NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

Purple text represents the responses from measure developers.

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

Brief Measure Information

NQF #: 3568

Corresponding Measures:

De.2. Measure Title: Person-Centered Primary Care Measure PRO-PM

Co.1.1. Measure Steward: American Board of Family Medicine

De.3. Brief Description of Measure: The Person-Centered Primary Care Measure instrument is an 11-item patient reported assessment of primary care. Patients complete the PCPCM instrument once a year. These instruments are used to calculate a performance score for the participating entity. That entity could be an individual clinician or a practice. The 11 items of the PCPCM assess primary care aspects rarely captured yet thought responsible for primary care effects on population health, equity, quality, and sustainable expenditures. These include: accessibility, comprehensiveness, integration, coordination, relationship, advocacy, family and community context, goal-oriented care, and disease, illness, and prevention management.

The target population of the PCPCM Performance Measure (PRO-PM) is all patients, active in a practice.

Patients are defined as active if they have had a documented interaction with the practice within 12 months of the patient's birth month. In the PCPCM PRO, patients are presented with 11 structured items. After each item, patients are asked to state their level of endorsement. The same scale is used for all 11 items: Definitely, Mostly, Somewhat, Not At All. Active patients receive the PCPCM PRO through mail, email, or patient portal, during the month of their birth (e.g., patients born in January will receive a request to complete the PCPCM PRO in January).

The PCPCM PRO-PM is calculated as a continuous variable on a 0 to 100 point scale, in which a higher value equates to better quality.

The time frame used to evaluate quality with the PCPCM PRO-PM is one year.

Receiving patient responses in the month of their birth allows a practice to receive monthly feedback in between quality reporting periods.

Scoring for the PCPCM PRO-PM is completed through a simple 4 step process using the PCPCM PRO to assess the broad scope of primary care from a patient's perspective.

Step One: Exclude incomplete patient responses.

Any PCPCM PRO instrument for which a patient failed to answer at least 8 of the 11 items is excluded from calculations.

Step Two: Calculate PCPCM PRO item specific mean scores.

Patients choose one of four response options for each item in the PCPCM PRO instrument. In scoring the PCPCM PRO, the first step requires determining an item mean score for each of the 11 items. Since the instrument scale is word based – Definitely, Mostly, Somewhat, Not At All – each response option must be assigned a value. Values are assigned as follows: Definitely = 4, Mostly = 3, Somewhat = 2, Not At All = 1.

Calculating the mean score for each item then requires looking across all PCPCM PRO instruments received for the entity being assessed during the analysis period. For example, if the entity is a clinician, then all completed (see Step One) PCPCM PRO instruments collected for that clinician are included in the calculation. If the entity is a practice, then all PCPCM PRO instruments collected for that practice are included in the analysis.

An entity's score for each PCPCM PRO item is calculated as a mean, i.e., the summary of all responses across PCPCM PRO instruments received for the entity, divided by the number of instruments received. This process leads to 11 item specific PCPCM PRO scores. Means should be reported to two decimal points.

Step Three: Calculate the PCPCM PRO total score.

The PCPCM PRO total score for the entity is calculated by determining the mean of the 11 scored PRO items. This is done by adding the mean scores of all 11 PRO items and then dividing by 11. PRO means should be reported to two decimal points.

Step Four: Converting PCPCM PRO total scores and to PCPCM PRO-PM performance score.

In order to use the PCPCM PRO as a performance measure for reporting, the 4 point PCPCM PRO scale must be converted to a 0-100 performance scale. To do this, the PCPCM PRO total score for an entity, as calculated in Step Three, is divided by 4 and then multiplied by 100.

Thus, a PCPCM PRO total score of 2.78 (based on a scale of 1-4) becomes a PCPCM PRO-PM performance score of 69.5 (on a scale of 0-100).

The monthly data collection allows for assessed entities to receive regular feedback during the course of the year. However, PCPCM PRO-PM performance scores are calculated based on quality reporting program requirements or a 12-month time frame.

There is no stratification required with the PCPCM.

1b.1. Developer Rationale: Rationale

The PCPCM PRO-PM fulfills the call from the Institute of Medicine and from CMS to create a stakeholder informed, meaningful measure that is an assessment of quality, low burden for implementation and collection, and provides adequate ability to compare performance across clinicians and practices while providing great face validity, transparency and actionable information.[1,2] The PCPCM PRO-PM does just that. It is unusual in its combination of robust internal consistency together with breadth and brevity. Its combination of parsimony - with a single item for each of 11 diverse primary care components - and conceptual coherence - exemplified by the fact that all 11 items load onto a single factor - is the result of an unusually broad and deep amount of preparatory work grounded in diverse stakeholder engagement.[3] This stakeholder engagement enabled the development of meaningful measure items and is the reason why the PCPCM PRO-PM covers 4 of the 8 "cross cutting connections" in the CMS Meaningful Measures Framework (identified as patient-centered and meaningful to patients; fulfill requirements in programs' statues; minimize level of burden for providers; significant opportunity for improvement).2 The PCPCM PRO-PM also addresses a critical quality measure gap as identified by the MACRA Measure Development Plan Technical Expert Panel, of which Dr. Etz – the developer of the PCPCM PRO-PM – was a part.[4]

Benefits and improvements in quality envisioned by use of this measure

The PCPCM PRO-PM is a performance measure that uses the PCPCM PRO instrument. The performance measure is used to assess quality of primary care from patient perspective, comparing an individual clinician's performance to national benchmarks and to other clinicians in their practice, locally, and regionally, as data availability allow.

The performance measure is calculated based on one year of data collection. PCPCM PRO instruments are received monthly by the practice. Active patients (defined as having had contact with the practice in the 12 months preceding their birth month) receive the PCPCM PRO instrument during the month of their birth. Receiving data on a monthly basis allows clinicians to receive feedback on their performance in between annual reporting periods. Such interim feedback enables constant attention and opportunities for correction of performance during any given performance year.

When validating the PCPCM PRO, we tested its concurrent validity with two existing and validated instruments: the Patient Enablement Instrument (PEI)5 and the What Matters Index (WMI).6,7 The PCPCM PRO is well correlated to both the PEI – an assessment of patient self-management – and the WMI – validated to correlate both retrospectively and prospectively with cost and utilization of services. Our comparative analyses used t tests and analysis of variance for continuous variables and c2 for categorical variables (see Table 1 below). PCPCM PRO-PM scores were strongly and positively associated with the WMI and PEI (both p = .0001). It is therefore envisioned that improvements in PCPCM PRO-PM scores, as facilitated by QI activities, will result in both improved patient self-management and reduced cost and utilization of services. See graph in Appendix A.1 (page 3)

The PCPCM PRO-PM is currently being piloted in several settings but is not yet widely implemented. The PCPCM PRO-PM is being piloted by health systems in Colorado, Missouri, Ohio, and Richmond, and within PRIME – a national primary care Qualified Clinical Data Registry (QCDR), hosted by the American Board of Family Medicine. The PCPCM PRO-PM has been endorsed by CMS for use as a QCDR measure in the 2020 MIPS reporting period and is being used by a subset of PRIME members for that purpose. On maintenance review of the PCPCM PRO-PM, we expect to have more information on implementation and scores.

1. Stange KC, Etz RS, Gullett H, et al. Metrics For Assessing Improvements In Primary Health Care. Annual review of public health. 2014;35:423-442.
2. In: Blumenthal D, Malphrus E, McGinnis JM, eds. Vital Signs: Core Metrics for Health and Health Care Progress. Washington (DC)2015.
3. Etz RS, Zyzanski SJ, Gonzalez MM, Reves SR, O´Neal JP, Stange KC. A New Comprehensive Measure of High-Value Aspects of Primary Care. Ann Fam Med. 2019 May;17(3):221-230.
4. Meaningful Measures Framework of CMS. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/CMS-Quality-Strategy Accessed June 27, 2020.

S.4. Numerator Statement: The PCPCM PRO-PM allows all patients to report their assessment of the quality of primary care received through responses to PCPCM PRO instrument.

The target population is all active patients in a practice during the performance reporting period. A patient is defined as active if the patient has had a documented interaction with the practice within 12 months of the patient's birth month. The PCPCM PRO is the same for all patients, regardless of age. Because the PCPCM PRO applies to all patients and is not particular to a clinical encounter, it is administered once a year to each patient during their birth month.

The target population is defined the same, regardless of unit of analysis (clinician or practice).

The numerator is the sum of all PCPCM PRO scores for active patients.

S.6. Denominator Statement: The target population for the denominator is the same as for the numerator.

The denominator is the total number of complete PCPCM PRO instruments received in the reporting period. A completed PRO instrument is defined as a PRO instrument for which the patient has responded to at least 8 of 11 items.

S.8. Denominator Exclusions: None.

De.1. Measure Type: Outcome: PRO-PM

S.17. Data Source: Instrument-Based Data

S.20. Level of Analysis: Clinician : Group/Practice, Clinician : Individual

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? N/A

Preliminary Analysis: New Measure

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcome measure include providing empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service; if these data not available, data demonstrating wide variation in performance, assuming the data are from a robust number of providers and results are not subject to systematic bias. For measures derived from patient report, evidence also should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.

