measure score was measured as the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in provider performance and the noise is the total variability in measured performance. Reliability at the level of the specific provider is given by:

Reliability = Variance (provider-to-provider) / [Variance (provider-to-provider) + Variance (provider-specific-error]

Reliability is the ratio of the provider-to-provider variance divided by the sum of the provider-to-provider variance plus the error variance specific to a provider. A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to real differences in provider performance.

Variance (provider-to-provider) = p(1-p)/n where p is the passing rate for a provider and n is the number of qualifying reporting events for that provider

Variance (provider-specific-error) = alpha*beta/((alpha + beta + 1)*(alpha + beta)^2)

Reliability testing was performed by using a beta-binomial model. The beta-binomial model assumes the provider performance score is a binomial random variable conditional on the provider's true value that comes from the beta distribution. The beta distribution is usually defined by two parameters, alpha and beta. Alpha and beta can be thought of as intermediate calculations to get to the needed variance estimates.

For this analysis Alpha = 6.6095 and Beta = 3.3067. These parameters are used to calculate the variance (provider-specific-error) which is approximately equal to 0.02. Reliability is then calculated for each provider using this value and the variance (provider-to-provider). Average reliability is reported by averaging reliability for each provider with at least 1 quality reporting event for the measure.

A reliability equal to zero implies that all the variability in a measure is attributable to measurement error. A reliability equal to one implies that all the variability is attributable to real differences in provider performance. A reliability of 0.70 – 0.80 is generally considered the acceptable threshold for reliability, 0.80 – 0.90 is considered high reliability, and 0.90 – 1.0 is considered very high. [1]

1. Adams JL, Mehrotra A, McGlynn EA, Estimating Reliability and Misclassification in Physician Profiling, Santa Monica, CA: RAND Corporation, 2010. www.rand.org/pubs/technical_reports/TR863. (Accessed on February 24, 2012.)

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*)

The average reliability for providers with at least one quality reporting event is 0.95. We also evaluated the reliability at each decile.

Table 1: Average Reliability Results

Decile	Reliability
1	0.94
2	0.96
3	0.96
4	0.96
5	0.97
6	0.97
7	0.97
8	0.98
9	0.98
10	0.99

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?*)

This measure has very high reliability when including providers with at least one quality reporting event.

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)*

Face Validity:

Input on the content validity of draft measures is obtained through soliciting comments from a panel of experts who were not involved in the development of the measure.

Face validity of the measure score as an indicator of quality was systematically assessed as follows:

After the measure was fully specified, the expert 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". The expert panel included 12 members. Panel members were comprised of experts from family medicine, geriatric medicine, palliative medicine, internal medicine, and from home health executive roles.

The scores obtained from the measure as specified will provide an accurate reflection of quality and can be used to distinguish good and poor quality.

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

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

Face Validity

Our expert panel included 12 members. The list of expert panel members is as follows:

1. Eric De Jonge, MD
2. Jen Hayashi, MD
3. Linda DeCherrie, MD
4. Steven Landers, MD
5. Mattan Schuchman, MD
6. Carla Perissinotto, MD
7. Thomas Cornwell, MD
8. Sharon Levine, MD
9. Carlos Weiss, MD
10. Mia Yang, MD
11. Eliza Shulman, DO, MPH
12. Alison Aubry, MSN, RN

Frequency Distribution of Ratings

1 (Strongly Disagree) - 0 responses (0%)

2 (Disagree) - 1 response (8.3%)

3 (Neither Agree nor Disagree) - 0 responses (0%)

4 (Agree) – 3 responses (25.0%)

5 (Strongly Agree) - 8 responses (66.7%)

The summary of the expert panel ratings of the validity statement were as follows:

N = 12; Mean rating = 4.50 with 91.7% of respondents with either an 'agree' or a 'strongly agree' response that this measure can accurately distinguish good and poor quality.

(a) Of 11 persons with either an 'agree' or 'strongly agree' response, 3 provided comments. One commenter that agreed with the measure noted that electronic records must adapt to include a data field in order to document the completion of these assessments. Another commented that they do screen function annually

and that it helps the provider on call and any inpatient team to understand the patient's baseline. A third commented that they do not currently collect information on ADLs/IADLs, but it does not mean that they provide lower quality of care.