Evidence Summary:

- This is a new outcome measure utilizing patient-reported data at the individual clinician and practice level gathered through the 11 items Person-Centered Primary Care Measure (PCPCM) instrument in order to assess primary care aspects rarely captured yet thought responsible for primary care effects on population health, equity, quality, and sustainable expenditures. These include: accessibility, comprehensiveness, integration, coordination, relationship, advocacy, family and community context, goal-oriented care, and disease, illness, and prevention management.
- Developer provides a logic model depicting improved patient experience and health outcomes after initial contact with primary care, evaluation of experience, review by clinicians, and improvement of care processes.
- Developer provides evidence of importance to patients. The PCPCM patient-reported outcome (PRO) was created through a multi-step approach and was evaluated through cognitive interviews among patients and pilot testing. Two groups were surveyed to determine if the 11 items of the PCPCM PRO were meaningful to participants: clinicians and patients.
    - Among the 285 clinicians surveyed, 79% said the items were important to their practice, 85% said the items were personally important to them and 84% said it would it be meaningful to them to receive their patients' responses to these items.
    - Among the first group of 1,140 patient's surveyed, 87% of patients said the survey was meaningful and 60% of patients said receipt of their answers would help with their care.
    - Among the second patient group of 1,344, every item in the PCPCM PRO was identified as meaningful by at least 80% of patients responding. Additionally, the majority of patients said their answers would not be affected by the pandemic.
- Developer provides health care actions for each of the eleven items on the PCPCM that can lead to improved performance on the measure. These were not supported by empirical evidence.

Exception to evidence

Not Applicable

Question for the Committee:

- Is there at least one thing that the provider can do to achieve a change in the measure results?
- If derived from patient report, does the target population value the measured outcome and finds it meaningful?

Guidance from the Evidence Algorithm

Outcome: PRO-PM (Box 1)→ Relationship between the PRO and at least one healthcare action not demonstrated by empirical data (Box 2) → Pass

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

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.

- Among 6 practices, there were significant differences (p=0.004) in PCPCM PRO-PM scores with a moderate effect size (at least .5 standard deviation). Performance scores for the six sites ranged from 0.84-0.90.
- Among 16 clinicians, there were significant differences (p=0.0001) in PCPCM PRO-PM scores with a large effect size (>1.0). Performance scores for the 16 clinicians ranged from 0.75-0.92.

Disparities

- The developer states that evidence to date shows no correlation between self-defined minority status and patient assessment of primary care, using the PCPCM

Questions for the Committee:

- Is there a gap in care that warrants a national performance measure?
- If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?

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

Committee Pre-evaluation Comments:
Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

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

- Nearly 80% of providers felt the survey items were important to their practices, to them personally, and thought gaining this information from their practices would be meaningful. Similarly, patients reported the survey is meaningful (84%) and did not feel the current pandemic would impact their responses.
- This is a patient reported measure of person centered primary care. The measure consists of 11 items that would be of value to patients and providers.
- na

- Survey and conference used to gather stakeholder input into development of the survey instrument. Specific evidence related to score on this instrument equating to a care outcome was inferred that better scores will mean more engaged patients and better self-management and reduced cost and utilization of services. Instrument is being piloted by several settings.
- No major concerns
- This is a patient reported outcome measure. The instrument to be completed by primary care patients was created with input from provider, patient, and employer focus groups. The survey for meaningfulness of the draft instrument was completed by 2484 patients. Each of the eleven items in the instrument were rated as meaningful by at least 80% of those surveyed. Twelve sources from medical and health services research are cited that demonstrate the strong connection between patient experience of care and traditional health care outcomes. There is no systematic review of the literature, as this is not required for a patient reported outcome performance measure (PRO-PM).
- Evidence seems low for needing such a patient centered evaluation and how it can improve quality.
- Evidence is strong for the measure and the results may have the ability to improve the patient experience of care and service delivery at the provider level.
- PRO aligns

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

- From comparison of 6 practices, and 16 providers there was moderate to large effect size - larger among the individual provider comparisons. Subgroup data is not apparent.
- new measure. There is variation in the site and individual performance scores. It's not clear though what this range really is. The measure defines the score on a 0-100 scale but the performance gap give scores that are all less than 1.
- na
- Variability in performance score on measure among clinicians and clinics is provided. Does this warrant a national performance measure??
- Yes, gap exists
- A pilot project with six practices and sixteen clinicians showed significant differences between the members of each of these categories. Using this project, there is no evidence to date showing a correlation between self-defined minority status and patient assessment of primary care.
- Not clear is any disparities or other issues, including those who do not complete their responses.
- The tool was tested with 6 practices and 16 clinicians. Data provided did not provide strong evidence in gap in care for the population, but gaps in care of the 6 practices and 16 practitioners.
- Moderate evidence provided for the opportunity for improvement

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

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.

Composite measures only:

2d. Empirical analysis to support composite construction. Empirical analysis should demonstrate that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct.

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

Evaluators: Marybeth Farquhar, Patrick Romano, Sean O'Brien, Terri Warholak, John Bott, Eric Weinhandl, Sherri Kaplan, ZQ Lin, David Nerenz

Methods Panel Review (Combined)

Methods Panel Evaluation Summary:

This measure was reviewed by the Scientific Methods Panel and discussed on the call. A summary of the measure and the Panel discussion is provided below.

Reliability

- Reliability testing conducted at the data element level:
    - Data element level testing was conducted using exploratory factor analysis, Rasch item fit statistics, and Cronbach's alpha testing
        - Exploratory Factor Analysis: This analysis resulted in the identification of a single factor: person centered primary care. This was further confirmed through calculated Eigenvalues of 6.9 for the patient online group and 4.7 for the patient point of care group.
        - Rasch Item Fit
            - Rasch item fit statistics ranged from 0.62 to 1.44 for the patient online group. Rasch item reliability was 0.99 for this group.
            - Rasch item fit statistics ranged from 0.55 to 1.49 for the patient point of care group. Rasch item reliability was 0.98 for this group.
        - Cronbach's alpha 0.91 for point of care group, 0.95 for the online group
    - Score level testing was conducted using ICC analysis between providers
        - Clinician: individual—ICC ranging between 0.76 and 0.94 and Guttman reliability ranging between 0.79 and 0.94
        - Group/practice—Split half reliability ranges from 0.86 and 0.95

Validity

- Validity testing conducted at the data element level:

- Data element testing was conducted using face validity for patients (n=30) who found the results to be meaningful to them and easy to answer. Clinicians (n=285) also found the results to be meaningful.
- Data elements were further tested against two external measures of quality at the patient level: the Patient Enablement Instrument (PEI) and the What Matters Index (WMI).
- Score level testing was conducted by comparing the PCPCM with
  - Clinician: individual—Pearson correlations between the PEI and the PCPCM at the clinician level ranging from 0.39 to 0.65
  - Group/practice—Pearson correlations between the PEI and the PCPCM are consistent from practice to practice, in the upper 0.5, and in agreement with the correlation for the total sample as well.
- Threats to validity:
  - "The testing form claims no exclusions, thus the exclusion related questions in 2b2 are skipped. However, the MIF notes that completed instruments with less than 8 of the items completed are excluded. It would be appropriate to state that exclusion in the testing form & respond to the corresponding questions regarding exclusions."
  - "The F-tests of homogeneity among practice scores and clinician scores are insufficient evidence of an ability to detect meaningful differences among units of analysis."

- "Data are very favorably skewed with very small differences between highest and lowest performers for the 6 practices and 16 individual physicians for which data were reported. It is not clear if the differences that were considered as meaningful (0.12 points different based on Cohen's d of 0.20) are actually meaningful as they are not calibrated against another validity variable. The paragraph describing statistical power is not sufficiently detailed to evaluate."

Questions for the Committee regarding reliability:

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

Questions for the Committee regarding validity:

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

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

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

Committee Pre-evaluation Comments:

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

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

- Is calculating a mean valid for questions in this tool? Is Median a better approach. Likewise, summing a score for each item assumes that each item has equal weight. An assumption on the part of the developers. At the individual data element level, factor analysis, Rasch Item fit and high Cronbach alpha do show reliability. At a total score level, there was good reliability note at the clinician and practice level.
- They are excluding those surveys that answer less than 8 questions. specifications seem straightforward.

- Any examination of patients responses <18? Any concerns? (looks like patient characteristics are 18+)
- No real concerns.
- No major concerns
- Data come from the measure's patient completed instrument. The steps involved in taking answers on the instrument's eleven items (using a four point scale) and converting them to a score (on a hundred point scale) are well described and clear.
- Many issues that perhaps the developer could discuss, but the single biggest it is the generally high level of "satisfaction" so more difficult to evaluate both individuals and improvement over time.
- The tool should be reliable.
- No concerns

2a2. Reliability - Testing: Do you have any concerns about the reliability of the measure?

- No, noting the challenges of assuming each item is of equal importance to the individual completing the survey.
- no
- testing conducted at the data element level
- No concerns.
- No major concerns
- Reliability testing was done on the critical data elements (the eleven questions of the instrument) and on the performance measure score. All the tests were interpreted as supporting reliability.
- Specifics of issues should be discussed with developer.
- I would have liked to see broader testing to determine reliability.
- No concerns

2b1. Validity -Testing: Do you have any concerns with the testing results?

- Face validity done at patient and clinician levels found results to be "meaningful". Comparison with two other measures (PEW and WMI) were reasonable
- no
- Data elements tested against two other measures
- No concerns
- No major concerns
- Validity testing was done on the critical data elements and on the performance measure score. In the test of the performance measure score, the measure instrument results were compared with the results of the older and established Patient Enablement Instrument. All tests support validity.
- Not sure the results really focus on "patient-centered" or more generalized satisfaction.
- No concerns with the validity testing results, though limited.
- No concerns

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

- Not presented. Does age, gender, SES, SDoH impact validity?
- no risk adjustment
- na
- Exclusions are not indicated but are surveys with missing items then excluded?
- No major concerns
- There were no exclusions or risk adjustments/stratification.
- Clarification of the exclusion is not completing the 8 of the 11 items on the survey.
- Measure is for patient-centered practice. No risk adjustment factors stated.
- No concerns

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

- No exclusions are noted.
- the presented data shows a very narrow range of performance.
- Panel is satisfied with the validity analyses
- How are incomplete surveys handled? Not clear that the small differences represent true differences in quality.
- No major concerns
- The results support the contention that PCPCM PRO-PM performance scores are able to distinguish between practices and between individual clinicians, providing that individual clinicians have an adequate sample size (at least 30) of patients and the practices have an adequate sample size (at least 50). Differences in mean performance scores were much greater for the clinicians than for the practices and this is logical, as the practice is an aggregate measure of clinicians. Since the PCPCM PRO-PM can differentiate among practices and clinicians, it could be used as an outcome measure to evaluate clinician or practice level performance. There is only one set of specifications. Less than 1% of respondents, among a total sample of 2,552 patients, had incomplete instruments (fewer than 8 items answered). Received instruments that had fewer than eight (8) responses were considered to have missing data and therefore were deemed incomplete. The amount of missing data (<1%) were insufficient to examine trends across providers.
- The issue of differences in the way the feedback is collected could be a threat to validity.
- Results should be comparable. No identified threats to validity. No identified missing data. Patients could refuse to participate in the survey resulting in missing data or no response. This could impact the validity of the results.
- No concerns

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.
    - Data elements used are collected directly from patients. Patients are invited to fill out the PCPCM PRO instrument electronically. In almost all cases, patients are sent an email with an embedded link either to an electronic survey platform, or to an electronic Patient Reported Outcomes (PRO) module as part

of the PRIME registry. The most likely format will be electronic sources, however paper-based instruments can be used.

- In all implementation of the PCPCM PRO-PM to date, performance scores and feedback are provided electronically to practices and clinicians. PCPCM PRO-PM scores are calculated at the point of data collection and then shared with the measured entity.

Questions for the Committee:

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

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

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

- Available as an online or paper survey. May create a selection bias for who would complete the instrument.
- data collection will be electronic directly from pts via a link. paper based forms will also be used. Will need to consider how the paper based forms are collected. Under-served populations might be disproportionately skewed to paper forms which may have lower response rates.
- Electronic or paper (has paper been tested?)
- Survey is not part of care delivery now. Plan an electronic format but how to reach patient's without electronic capability.
- No major concerns
- The measure instrument has eleven questions. The most likely format will be electronic sources but paper-based instruments can be used. Based on electronically collected PCPCM PRO instruments from over 4,000 patients, the average time to complete the PCPCM is 60-90 seconds. The PCPCM PRO-PM was used in 10 practices that do not share an EHR, patient portal, or registry. PRO-PM scores are calculated at the point of data collection and then shared with the measured entity. There are no fees or other requirements to use any aspect of the measure.
- Seems feasible.
- Measure is feasible as it is an 11 question survey completed by the patient electronically. Non-response rate could be high. Not all patients may look at the email or click on the link. It is, however, feasible.
- Patient completed survey, moderate feasibility

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

- Part of the PRIME Qualified Clinical Data Registry (QCDR) in the CMS Quality Payment Program – Merit-based Incentive Payment System (MIPS).
- Measure was submitted for consideration by the measure applications partnership (MAP) for the MIPS program during the 2020-21 measure review cycle. Feedback will be added once MAP review is complete.

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

- Clinicians and practices were provided with summary performance scores, as well as PCPCM PRO item level scores (as compared with benchmarks) to enable easy identification of areas in need of improvement.
- Clinicians have access to their data on their PRIME dashboard on an on-going basis and dashboards are updated as new data becomes available

***Questions for the Committee**:*

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

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

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:

- Measure has not been implemented and therefore does not have year-over-year performance data for review.

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 did not receive any comments regarding burden or unexpected negative consequences related to adoption. A more thorough analysis of use experience by clinicians and patients is planned over the next 12 months.

Potential harms: The developer did not identify any potential harm.

***Questions for the Committee**:*

- How can the performance results be used to further the goal of high-quality, efficient healthcare?

- Do the benefits of the measure outweigh any potential unintended consequences?

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

Committee Pre-evaluation Comments:

Criteria 4: Usability and Use

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

- Results available thru the PRIME dashboard (MIPS data). Feedback has been sought from those being surveyed, clinicians being measured - performance and implementation.
- feedback provided
- Used in an accountability program
- Not publicly reported. Part of the PRIME Qualified Clinical Data Registry (QCDR) in the CMS Quality Payment Program – Merit-based Incentive Payment System (MIPS). No feedback from those being measured is provided.
- No major concerns
- Clinicians and practices were provided with summary performance scores, as well as item level scores (as compared with benchmarks) to enable easy identification of areas in need of improvement. Clinicians have access to their data on their PRIME dashboard on an on-going basis and dashboards are updated as new data becomes available. The PCPCM PRO-PM has been endorsed for use as a PRIME QCDR measure in the CMS QPP MIPS program. The creators expect to have data to report the implementation and adoption of the measure within 12 months. Also, feedback on measure performance and implementation from the measured entities is planned within the next 12 months. The lead measure developer and researcher held a live webinar for all PRIME practices.
- Meets the requirements for Use.
- Not currently publicly reported. Although providers are given access to their results, it is not clear if there was a feedback loop for the performance measure or how feedback would be considered.
- Planned use in MIPS for 2020-2021

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

- Probably no unintended consequences. Some of the questions may provide some conceptual challenges for individuals completing the survey (e.g. - what does it mean to patients that my doctor stands up for me?)
- no obvious harms.
- na
- Not in use. No clear harms.
- No major concerns
- Each item of the PCPCM is actionable. For example, if a clinician scores poorly on the item “Over time, this practice helps me to meet my goals”, clinicians can use that feedback to make sure discussion of patient

goals is part of the clinical encounter. Each clinician or practice can create quality improvement activities best suited to their context. The items within the measure instrument used to calculate the performance measure are also each individually supported by evidence from the literature as having a strong effect on desirable health outcomes. No practices reported negative experience or consequences.

- Appears to meet the requirements for Usability.
- If implemented correctly and patients complete the survey, if could improve high-quality care. No harms were cited by the developer.
- No concerns

Criterion 5: Related and Competing Measures

Related or competing measures

- None

Harmonization

- None

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

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

- None
- none
- na
- None
- None identified
- The CAHPS Clinician & Group Surveys (CG-CAHPS) Version 3.0 - Adult, Child (NQF #0005) targets all patients who have been seen in ambulatory care settings. The PCPCM PRO PM targets all patients who have been to a primary care practice. The CG CAHPS measure focuses on patient experience of care delivery that is encounter specific and is limited to domains such as communication and access. The PCPCM PRO PM measure is conceptually dissimilar to CG CAHPS and does not need to be harmonized with it.
- Doesn't apply as New Measure with no related or competing measures.
- No competing measures.
- NA

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 01/21/2021

- **Comment by: American Academy of Family Physicians**

The American Academy of Family Physicians is highly supportive of endorsement of the person-centered primary care measure (PCPCM). This measure evaluates the key functions of primary care that patients, clinicians, employers, communities, and health systems value most. Primary care measures must move beyond disease-specific criteria to assess the unique features of primary care most responsible for better outcomes and lower costs and value. The measure recognizes the patient as a valuable source of knowledge about many important aspects of care.

The heart of primary care does not focus on a diagnosis, yet current measures continue to emphasize diagnosis and procedures. The PCPCM focuses on integrating, personalizing, and prioritizing care. Its eleven items (plus one optional question) form an evaluation of access, continuity, comprehensiveness, coordination, advocacy, family and community context, and goal-oriented care. These fundamental elements are associated with better health, equity, quality, and sustainable health care expenditures and are unique to primary care.

The measure is brief, has high face validity, and is understandable by patients and clinicians (e.g., high transparency). It has been tested in many cultures and developers have noted their analysis does not indicate a need for case or risk-adjustment. The measure will help clinicians identify areas of primary care in which their performance is weak to help direct improvement efforts. Over forty improvement activities have been identified that are relevant to the measure.

- **Comment by: Blue Cross BlueShield of Massachusetts**

NQF Measure #3568 Person-Centered Primary Care Measure Patient Reported Outcome Performance Measure (PCPCM PRO-PM) requires further development before it should receive NQF endorsement and be considered ready for high-stakes uses such as performance-based payment and public reporting. For this measure, the most critical area in need of development is case-mix adjustment. To our knowledge, the PCPCM has not undergone empirical analysis to assess the need for case-mix adjustment and to develop case-mix adjustment methods. It is plausible that PCPCM scores, which include items that implicitly assume a need for care "from multiple places" and a long enough relationship to "have been through a lot together" vary substantially according to patient age, health status, and tenure with the index practice. The clinician-level ICCs reported for the PCPCM are likely to be misleading when the underlying measure is not valid for interunit comparisons--for example, because case-mix adjustment is needed but has not been developed. In the absence of case-mix adjustment, high ICCs can result from differences in case-mix rather than differences in providers' true performance. To investigate and remediate this threat to validity, we suggest that the measure developers analyze, based on a large PCPCM fielding that reflects a wide array of practices, the relationships between standard CAHPS case-mix adjustment variables (at a minimum) and PCPCM scores--and then develop case-mix adjustment methods and re-estimate the interunit reliabilities of PCPCM PRO-PM scores based on valid (i.e., case-mix adjusted) comparisons. As a secondary concern, practice-level interunit reliabilities should be calculated if this measure is intended to be applicable to practices (i.e., not be restricted to measurement of individual clinicians).

- Of the 2 NQF members who have submitted a support/non-support choice:
  - 1 supports the measure
  - 1 does not support the measure

Combined Methods Panel Scientific Acceptability Evaluation

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 3568

Measure Title: Person-centered primary care measure

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.*

   Panel Member #8: Measure description in MIF De.3 is quite confusing. Step two indicates item specific mean score, this has to be done at group level (individual or group practice). This will lead to 11 group level PRO scores. Step three indicates total score is obtained through averaging all 11 group level item scores. Based on this, it is not clear what Step four does. From step one to step three, there is no calculation of patient level PCPCM PRO total scores. So it is not clear how to aggregate PCPCM PRO total scores.