(b) The 1 person with a 'disagree' response, did not provide comments.

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?*)

Based on the mean rating by the expert panel considered with the comments, overall this measure is valid as specified.

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*)

Exceptions include:

- Most recent new patient encounter (with subsequent established patient encounter) occurs within the last 90 days of the measurement period

Exceptions were analyzed for frequency across providers and deciles of exceptions were reported.

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*)

Amongst the 220 providers with the minimum (10) number of quality reporting events, there were a total of 807 exceptions reported. The average number of exceptions per provider in this sample is 3.67. The proportion of exceptions to patients is 0.01.

Amongst the 221 included providers, there were a total of 807 exceptions reported. The average number of exceptions per provider in this sample is 3.65. The proportion of exceptions to quality events is 0.01. Exception deciles illustrate the spread of exceptions amongst providers. According to the results, 50% of providers had 2 or fewer exceptions across eligible patients for the year under study.

Decile	Exceptions
1	0
2	0
3	1
4	2
5	2
6	3
7	5
8	6
9	9
10	21

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*)

Exceptions are necessary to account for those situations when the most recent new patient encounter (with subsequent established patient encounter) occurs within the last 90 days of the measurement period. Exceptions are discretionary and the methodology used for measure exception categories are not uniformly relevant across all measures; for this measure, there is a clear rationale to permit an exception. New patient encounters between 10/1 and 12/31 of the measurement period will not have the full 90 day enrollment period to complete the measure criteria and will be less likely to pass the measure.

Some have indicated concerns with exception reporting including the potential for physicians to inappropriately exclude patients to enhance their performance statistics. Research has indicated that levels of exception reporting occur infrequently and are generally valid (Doran et al., 2008), (Kmetik et al., 2011). Furthermore, exception reporting has been found to have substantial benefits: "it is precise, it increases acceptance of [pay for performance] programs by physicians, and it ameliorates perverse incentives to refuse care to "difficult" patients." (Doran et al., 2008).

Although this methodology does not require the external reporting of more detailed exception data, the measure developer recommends that physicians document the specific reasons for exception in patients' medical records for purposes of optimal patient management and audit-readiness. We also advocate for the systematic review and analysis of each physician's exceptions data to identify practice patterns and opportunities for quality improvement.

Without exceptions, the performance rate would not accurately reflect the true performance of that physician. This would result in an increase in performance failures and false negatives. The additional value of increased data collection of capturing an exception greatly outweighs the reporting burden.

References:

Doran T, Fullwood C, Reeves D, Gravelle H, Roland M. Exclusion of pay for performance targets by English Physicians. New Engl J Med. 2008; 359: 274-84.

Kmetik KS, Otoole MF, Bossley H et al. Exceptions to Outpatient Quality Measures for Coronary Artery Disease in Electronic Health Records. Ann Intern Med. 2011;154:227-234.

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.

Not applicable

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.

Not applicable

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?

Not applicable

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?

Not applicable

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.**

Not applicable

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*)

Not applicable

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*)**:**

Not applicable

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

Not applicable

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

Not applicable

2b3.9. Results of Risk Stratification Analysis:

Not applicable

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*)

Not applicable

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*)

Not applicable

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)*

Measures of central tendency, variability, and dispersion were calculated.

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*)

Based on the sample of 221 included providers, the mean performance rate is 0.67, the median performance rate is 0.71 and the mode is 0.67. The standard deviation is 0.15. The range of the performance rate is 0.77, with a minimum rate of 0.16 and a maximum rate of 0.93. The interquartile range is 0.20 (0.77–0.57). Deciles are provided below:

Decile	Performance
1	0.46
2	0.54
3	0.61
4	0.66
5	0.71
6	0.74
7	0.76
8	0.80
9	0.83
10	0.93

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?*)

The range of performance from 0.16 to 0.93 suggests there's clinically meaningful variation across providers' performance. Outliers are considered to be values less than quartile 1 (0.57) or greater than quartile 3 (0.77) by more than 1.5 the IQR (0.20) and there were approximately 4 outliers in the data set.