2. **Briefly summarize any concerns about the measure specifications.**

   Panel Member #1: No concerns.

   Panel Member #6: I did understand the logic of the measure specification, but I do think that a schematic would be a very helpful addition, as a visual may be complementary.

   I do wonder about whether the conversion of the response scale to the values of 1, 2, 3, and 4 has a conceptual basis in this application. I also wonder about the conceptual basis of equal weighting for all 11 items.

   Panel Member #4: No concerns.

   At first, I had concerns about the use of averages for an ordinal scale but the fact that the items fit the Rasch model and approximated interval level data, made me feel better about this.

   Panel Member #6:

   Panel Member #7: It is not clear how patients will be screened for the method of receiving the instrument – by mail, email or patient portal, how they will be followed up for return of the instrument, how or to whom practices will return the instrument for scoring, how data will be processed and reported (see S.18), especially if integrating paper vs emailed or patient portal data. There does not appear to be a Measures Steward. It is also not clear how practices are to monitor numbers of instruments returned – at the individual clinical, overall practice, etc.

   Panel Member #8: See above

   Panel Member #9: The specifications for the survey instrument itself seem clear and unambiguous, but the testing seems to have been done on a convenience sample, and there is no description of any formal sampling method to be used in future uses of the measure. Even if the intended future uses only involve convenience sampling, that point could have been made explicitly.

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**

   Panel Member #3: Note: I believe the developers use "measure score" to refer to the average across 11 items in a single patient. They do not directly estimate the reliability of scores calculated at the provider level i.e. averages across patients within providers. Analyses reported in Section 2b4 (identification of statistically significant and meaningful differences in performance) shed light on the reliability of provider-level averages of interest indirectly.

4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure**
   ☒ **Yes** ☒ **No**

   Panel Member #5: While testing was conducted with the data source employed for the measure, the route of administration of the instrument differed to some degree between. Specifically, the MIF states: "…patients receive the PCPCM PRO through mail, email, or patient portal…" [p1]. In comparison, testing form states the reliability testing was performed on "the patient online group" [p9].

5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of patient-level data conducted?

   ☒ **Yes** ☐ **No**

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

   Submission document: Testing attachment, section 2a2.2

   Panel Member #1: Used factor analysis and Cronbach's alpha to test survey. Adequate.

   Panel Member #3: Developers used psychometric methods (factor analysis, Rasch item fit statistics, Cronbach's α) to assess instrument properties such as dimensionality, internal consistency, and the reliability of summary scores calculated across 11 items at the patient level.

   In Section 2b4 (identification of statistically significant and meaningful differences in performance), the developers use ANOVA to assess evidence of between-provider variation. The F statistics and other results from this section shed light on the reliability of scores aggregated at the provider level (clinician or practice).

   Panel Member #4: Critical data element: Factor analysis and Rasch item fit and item reliability were used.

   Measure score: Cronbach's alpha.

   Samples included point of care and electronically gathered data.

   These methods seem appropriate.

   Panel Member #5: The methods used for reliability testing was appropriate for the instrument / measure.

   '**critical data elements**
   exploratory principal axes factor analysis
   -…exploratory principal axes factor analysis within the patient online group (n=2,229). We also examined the scree plot and Eigenvalues.
   Rasch item fit statistics
   -Rasch item fit statistics were computed for each PCPCM PRO item.
   performance measure score
   Cronbach's α reliability coefficient
   -…estimate of reliability best fits what we expect the total score to provide, i.e., we expect every item to contribute independently to the total score.' [p9]

   Panel Member #6: The measure steward reports results in a peer-reviewed publication. That publication included principal factor analysis of critical data elements, Rasch modeling, and Cronbach's alpha coefficient estimation for the performance measure reliability.

   Panel Member #7: The application reports patient level reliability only, and reports construct validity as evidence of reliability.

Panel Member #8: Critical data elements testing included exploratory factor analysis and Rasch item ti statistics. The developer used Cronbach's alpha as performance measure score reliability. However, Cronbach's alpha is still really about critical data element reliability as it is about instrument. No real measure reliability testing was included.

Panel Member #9: The measure developer was clearly confused about the NQF terminology of data element and measure score. The developer's use of "data element" referred to individual survey items; their use of "measure score" referred to the total survey instrument score. No analysis at all was done with aggregate data at the individual clinician or practice level, in spite of the developer's description of the sites of data collection and the statement about a minimum required sample size of 30 at both individual clinician and practice levels. The only reliability testing done was about the instrument itself; not about the measure as a quality measure at practice or individual clinician level.

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

Submission document: Testing attachment, section 2a2.3

Panel Member #1: Acceptable

Panel Member #3: Analyses of instrument properties appear to be rigorous but I don't have a strong appreciation for the importance of properties such as one-dimensionality so I will defer to the developers and other reviewers.

Analyses of meaningful differences suggest that there was evidence of between-provider signal variation that was large enough to detect statistically.

Panel Member #4: For the most part, I have no concerns. However, the authors sate "The Cronbach alpha internal consistency reliability coefficient was computed to be 0.95 for the patient online group and 0.91 for the patient point of care group. The corrected item-total correlations for the patient online group ranged from 0.66 to 0.84 for the 11 items in the scale. The corrected item-total correlations for the patient point of care group ranged from 0.27 to 0.86 for the 11 items in the scale." This makes me wonder… Why is the point of care lower bounds of the item total correlations so much different than electronic? Are the 2 forms equivalent?

Panel Member #5: The reliability testing were high in most of the test findings.

'critical data elements
exploratory principal axes factor analysis

Eigenvalues of 6.9 for the patient online group and 4.7 for the patient point of care group.

Rasch item fit statistics

-Rasch item fit statistics ranged from 0.62 to 1.44 for the patient online group. Rasch item reliability was 0.99 for this group.

-Rasch item fit statistics ranged from 0.55 to 1.49 for the patient point of care group. Rasch item reliability was 0.98 for this group.

performance measure score
Cronbach's α reliability coefficient

-…0.95 for the patient online group and 0.91 for the patient point of care group. The corrected item-total correlations for the patient online group ranged from 0.66 to 0.84 for the 11 items in the scale. The corrected item-total correlations for the patient point of care group ranged from 0.27 to 0.86 for the 11 items in the scale.

-…0.91 for the clinician and group practice dataset. The corrected item-total correlations for this group ranged from 0.58 to 0.79 for the 11 items in the scale.' [p12]

Panel Member #6: Principal factor analysis identified only one factor. Rasch item reliability was 0.98 and 0.99 in the point of care and online groups, respectively. Cronbach alpha coefficient values exceeded 0.9.

Panel Member #7: Not demonstrated at the physician or group practice level.

Panel Member #8: Instrument reliability including Chronbahc's alpha is acceptable, however, measure score reliability is missing.

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.**

    Panel Member #1: Testing results. Point of care tends to skew the results to the positive.

    Panel Member #2:

    Panel Member #4: No concerns except those mentioned above.

    Panel Member #5: As noted in Q6 & Q7: The methods used for reliability testing was appropriate for the instrument / measure, and the results were high in most of the test findings.

    Panel Member #6: The constellation of analyses indicate high internal reliability of critical data elements and the measure itself.

    Panel Member #7: The application is incomplete, does not report reliability at the appropriate level and confuses construct validity with reliability assessment.

    Panel Member #8: No assessment of measure score reliability was performed.

    Panel Member #9: The survey instrument seems to have good reliability. The measure score-level reliability was not tested and is therefore unknown.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

    Panel Member #1: No exclusions other than incomplete survey (must answer 8 of 11 items to be included).

    Panel Member #2:

    Panel Member #3: None

Panel Member #4: There are no exclusions. I have no concerns.

Panel Member #5: The testing form claims no exclusions, thus the exclusion related questions in 2b2 are skipped. However, the MIF notes that completed instruments with less than 8 of the items completed are excluded. It would be appropriate to state that exclusion in the testing form & respond to the corresponding questions regarding exclusions.

Panel Member #6: There are no exclusions, aside from nonresponse. However, the influence of nonresponse is unclear. It is entirely possible that some individual components of the score are more likely to elicit response than others.

Panel Member #7: None.

Panel Member #8: My concern is more with non-response bias.

Panel Member #9: The developers propose no measure exclusions, but it would seem a bit odd that a patient who had just one drop-in visit with a practice or clinician would be included in the denominator for a measure, since he/she would have no real basis for evaluating the practice on most of the dimensions included in the survey.

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

    Submission document: Testing attachment, section 2b4.

    Panel Member #1: No concerns.

    Panel Member #3: Provider-level performance testing was based on a limited sample size (n = 7 practice groups, n =~50 clinicians) and were not necessarily representative of the national experience. I am not particularly concerned about this but seemed worth mentioning as a limitation.

    Panel Member #4: No concerns.

    Panel Member #5: The concern is that the testing was performed with a small set of providers: 6 groups & 16 individual clinicians. Further, we don't receive an explanation as to the representativeness of those providers sampled. I think this is important given the relatively modest testing results regarding differences in performance scores.

    Panel Member #6: The F-tests of homogeneity among practice scores and clinician scores are insufficient evidence of an ability to detect meaningful differences among units of analysis.

    Panel Member #7: Data are very favorably skewed with very small differences between highest and lowest performers for the 6 practices and 16 individual physicians for which data were reported. It is not clear if the differences that were considered as meaningful (0.12 points different based on Cohen's d of 0.20) are actually meaningful as they are not calibrated against another validity variable. The paragraph describing statistical power is not sufficiently detailed to evaluate.

    Panel Member #9: There is no information provided on this point – we have no idea how well the measure can identify meaningful differences in performance.

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.
    Panel Member #1: N/A
    Panel Member #4: N/A
    Panel Member #5: No concern in this regard.
    Panel Member #6: This is not applicable.
    Panel Member #8: Accounting for different response rates across practices will be key that is missing.
    Panel Member #9: The on-line and in-person versions of the survey appear to perform in similar ways.

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

Submission document: Testing attachment, section 2b6.

Panel Member #1: No concerns.

Panel Member #2:

Panel Member #3: A limitation that applies to all voluntary surveys is the potential for nonresponse bias. The developers did not report the response rate in their performance measure validation but presumably the number of responders (n=1205) was less than the number of email invitations.

Panel Member #4: With respect to missing data on complete questionnaires, the developers state at less than 1% of the questionnaires were incomplete and that this will have a negligible impact.

I agree with this.

However, I believe non-response error is a significant threat and should be investigated.

Panel Member #5: As noted in Q12 (above): The testing form claims no exclusions, thus the exclusion related questions in 2b2 are skipped. However, the MIF notes that completed instruments with less than 8 of the items completed are excluded. It would be appropriate to state that exclusion in the testing form & respond to the corresponding questions regarding exclusions.

Panel Member #6: The measure steward suggests that surveys with between one and seven responses are rare.

Panel Member #7: The data on response rates are missing and that provided for missing data at the patient level are insufficiently detailed to evaluate.

Panel Member #8: For a PRO-PM measure, it is important to address potential non response bias issue.

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**

Panel Member #1: Acceptable.

Panel Member #4: The developer states "In our published study using the patient dataset,[1] the PCPCM showed no association with patient's age, sex, or number of years knowing the physician, and in online samples that asked about race, the PCPCM scores showed no association with race. Therefore, to foster ease of scoring and interpretation, we do not recommend adjusting for these factors. Because the recommended use of the PCPCM is for individual clinicians or practices to conduct a self-assessment, and to use their total and individual item scores to improve the patient-centeredness of their care, in

comparison to themselves rather than to others, we do not recommend any further risk adjustment or stratification."

I do not agree with this… While it might be acceptable to not risk adjust for this initial submission, I think that stratification could be helpful to sites and to clinicians to assess how patients in different groups view their primary care quality. I have a feeling this could be very instructive.

Panel Member #5: The testing of several potential risk factors, as noted in 2b3.2, is appropriate. Further, given the findings it's logical to then not include these variables in risk adjustment.

However, the instrument can be completed in several manners. . Specifically, the MIF states: "…patients receive the PCPCM PRO through mail, email, or patient portal…" [p1]. We've seen other instruments have tested & included risk factors for the mode of completing the instrument (e.g. HCAHPS). Knowing that mode of administration can impact one's scoring, the measure developer should have tested their difference modes of completing the instrument.

Panel Member #6: No risk adjustment is employed**.** Whether this is an logical strategy is unclear, considering my lack of subject matter expertise.

Panel Member #7: They do not risk adjust**,** despite evidence from other PROs such as CAPHS, that age, gender, education status and overall health are important risk factors requiring adjustment for fair comparisons.

Panel Member #8: The developer's reasons for not risk-adjusted didn't address issues like why proxy response should not be accounted for and non-response bias.

Panel Member #9: The developers propose no risk adjustment, and in the context of self-assessment by individual clinicians or practices, this may be acceptable. If the measure were ever used for comparative purposes, though, like either public reporting or pay-for-performance, some attention to risk adjustment would have to be made, as the measure scores would almost certainly be affected by some combination of clinical case mix and social case mix variables**.**

For cost/resource use measures ONLY:

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

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

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

VALIDITY: TESTING

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

    Panel Member #3: Note: I believe the developers use "measure score" to refer to the average across 11 items in a single patient. They did not directly assess the validity of scores calculated by aggregating across patients within providers. This is why I left the "measure score" box above unchecked.

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

    Panel Member #1: Appropriate

    Panel Member #3: Validity was assured through qualitative (cognitive testing) interviews among patients and pilot testing surveys of patients and clinicians to determine if they agreed that individual questions were meaningful and relevant. The developers also assessed patient-level correlation between their score and two existing PRO's – the Patient Enablement Instrument and the What Matters Index.

Panel Member #4: Critical Data Elements – Validity Testing: Cognitive interviews among patients, Pilot testing of the PCPCM PRO and Survey for meaningfulness

These methods seem appropriate to me.

Performance Measure Score – Validity Testing: concurrent validity

It is not clearly stated what test is being done.

Panel Member #5: The validity testing methods appear to be appropriate for this given measure.

Data element

[1] Cognitive interviews among patients:

- we asked participants to read each item aloud and then to explain in their own words what the item is asking
- we asked participants to pick one of the four response options for each item, followed by asking them to share in their own words why they chose that answer.

[2] Pilot testing of the PCPCM PRO

- we fielded the PCPCM PRO to a pilot set of online participants
- online pilot fielding included each of the 11 items, as well as questions following each of the 11 items that asked whether the question was hard to understand, whether the question was hard to answer, and whether the question seemed relevant to their primary care experience.

[3] Survey for meaningfulness

- groups were surveyed to determine if the 11 items .. were meaningful to participants: clinicians and patients.

Measure Score

- We used concurrent validity to test the validity of the … performance measure score.

conducted this test using all participants in the patient dataset described above (n=2,552).

- Performance scores were fielded alongside previously validated patient reported PRO-PMs in a single survey of patient participants:
- Patient Enablement Instrument (PEI)
- What Matters Index (WMI)' [p15]

Panel Member #6: Interviews with 10 patients, pilot testing of online participants, and a validation survey in 285 clinicians were all performed. Correlations of the measure with two other validated instruments, PEI and WMI, were also estimated.

Panel Member #7: Face validity was determined with cognitive interviews of 10 patients and surveys of 30 patients and 285 clinicians. No detail is provided on methods for conducting either of those efforts. Detail on how empirical validation analyses were performed (linking the PCPCM PRO with the What Matters Index and the Patient Enablement Index) , especially for the Patient Point of Care Group nesting patients within practice was not provided and would be necessary to evaluate the adequacy of validation efforts.

Panel Member #8: Item level validity focused on content validity including patient cognitive interview and pilot testing. Additionally, clinicians were also surveyed to determine the meaningfulness of the survey items.

The developer used concurrent validity for performance score validity testing. Results of PCPCM PRO-PM were compared with other two measures. However, this assessment was not conducted at the measure entity level.

Panel Member #9: The authors have done a reasonable job of assessing validity of the survey instrument, although the underlying concept of "patient-centered" is not carefully defined or included explicitly at

every step of the validation analysis. At some steps, for example, patients or clinicians are asked if the questions are "meaningful", but that is not the same as asking whether the questions clearly reflect the concept of "patient-centered". The survey does seem to address some core concepts of "patient-centeredness", but the focus of the measure on that one concept does not seem to have been formally tested during validity testing**.**

As far as I can tell, there is no conceptual relationship between the concepts tested in the What Matters Index and the survey under consideration here. Any relationships found do not support the validity of the proposed survey measure. The Patient Enablement Index is only slightly more conceptually related, as it assesses the results of a primary care visit on six elements of patients' understanding of illness and ability to manage illness. Any relationship between the proposed survey and the PEI gives only very weak evidence about validity, even if the statistical relationship is significant. In fact, a strong correlation between the two surveys suggests that both are tapping into some general, broad feeling of satisfaction and NOT into some more precise concept of patient-centeredness. However, given that the developers claim that the measure is " a broad scope assessment of primary care", they cannot be held to a high standard of precision for focusing the measure on a narrower concept like patient-centeredness.

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

Submission document: Testing attachment, section 2b2.3

Panel Member #1: Appropriate

Panel Member #3: The cognitive interviews led to refinements in question wording. Pilot testing indicted that most responders did not find questions difficult to answer and said they were relevant to their experience. 84% of clinicians surveyed (239/285) reported that it would be meaningful to receive patients' feedback on items included in the instrument.

60% of patients reported that their doctor's care would be improved by receiving their response to the instrument.

Panel Member #4: I am not familiar with the scoring for the Patient Enablement Instrument (PEI) (for which, via the internet, I found that a score that is higher, is better),[1] and the What Matters Index (WMI) (for which, via the internet, I found that a score that is lower, is better). However, given the information gleaned from the internet, I am ok with these results. However, it would have been nice to have the information I needed in the application.

Validation Item	**Patient Online Group** (N=2,229): **N**	**Patient Online Group** (N=2,229): **Mean Scores**	**Patient Online Group** (N=2,229): **S.D.**	**Patient Online Group** (N=2,229): **P**	**Patient Point of Care Group** (N=323): **N**	**Patient Point of Care Group** (N=323): **Mean Scores**	**Patient Point of Care Group** (N=323): **S.D.**	**Patient Point of Care Group** (N=323): **P**
What Matters Index: 0	315	3.23	.56	*	106	3.62	.39	*
What Matters Index: 1	348	2.73	.87	0.0001	57	3.47	.49	0.08
What Matters Index: ≥2	396	2.59	.82	*	37	3.63	.52	*

Validation Item	Patient Online Group (N=2,229): N	Patient Online Group (N=2,229): Mean Scores	Patient Online Group (N=2,229): S.D.	Patient Online Group (N=2,229): P	Patient Point of Care Group (N=323): N	Patient Point of Care Group (N=323): Mean Scores	Patient Point of Care Group (N=323): S.D.	Patient Point of Care Group (N=323): P
Patient Enablement Index: 0 1-5	488 543	1.94 2.56	.61 .57	0.0001	23	3.04 3.25	.58	0.0001
Patient Enablement Index: 6-11	764	3.15	.49	*	81	3.50	.44	*
Patient Enablement Index: 12	386	3.67	.39	*	81	3.78	.30	*

*cell intentionally left blank

Panel Member #5: Data element testing Concerning that 24 of 30 people said "yes" to the question of "Was this item relevant to their primary care experience?". In other words, 20% of the people said the item wasn't relevant to their primary care experience.