Quartile	Performance
1	0.57
2	0.71
3	0.77
4	0.93

Excluding those outliers, the range of performance is 0.27 to 0.93 with a mean of 0.68, median of 0.71, and standard deviation of 0.14. Looking at the performance percentiles without outliers, 50% of the data falls at or below a performance score of 0.71 which demonstrates additional meaningful variation across providers' performance. See below for performance percentiles with outliers excluded:

Decile	Performance
1	0.48
2	0.56
3	0.62
4	0.67
5	0.71
6	0.74
7	0.76

Decile	Performance
8	0.80
9	0.83
10	0.93

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*)

Not applicable

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*)

Not applicable

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*)

Not applicable

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 nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

The registry dataset provided to us by the National Home-Based Primary Care & Palliative Care Registry did not contain missing data so this test was not performed. Nevertheless, missing data may have been rejected when submitted to the registry in which case those values would not be counted towards measure performance. The registry does not provide data on the rate of rejections that may have occurred for data submitted. Therefore, we cannot obtain missing data for analysis. There is no indication that missing data was systematic, thus their omission would lead to unbiased performance results.

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*)

This test was not performed for this measure. There was no missing data.

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 nonresponders) 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*)

The registry dataset provided to us by the National Home-Based Primary Care & Palliative Care Registry did not contain missing data so this test was not performed. Nevertheless, missing data may have been rejected when submitted to the registry in which case those values would not be counted towards measure performance. The registry does not provide data on the rate of rejections that may have occurred for data submitted. Therefore, we cannot obtain missing data for analysis. There is no indication that missing data was systematic, thus their omission would lead to unbiased performance results.

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.

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

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 electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)

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).

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.

There were no difficulties 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. The use of this measure by 221 providers suggests that the measure is feasible.

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)*.

The Measures, while copyrighted, can be reproduced and distributed, without modification, for noncommercial purposes (e.g., use by healthcare providers in connection with their practice). Commercial use is defined as the sale, license, or distribution of the Measures for commercial gain, or incorporation of the Measures into a product or service that is sold, licensed or distributed for commercial gain. Commercial uses of the Measures require a license agreement between the user and The Johns Hopkins University and the University of California, San Francisco.

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	Payment Program MIPS QCDR www.medconcert.com/NHBCPCR Professional Certification or Recognition Program American Board of Internal Medicine Approved Quality Improvement Activity http://www.abim.org Quality Improvement (external benchmarking to organizations) National Home-based Primary Care and Palliative Care Learning Collaborative www.nhpc2pc.org Quality Improvement (Internal to the specific organization) National Home-based Primary Care and Palliative Care Learning Collaborative http://www.nhpc2pc.org

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

Name of program and sponsor: MIPS QCDR
Purpose: MIPS Performance Reporting
Geographic area and number and percentage of accountable entities and patients included: National
Level of measurement and setting: Patient level; home-based primary care and palliative care
Name of program and sponsor: American Board of Internal Medicine Approved Quality Improvement Activity
Purpose: Practice-based quality improvement
Geographic area and number and percentage of accountable entities and patients included: National
Level of measurement and setting: Patient level; home-based primary care and palliative care
Name of program and sponsor: National Home-based Primary Care and Palliative Care Learning Collaborative
Purpose: Practice-based quality improvement learning collaborative (internal and external)
Geographic area and number and percentage of accountable entities and patients included: National
Level of measurement and setting: Patient level; home-based primary care and palliative care

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

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.*)

According to the CY 2019 Quality Payment Program final rule, CMS intends to "make all measures under MIPS quality performance category available for public reporting on Physician Compare in the transition year of the Quality Payment Program, as technically feasible." These measures include those reported via all available submission methods for MIPS-eligible clinicians and groups.

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.