Measure score

The measure score was found to generally have strong concurrent validity.

Panel Member #6: Majorities of patients and clinicians who participated in pilot testing and a validation survey, respectively, indicated that the survey was relevant and important. The measure was positively correlated with other instruments, although strength of association was greater in the online setting.

Panel Member #7: There is insufficient detail to interpret results at the practice level.

Panel Member #8: Instrument validity testing is reasonable but performance score validity testing at measure entity level is missing.

Panel Member #9: The results show that the survey measure is a reasonable measure of some broad satisfaction with, or assessment of, primary care. The phase "patient-centered" in the measure title is something of a misnomer then, as the measure is not really about patient-centeredness per se, vs. other possible dimensions of primary care. As a general measure of positive feelings about primary care, the measure is reasonably valid. As was the case for reliability, there is no assessment of measure score validity.

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.**

Panel Member #1: Test results.

Panel Member #4: No additional concerns.

Panel Member #5: Generally, the validity testing results were high. The rating of "moderate" is due to several issues:

[a] No acknowledgement of the exclusion of less than 8 completed items. Thus, no rationale nor analysis of this exclusion [Q15]

[b] No testing regarding the potential use of the mode of completing the instrument as a risk variable. [Q16e]

[c] Regarding data element testing: 20% of the people said the item wasn't relevant to their primary care experience. [Q22]

Panel Member #6: The measure appears to have reasonable validity, primarily on the basis of correlations with existing instruments.

Panel Member #7: The application has insufficient detail on statistical methods to provide an appropriate assessment

Panel Member #8: Performance score validity testing at measure entity level is missing.

Panel Member #9: The evidence of validity at the data element level (i.e., individual patient survey) is reasonable, once it is established that the measure is about general positive feelings about primary care and not about patient-centeredness as a specific aspect of primary care. There is no testing of validity at the measure score level

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.**

Panel Member #4: I have some basic concerns with the instrument itself… Specifically, the item anchors do not seem to be consistent (Definitely, Mostly, Somewhat, Not at all). It seems to me that it would be better to replace Definitely with Always for consistency and because I'm not convinced that all patients will make sense of this scale as is. Also, the instrument only references doctors. While I see that the instrument has been tested in a sample that included NPs and PAs, I worry that patients will not understand to include these mid-level providers as doctors.

Panel Member #5: While I noted a couple of issues, none are substantial enough to call for further discussion.

Panel Member #9: This is a perfect example of the question of whether a measure should receive NQF endorsement if there is no evidence whatsoever about its reliability and validity at the measure score level (i.e., comparison of performance among entities like individual clinicians or practices). There is a survey here that assesses broad positive feelings about primary care, and in that limited context, reliability and validity are adequate. The measure developers state clearly that the measure is to be used for self-assessment at the clinician or practice levels – they make no claim for the ability of the measure to compare clinicians or practices to one another. Is such a measure suitable for NQF endorsement?

Developer Submission

Additional evaluations and submission materials attachments…

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

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

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

NQF_evidence_attachment_11_7_2020.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.

No

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 3568

Measure Title: Person-Centered Primary Care Measure PRO-PM

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

Date of Submission: **11/8/2020**

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

Outcome

☐ Outcome:

☒ Patient-reported outcome (PRO): Person-Centered Primary Care Measure PRO-PM

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*):

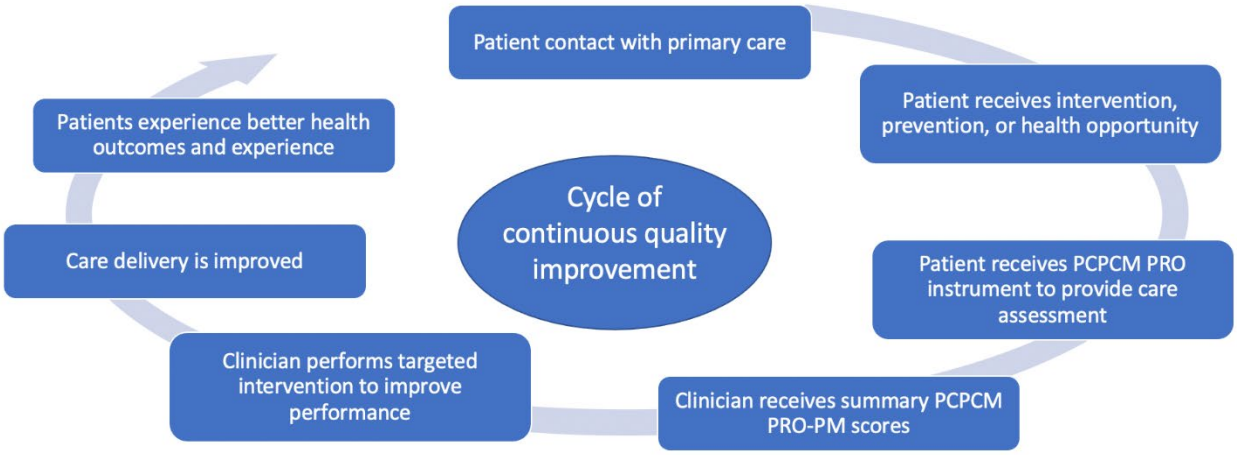
☐ Process:

☐ 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.



Any patient active in a primary care practice will receive the Person-Centered Primary Care Measure (PCPCM) instrument once a year. Patients are defined as active if they are identified by their primary care practice as having had at least once documented interaction during the 12 months preceding their use of the PCPCM.

Active patients interact with a practice during the course of a year. That interaction can include, but is not limited to, in person visits, telehealth visits, or responses to needs for medication refills, questions answered, or health forms filled out. Active patients will receive an invitation to fill out the PCPCM instrument either by email, web portal, or mail, as determined by the practice. Patients are invited to fill out the instrument during the month of their birth. Clinicians receive regular feedback based on their PCPCM PRO-PM score. Clinicians receive their scores in conjunction with national and regional or local benchmarks, as deemed appropriate by the practice. Using the PCPCM PRO-PM score, and their score on individual items, clinicians are able to follow actionable steps to improve their performance relative to such things as care coordination, access, comprehensiveness, patient health goals, etc. As adjustments are made to their behaviors and care delivery based on this feedback, overall care can be improved. Whether care has improved or not can be determined by assessing changes in PCPCM PRO-PM scores over time, or changes in other health outcomes over time.

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.)

A core component of the PCPCM and central to constructing the instrument itself was the value and meaningfulness of the instrument to patients. In essence, we enlisted the help of hundreds of patients as co-developers of this instrument.

The PCPCM PRO-PM is calculated using the PCPCM PRO instrument. The 11 items of the PCPCM PRO are the critical data elements of the PCPCM PRO-PM

Background. The PCPCM PRO was created through a multi-step approach. First, we conducted a survey to identify a preliminary set of quality indicator areas of greatest importance to primary care clinicians (n=525), patients (n=412), and employers (n=85). Qualitative content analysis was used to identify 18 quality indicator areas among the ~10,000 written comments received.[1] Next, we hosted a national conference (http://www.starfieldsummit.com/starfield3) in which we engaged a diverse range of 70 national leaders in primary care quality and measurement, including patients and patient advocates.[2] Qualitative analyses of conference data, combined with previously conducted literature reviews, resulted in a reduction of quality areas from 18 to 11. A single item was created to reflect each quality area as identified. These 11 items resulted in the first draft of the PCPCM PRO.

1. Etz RS, Gonzalez MM, Brooks EM, Stange KC. Less AND more are needed to assess primary care. J Am Board Fam Med. 2017; 30(1): 13–15.

2. Etz, RS and the Starfield Writing Team. Conference Brief: Framework of PC Measure Domains and Key Elements. 2017. Starfield Summit III: Washington, DC. Accessed January 6, 2020, http://www.starfieldsummit.com/resources3

We used three different analyses to test the validity and meaningfulness of our critical data elements.

Cognitive interviews among patients

We conducted cognitive interviews among 10 patients regarding the 11 items of the PCPCM PRO. For these interviews, we asked participants to read each item aloud and then to explain in their own words what the item is asking. We then asked participants to pick one of the four response options for each item, followed by asking them to share in their own words why they chose that answer. This process allowed us to understand if the items were likely to be interpreted as intended by participants, and whether the answers they chose were providing the type of information we expected.

Responses from patients during the cognitive interviews were compared with the previously written detailed descriptions of each PCPCM PRO item. Nine of 10 patients successfully identified the intended quality indicator area for each item. The language patients used in their description of indicator areas resulted in minor revisions to the language of 3 items. Patient descriptions of their intention when selecting response options for each item were consistent with the development team's intentions regarding the established scale.

Pilot testing of the PCPCM PRO

We fielded the PCPCM PRO to a pilot set of online participants. The invitation to respond to the PCPCM PRO was given to an existing set of patient respondents that had participated in previous ACORN studies. We accepted all responses, stopping data collection at 30. This online pilot fielding included each of the 11 items, as well as questions following each of the 11 items that asked whether the question was hard to understand, whether the question was hard to answer, and whether the question seemed relevant to their primary care experience.

Among the 30 patient respondents, responses to the two questions fielded with the piloting of the PCPCM PRO:

- Was this item hard to answer: 27 said no.
- Was this item relevant to their primary care experience: 24 said yes.

Survey for meaningfulness

Two groups were surveyed to determine if the 11 items of the PCPCM PRO were meaningful to participants: clinicians and patients.

Drawing on a database of clinicians that have participated in previous ACORN studies, we sent an invitation to participate in a quick validation survey of the PCPCM PRO. We accepted all responses within a one-week time period. We received 285 responses. The clinician validation survey presented the 11 items to clinicians and then asked 1) if the item appeared meaningful to them, 2) if they currently collected that information in their practice, and 3) if they would find it meaningful to receive patient responses to these items.

Among the 285 clinicians surveyed, responses to the three questions asked after sharing the 11 PCPCM PRO items:

- Is this item important to your practice: 225 (79%) said yes
- Is this item personally important to you: 242 (85%) said yes
- Would it be meaningful to you to receive your patients' responses to these items: 239 (84%) said yes

Additionally, we fielded the PCPCM PRO among two patient populations.

The first was a patient population of 1,140 among whom we also asked two questions: Is this measure meaningful to you, and, if your doctor had these survey results, would it help your care? 87% of

patients said the survey was meaningful and 60% of patients said receipt of their answers would help with their care.

The second was a patient population of 1,344 (March 31, 2020) in which we asked three questions after each item in the PCPCM PRO: 1) is this item meaningful to you, 2) would you have answered this question differently before the COVID-19 pandemic, and 3) please briefly explain your answer. Every item in the PCPCM PRO was identified as meaningful by at least 80% of patients responding. Additionally, the majority of patients said their answers were not different as a result of the pandemic. Open text responses typically pointed to the importance of content area addressed by the item.

****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.

Each item of the PCPCM is actionable. For example, if a clinician scores poorly on the item "Over time, this practice helps me to meet my goals", clinicians can use that feedback to make sure discussion of patient goals is part of the clinical encounter. The PCPCM is freely available online, as is some advice regarding quality improvement activities and action reflection items to assist with quality improvement efforts.[12] Examples related to each of the PCPCM items include:

PCPCM Item	Example Actions
The practice makes it easy for me to get care	• Altering scheduling options, availability, or who does the scheduling • Providing options for asynchronous communication or telehealth visits
This practice is able to provide most of my care	• Schedule longer visits for more complicated problems or patients so that you can provide more of the care rather than referring it out • Refer in-house to staff or clinicians with specialized expertise or interests
In care for me, my doctor considers all of the factors that affect my health	• Consider starting visits by asking patients what matters to them for this visit • Ask patients, "What one thing would you like someone taking care of you to know?" and add this to the medical record in a consistent and easy-to-see place
My practice coordinates the care I get from multiple places	• The medical assistant asks and documents any care received elsewhere since the last visit • When doing medication reconciliation, ask about care received elsewhere
My doctor or practice knows me as a person	• Try to talk about at least one non-medical item during each visit • Ask the patients what matters to them • Link recommended treatments to what gives meaning in the patient's life
My doctor and I have been through a lot together	• Do phone follow up after hospital discharges • Consider other ways you might connect with patients' important health and life events
My doctor or practice stands up for me	• Let patients know when you spend time doing prior authorizations • Discuss options regarding medications with patients to show them you are aware of patient costs and taking that into account

PCPCM Item	Example Actions
The care I get takes into account knowledge of my family	• Do a quick and dirty family tree and update it periodically – try to find a consistent place in the EHR to keep this information • More routinely ask about the family as a resource or the impact of the patient's illness on the family
The care I get in this practice is informed by knowledge of my community	• Participate in community events and include that in posters or on the practice website • Ask about the patient's neighborhood
Over time, this practice helps me to meet my goals	• Frame care plans around patients' goals or what matters to them • Do HOPE notes: https://drwaynejonas.com/wp-content/uploads/2018/01/HOPENoteQuestions_WEB.pdf
Over time, my practice helps me to stay healthy	• Look for teachable moments when the patient is open to a health behavior change • Standing orders for immunizations

Each clinician or practice can create quality improvement activities best suited to their context.

There is a clear and large body of evidence that demonstrates the strong connection between patient experience of care and traditional health care outcomes, such as improved intermediate outcomes, greater adherence to recommended treatment, and reduced use of health care services. Two systematic reviews – one in the US and one in the UK – provide clear evidence to that effect.[1-3]

The items within the PCPCM PRO instrument used to calculate the PCPCM PRO-PM performance measure are also each individually supported by empirical resource as having a strong effect on desirable health outcomes. For instance, continuity of care is associated with improved intermediate outcomes and reductions in cost of care.[4,5] Comprehensiveness has been shown to be associated with lower hospitalization rates, greater use of preventive services, greater adherence to recommended treatment and reduction of burnout among clinicians.[3,6-10]

The ability to assess those aspects of primary care that uniquely contribute to primary care's proven ability to improve patient health outcomes and experience while reducing health burden and costs warrants a national measure. A 2014 review of measures used to assess primary care shows many aspects of care remain unassessed by current measures. The PCPCM allows for patient reported assessment of those aspects of primary care identified by patients and clinicians as most important.[11]

[1] Anhang Price R, Elliott MN, Zaslavsky AM, Hays RD, Lehrman WG, Rybowski L, Edgman-Levitan S, Cleary PD. Examining the role of patient experience surveys in measuring health care quality. Med Care Res Rev. 2014 Oct;71(5):522-54.

[2] Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. BMJ Open. 2013 Jan 3;3(1):e001570.

[3] Stange KC, Etz RS, Gullett H, Sweeney SA, Miller WL, Jaén CR, Crabtree BF, Nutting PA, Glasgow RE. Metrics for assessing improvements in primary health care. Annu Rev Public Health. 2014;35:423-42.

[4] Starfield B, Shi L, Macinko J. Contribution of primary care to health systems and health. Milbank Q. 2005;3:457-502.

[5] Institute of Medicine (U.S.). Division of Health Care Services. Committee on the Future of Primary Care., Donaldson MS. Primary care : America's health in a new era. Washington, D.C.: National Academy Press; 1996.

[6] McWhinney IR, Freeman T. Textbook of family medicine. Oxford ; New York: Oxford University Press; 2009.

[7] Institute of Medicine (U.S.). Committee on Core Metrics for Better Health at Lower Cost, Blumenthal D, Malphrus E, et al. Vital signs : core metrics for health and health care progress. Washington D.C.: National Academies Press; 2015.

[8] Stange KC. The paradox of the parts and the whole in understanding and improving general practice. Int J Qual Health Care. 2002 Aug;4:267-8.

[9] Starfield B. Is patient-centered care the same as person-focused care? Perm J. 2011 Spring;2:63-9.

[10] Soler JK, Okkes I, Wood M, et al. The coming of age of ICPC: celebrating the 21st birthday of the International Classification of Primary Care. Fam Pract. 2008 Aug;4:312-7.

[11] Etz RS, Zyzanski SJ, Gonzalez MM, Reves SR, O'Neal JP, Stange KC. A New Comprehensive Measure of High-Value Aspects of Primary Care. Ann Fam Med. 2019 May;17(3):221-230.

[12] www.green-center.org/pcpcm

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.

Per instruction from NQF staff, this item is not required for a PRO-PM. We have therefore skipped responding item 1a.3 and all sub-items of 1a.3.

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

Systematic Review	Evidence
Source of Systematic Review: • Title • Author • Date • Citation, including page number • URL	*
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.	*
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	*
Provide all other grades and definitions from the evidence grading system	*

Systematic Review	Evidence
Grade assigned to the **recommendation** with definition of the grade	*
Provide all other grades and definitions from the recommendation grading system	*
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	*
Estimates of benefit and consistency across studies	*
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	*

*cell intentionally left blank

_____

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.

Per instruction from NQF staff, this item is not required for a PRO-PM. We have therefore skipped responding item 1a.4 and all sub-items of 1a.4.

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

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

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

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.

Rationale

The PCPCM PRO-PM fulfills the call from the Institute of Medicine and from CMS to create a stakeholder informed, meaningful measure that is an assessment of quality, low burden for implementation and collection, and provides adequate ability to compare performance across clinicians and practices while providing great face validity, transparency and actionable information.[1,2] The PCPCM PRO-PM does just that. It is unusual in its combination of robust internal consistency together with breadth and brevity. Its combination of parsimony - with a single item for each of 11 diverse primary care components - and conceptual coherence - exemplified

by the fact that all 11 items load onto a single factor - is the result of an unusually broad and deep amount of preparatory work grounded in diverse stakeholder engagement.[3] This stakeholder engagement enabled the development of meaningful measure items and is the reason why the PCPCM PRO-PM covers 4 of the 8 "cross cutting connections" in the CMS Meaningful Measures Framework (identified as patient-centered and meaningful to patients; fulfill requirements in programs' statues; minimize level of burden for providers; significant opportunity for improvement).2 The PCPCM PRO-PM also addresses a critical quality measure gap as identified by the MACRA Measure Development Plan Technical Expert Panel, of which Dr. Etz – the developer of the PCPCM PRO-PM – was a part.[4]

Benefits and improvements in quality envisioned by use of this measure

The PCPCM PRO-PM is a performance measure that uses the PCPCM PRO instrument. The performance measure is used to assess quality of primary care from patient perspective, comparing an individual clinician's performance to national benchmarks and to other clinicians in their practice, locally, and regionally, as data availability allow.

The performance measure is calculated based on one year of data collection. PCPCM PRO instruments are received monthly by the practice. Active patients (defined as having had contact with the practice in the 12 months preceding their birth month) receive the PCPCM PRO instrument during the month of their birth. Receiving data on a monthly basis allows clinicians to receive feedback on their performance in between annual reporting periods. Such interim feedback enables constant attention and opportunities for correction of performance during any given performance year.