The developers have obtained feedback from practices via the QCDR on measure feasibility in the following domains: data availability, data accuracy, data standards, and workflow to guide future modifications to the measure. During this process, we regularly receive recommendations to improve the experience of those implementing and reporting on this measure and we follow up on any questions or concerns received by those completing the feasibility assessment. Doing so addresses any issues with interpretation and serves as an important step in the measure development process.

Data from the registry has been shared with the 221 providers involved in submitting data. They receive monthly reports of their performance to assist with their implementation. In addition, the measure developers work with providers and practices to engage in practice-based quality improvement activities using performance data from these measures.The measure development and maintenance process used by the measure developers was a rigorous, evidence-based process that adhered to several key principles, including the following which underscore the role those being measured have played in the development and maintenance process and in providing feedback based on measure implementation: (1)

Collaborative Approach to Measure Development

The measure was developed and maintained through iterative involvement by and input from clinical practices, research (2-5), as well as cross-specialty, multi disciplinary technical expert panels. Representatives of relevant clinical practices of home-based primary care and home-based palliative care, and the specialties of geriatrics, palliative medicine, internal medicine and family medicine participated in our expert panels to advise us throughout the measure development process and as questions arose during measure implementation. Additionally, stakeholders participated in our panels as equal contributors to the measure development process, including stakeholders from relevant professional societies (The American Academy of Home Care Medicine, The American Geriatrics Society, and the American Academy of Hospice and Palliative Medicine). The measure developers also included on its panels individuals representing the perspectives of patients, consumers, private health plans, and employers, including organizations such as the Kaiser Family Foundation, the AARP Public Policy Institute, and the National Partnership for Women and Families. Measure methodologists and coding and informatics experts are also considered important members of the expert panel and were included. This broad-based approach to measure development maximizes the input from those being measured and other stakeholders to develop evidence-based, feasible and clinically meaningful measures.

Data from the registry has been shared with the 221 providers involved in submitting data. They receive monthly reports of their performance to assist with their implementation. In addition, the measure developers work with providers and practices to engage in practice-based quality improvement activities using performance data from these measures.

Members of measure development technical expert panel:

Gresham Bayne, MD, Nant Health; Lynn Beatty, MD, Visiting Physicians Association; Peter A. Boling, MD, Virginia Commonwealth University; Tom Cornwell, MD, Home Centered Care Institute; Eric De Jonge, MD, Medstar Washington Hospital Center Medical House Call Program; Tom Edes, MD, FACP, U.S. Department of Veterans Affairs; Lynn Friss Feinberg, AARP Public Policy

Institute; Alanna Goldstein, MPH, American Geriatrics Society; Jen Hayashi, MD, Johns Hopkins Elder House Call Program; Benneth Husted, DO, Housecall Providers; Julia Jung, CPA, House Call Doctors; Tricia Neuman, ScD, Kaiser Family Foundation; Patricia Tomsko Nay, MD, CMD, American Academy of Hospice and Palliative Medicine; Tom Reed, Senior Advocate Resources; Constance Row, American Academy of Home Care Medicine; Christine Broderick, National Partnership For Women & Families; Theresa Soriano, MD, MPH, Mount Sinai Visiting Doctors Program; Robert Sowislo, Visiting Physicians Association.

Members of the measure validity testing technical panel:

Eric De Jonge, MD, Jen Hayashi, MD, Linda DeCherrie, MD, Steven Landers, MD, Mattan Schuchman, MD, Carla Perissinotto, MD, Thomas Cornwell, MD, Sharon Levine, MD, Carlos Weiss, MD, Mia Yang, MD, Eliza Shulman, DO, MPH

1. Leff B, Carlson CM, Saliba D, Ritchie C. The invisible homebound: setting quality-of-care standards for home-based primary and palliative care. Health Aff (Millwood). 2015;34:21-9.

2. Huber K, Patel K, Garrigues S, Leff B, Ritchie C. Interdisciplinary Teams and Home-Based Medical Care: Secondary Analysis of a National Survey. J Am Med Dir Assoc. 2019 Epub ahead of print PubMed PMID: 30738821.

3. Ritchie CS, Leff B, Garrigues SK, Perissinotto C, Sheehan OC, Harrison KL. A Quality of Care Framework for Home-Based Medical Care. J Am Med Dir Assoc. 2018:818-823.