When validating the PCPCM PRO, we tested its concurrent validity with two existing and validated instruments: the Patient Enablement Instrument (PEI)5 and the What Matters Index (WMI).6,7 The PCPCM PRO is well correlated to both the PEI – an assessment of patient self-management – and the WMI – validated to correlate both retrospectively and prospectively with cost and utilization of services. Our comparative analyses used t tests and analysis of variance for continuous variables and c2 for categorical variables (see Table 1 below). PCPCM PRO-PM scores were strongly and positively associated with the WMI and PEI (both p = .0001). It is therefore envisioned that improvements in PCPCM PRO-PM scores, as facilitated by QI activities, will result in both improved patient self-management and reduced cost and utilization of services. See graph in Appendix A.1 (page 3)

The PCPCM PRO-PM is currently being piloted in several settings but is not yet widely implemented. The PCPCM PRO-PM is being piloted by health systems in Colorado, Missouri, Ohio, and Richmond, and within PRIME – a national primary care Qualified Clinical Data Registry (QCDR), hosted by the American Board of Family Medicine. The PCPCM PRO-PM has been endorsed by CMS for use as a QCDR measure in the 2020 MIPS reporting period and is being used by a subset of PRIME members for that purpose. On maintenance review of the PCPCM PRO-PM, we expect to have more information on implementation and scores.

1. Stange KC, Etz RS, Gullett H, et al. Metrics For Assessing Improvements In Primary Health Care. Annual review of public health. 2014;35:423-442.
2. In: Blumenthal D, Malphrus E, McGinnis JM, eds. Vital Signs: Core Metrics for Health and Health Care Progress. Washington (DC)2015.
3. Etz RS, Zyzanski SJ, Gonzalez MM, Reves SR, O´Neal JP, Stange KC. A New Comprehensive Measure of High-Value Aspects of Primary Care. Ann Fam Med. 2019 May;17(3):221-230.
4. Meaningful Measures Framework of CMS. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/CMS-Quality-Strategy Accessed June 27, 2020.

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.*

During validation reliability testing, variation of PCPCM PRO-PM scores, which illustrate a difference of moderate to large effect size among clinicians in our validation tests, demonstrated a performance gap and opportunities for improvement. Among 6 practices, there were significant differences (p=0.004) in PCPCM PRO-PM scores with a moderate effect size (at least .5 standard deviation).See graph in Appendix A.1 (page 3-4)

Among 16 clinicians, there were significant differences (p=0.0001) in PCPCM PRO-PM scores with a large effect size (>1.0).See graph in Appendix A.1 (page 3-4)

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.

While empirical studies of the PCPCM PRO-PM and its connection to improved outcomes as resulting from quality improvement efforts is limited, its strong concurrent validity with known and validated measures support the benefits likely to result from its use. In addition, existing literature supports both the need for clinician and practice improvement in the areas covered by the PCPCM PRO-PM and the absence of current measures able to support that work.

Current measure sets presume that quality primary care is the sum of quality measures for individual diseases and health screening. Value-based payments to primary care physicians frequently employ measures that are not aligned or recognize the higher-level integrating, personalizing, and prioritizing functions of primary care and the needs of patients, communities or health care systems.[1,2] These measures are then tied to financial incentives which drive behavior to maximize these rudimentary measures. Driving clinicians' behavior toward low-value measures produces burnout and diminishes the value of primary care for people and populations.[3-5] The 2001 Crossing the Quality Chasm report, issued by the Institute of Medicine, revealed a worrying gap in performance between what we know to be good quality care, and care as measured by current forms of assessment.[6] More recent studies of the basic aspects of primary care, such as continuity and comprehensiveness, continue to demonstrate weaknesses in existing attempts both to measure primary care effectively, and to reach measurement targets associated with high quality.[7-9] In addition, measures that focus on single aspects of primary care may inadvertently cause harm to other key aspects of primary care. For instance, studies have shown that quality improvement activities focused on access to care may cause reduced performance in continuity – both of which are critical to overall primary care quality.[10-11]

The PCPCM PRO-PM fulfills the call from the Institute of Medicine and from CMS to create a stakeholder informed, meaningful measure that is an assessment of quality, low burden for implementation and collection, and provides adequate ability to compare performance across clinicians and practices while providing great face validity, transparency and actionable information.[12,13]

1. Etz RS, Gonzalez MM, Brooks EM, Stange KC. Less AND More Are Needed to Assess Primary Care. J Am Board Fam Med. 2017;30(1):13-15.
2. Stange KC, Etz RS, Gullett H, et al. Metrics for assessing improvements in primary health care. Annu Rev Public Health. 2014; 35: 423–442.
3. Berenson RA. If you can't measure performance, can you improve it? Jama. 2016;315(7):645-646.
4. McWilliams J. Michael. Professionalism Revealed: Rethinking Quality Improvement in the Wake of a Pandemic. NEJM Catalyst. 1(5). doi:10.1056/CAT.20.0226
5. Phillips RL. The Built Environment for Professionalism. J Am Board Fam Med. 2020;33(Supplement):S57.
6. Institute of Medicine (U.S.). Committee on Quality of Health Care in America. Crossing the quality chasm : a new health system for the 21st century. Washington, D.C.: National Academy Press; 2001.
7. Bazemore A, Neale AV, Lupo P, Seehusen D. Advancing the Science of Implementation in Primary Health Care. J Am Board Fam Med. 2018 May-Jun;31(3):307-311.

8. Gillam SJ, Siriwardena AN, Steel N. Pay-for-performance in the United Kingdom: impact of the quality and outcomes framework: a systematic review. Ann Fam Med. 2012 Sep-Oct;5:461-8.

9. Berwick DM. The Moral Determinants of Health. Jama. 2020 Jun 12. PMID: 32530455.

10. Casalino LP, Khullar D. Value-Based Purchasing and Physician Professionalism. JAMA. 2019;322(17):1647.

11. Campbell SM, Reeves D, Kontopantelis E, Sibbald B, Roland M. Effects of Pay for Performance on the Quality of Primary Care in England. New England Journal of Medicine. 2009;361(4):368-378.

12. In: Blumenthal D, Malphrus E, McGinnis JM, eds. Vital Signs: Core Metrics for Health and Health Care Progress. Washington (DC)2015.

13. Meaningful Measures Framework of CMS. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/CMS-Quality-Strategy Accessed June 27, 2020.

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.*

Evidence to date shows no correlation between self-defined minority status and patient assessment of primary care, using the PCPCM.1 Our findings when validating the PCPCM PRO-PM show neither a positive nor negative association between PCPCM PRO-PM scores and patient self-defined minority status. While unexpected, this does not call into question the validity of patient assessments using the PCPCM PRO-PM.

The existence of health disparities within the US health care system, and in primary care settings, is without question. High quality primary care has previously been shown to mitigate the negative impact associated with negative social drivers of health. 2,3 The PCPCM PRO-PM works to assess the quality of primary care, not the status of population health. The two are associated, but not mutually dependent. For example, Sweden is known to have one of the best population health outcomes, globally, and yet is also known for having a poor primary care system. In a recent study conducted by our team to validate the PCPCM PRO-PM in 28 languages and 34 country settings, including Sweden, the PCPCM PRO-PM was able to detect Sweden's lower performing primary care.[4] It supports the ability of the PCPCM PRO-PM to appropriately assess primary care in spite of social drivers that may disproportionately impact the health of minority populations.

During the maintenance of endorsement period, we intend to focus research efforts to better understand the relationship between the PCPCM PRO-PM and the health outcomes of minority populations so that we may understand how to best use the PCPCM PRO-PM to combat health disparities within the US.

In our validation study, we published findings regarding the PCPCM PRO-PM and known sub-populations[1]:

"Based on prior research, and clinical and research experience, we hypothesized there would be a positive association between a higher PCPCM score and patients of greater age, patients receiving most of their care from a single physician, the more years a patient knew the physician, a higher What Matters Index score, and a higher Patient Enablement Index score. In contrast, a negative association was hypothesized for patients with minority status and the type of device used to administer the questions was anticipated to be neutral."

"These associations are in the hypothesized direction … except for minority status which was nonsignificant. Moreover, rank-ordered associations were observed for income, whether the survey was hard to complete, and whether respondents felt that clinician awareness of their PCPCM responses would positively inform their care. No associations were observed for region, mode of administration, or sex." See graph in Appendix A.1 (page 4-5)

1. Etz RS, Zyzanski SJ, Gonzalez MM, Reves SR, O´Neal JP, Stange KC. A New Comprehensive Measure of High-Value Aspects of Primary Care. Ann Fam Med. 2019 May;17(3):221-230.
2. Starfield B. Primary care and equity in health: the importance to effectiveness and equity of responsiveness to people´s needs. Humanity and Society. 2009;33(1/2):56-73.
3. Starfield B. Equity in health. J Epidemiol Community Health. 2002;56(7):483-484.
4. Zyzanski SJ, Gonzalez MM, O'Neal JP, Etz RS, Reves SR, Stange KC. Person-Centered Primary Care in 35 OECD Countries. Ann Fam Med. 2020;(in press).

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

Equity-oriented health care is associated with better patient outcomes,[2] and primary care is associated with health and health care equity.[3-13] The PCPCM measures the mechanisms by which primary care results in greater health and health care equity, and thus will be very useful in efforts to improve equity.

1. Ford-Gilboe M, Wathen CN, Varcoe C, et al. How Equity-Oriented Health Care Affects Health: Key Mechanisms and Implications for Primary Health Care Practice and Policy. Milbank Q. 2018;96(4):635-671.
2. Homa L, Rose J, Hovmand PS, et al. A participatory model of the paradox of primary care. Ann