4. Sheehan OC, Ritchie CS, Fathi R, Garrigues SK, Saliba D, Leff B. Development of Quality Indicators to Address Abuse and Neglect in Home-Based Primary Care and Palliative Care. J Am Geriatr Soc. 2016;64:2577-2584.

5. Fathi R, Sheehan OC, Garrigues SK, Saliba D, Leff B, Ritchie CS. Development of an Interdisciplinary Team Communication Framework and Quality Metrics for Home-Based Medical Care Practices. J Am Med Dir Assoc. 2016;:725-729.

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.

As indicated in 4a2.1.1 above: Data from the registry has been shared with the 221 providers involved in submitting data. They receive monthly reports of their performance to assist with their implementation. In addition, the measure developers work with providers and practices to engage in practice-based quality improvement activities using performance data from these measures.

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.

Feedback Mechanisms

There are several mechanisms to receive measure-related comments and questions from implementers. Comments may be received via email inquiries sent to the developers via the website of the National Home-based Primary Care and Palliative Care Consortium, which has developed a learning community of practices to use the measures and the QCDR to engage in quality improvement activities using the measure. In addition, comments may be sent via the QCDR portal. Finally, the CMS annual QCDR self nomination process is a feedback mechanism. As comments and questions are received, they are reviewed by the developers and, if needed, changes are made in the measure. If comments or questions require expert input, these are shared with the development team and technical expert panel to determine if measure modifications may be warranted.

Feasibility Assessments

The developers have obtained feedback from practices via the QCDR on measure feasibility in the following domains: data availability, data accuracy, data standards, and workflow to guide future modifications to the measure. During this process, we regularly receive recommendations to improve the experience of those implementing and reporting on this measure and we follow up on any questions or concerns received by those completing the feasibility assessment. Doing so addresses any issues with interpretation and serves as an important step in the measure development process.

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

In initial iteration of the measure, the measures for evaluation of activities of daily living (ADLs) and instrumental activities of daily living (IADLs) were two separate measures. Key feedback from the CMS resulted in combining ADL and IADL into a single measure, as well as specification of specific ADLs and IADLs most appropriate for the home care setting to meet measure performance.

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

In initial iteration of the measure, the measures for evaluation of activities of daily living (ADLs) and instrumental activities of daily living (IADLs) were two separate measures. Key feedback from the CMS resulted in combining ADL and IADL into a single measure, as well as specification of specific ADLs and IADLs required to meet measure performance.

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.

See 4a2.2.2. above

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.

The intent of this measure is to improve the care of patients receiving home-based primary care and / or palliative care who have high prevalence of functional impairment. (1) In the years 2016, 2017, and 2018, the number of providers reporting on the measure were 303, 271, and 246, respectively, with rates of measure completion at 56.0%, 70.3%, and 57.5%, respectively. These represent a variable trend in reporting and performance. The variable trend may be related to practice management-related issues, including turnover in practice management, need to focus on completion of other quality measures in the context of the need for practices to prioritize their quality improvement focus on measures identified by practice leadership in the context of performance reporting for various CMS quality programs. However, reporting rates represent but one facet of the quality improvement process.

While the measure developer created the measure with an ultimate goal of improving the quality of care, measurement is a mechanism to drive improvement but does not equate with improvement. Measurement can help identify opportunities for improvement with actual improvement requiring making changes to health care processes and/or structure. In order to promote improvement, quality measurement systems need to provide feedback to front-line clinical staff in as close to real time as possible and at the point of care whenever possible. (2)

1.Ornstein KA, Leff B, Covinsky KE, Ritchie CS, Federman AD, Roberts L, Kelley AS, Siu AL, Szanton SL. Epidemiology of the Homebound Population in the United States. JAMA Intern Med. 2015 Jul;175(7):1180-6.

2. Conway PH, Mostashari F, Clancy C. The future of quality measurement for improvement and accountability. JAMA. 2013 Jun 5;309(21):2215-6.

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.

We have not received reports of unexpected findings resulting from the implementation of this measure. The measure developer has various mechanisms in place for measure users to provide feedback and to identify issues related to the maintenance and implementation of this measure as noted above. We convened several topic-specific technical expert panels comprised of various stakeholders including those being measured to advise us regarding any unexpected findings and actions that can be taken to mitigate them.

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

No unexpected benefits

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.

Yes

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

2524 : Rheumatoid Arthritis: Patient-Reported Functional Status Assessment

2624 : Functional Outcome Assessment

2631 : Percent of Long-Term Care Hospital (LTCH) Patients With an Admission and Discharge Functional Assessment and a Care Plan That Addresses Function

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?

Yes

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.)

Three measures were identified as related to this measure. However, the target population and/or setting for this measure (home based primary care and home based palliative care) differs from each of those identified and listed here. There were no competing measures identified.

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): American Academy of Home Care Medicine

Co.2 Point of Contact: Brent, Feorene, bfeorene@aahcm.org, 440-871-2756-

Co.3 Measure Developer if different from Measure Steward: Johns Hopkins University School of Medicine

Co.4 Point of Contact: Bruce, Leff, bleff@jhmi.edu, 410-550-2654-

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.

Measures are developed and maintained under the aegis of topic-specific technical expert panels (TEPs). The TEPs are comprised of clinicians and other healthcare professionals representing medical specialty societies and other stakeholders. The TEPs provide clinical expertise as well as advice on methodologic questions and review the measures annually to ensure accuracy and adherence to the most current evidence.

Members of measure development technical expert panel:

Gresham Bayne, MD, Nant Health; Lynn Beatty, MD, Visiting Physicians Association; Peter A. Boling, MD, Virginia Commonwealth University; Tom Cornwell, MD, Home Centered Care Institute; Eric De Jonge, MD, Medstar Washington Hospital Center Medical House Call Program; Tom Edes, MD, FACP, U.S. Department of Veterans Affairs; Lynn Friss Feinberg, AARP Public Policy

Institute; Alanna Goldstein, MPH, American Geriatrics Society; Jen Hayashi, MD, Johns Hopkins Elder House Call Program; Benneth Husted, DO, Housecall Providers; Julia Jung, CPA, House Call Doctors; Tricia Neuman, ScD, Kaiser Family Foundation; Patricia Tomsko Nay, MD, CMD, American Academy of Hospice and Palliative Medicine; Tom Reed, Senior Advocate Resources; Constance Row, American Academy of Home Care Medicine; Christine Broderick, National Partnership For Women & Families; Theresa Soriano, MD, MPH, Mount Sinai Visiting Doctors Program; Robert Sowislo, Visiting Physicians Association.

Members of the measure validity testing technical panel:

Eric De Jonge, MD, Jen Hayashi, MD, Linda DeCherrie, MD, Steven Landers, MD, Mattan Schuchman, MD, Carla Perissinotto, MD, Thomas Cornwell, MD, Sharon Levine, MD, Carlos Weiss, MD, Mia Yang, MD, Eliza Shulman, DO, MPH.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2016

Ad.3 Month and Year of most recent revision: 2017

Ad.4 What is your frequency for review/update of this measure? Supporting guidelines, specifications, and coding for this measure are reviewed annually

Ad.5 When is the next scheduled review/update for this measure? 08, 2019

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

The Measures, while copyrighted, can be reproduced and distributed, without modification, for noncommercial purposes, e.g., use by health care providers in connection with their practices. Commercial use is defined as the sale, license, or distribution of the Measures for commercial gain, or incorporation of the Measures into a product or service that is sold, licensed or distributed for commercial gain.

Commercial uses of the Measures require a license agreement between the user and The Johns Hopkins University and the University of California, San Francisco.

Use of the Measures by other health care professionals, where appropriate, is encouraged

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

Limited proprietary coding is contained in the Measure specifications for convenience. Users of the proprietary code sets should obtain all necessary licenses from the owners of these code sets. The developer disclaims all liability for use or accuracy of any Current Procedural Terminology (CPT®) or other coding contained in the specifications.

Ad.8 Additional Information/Comments: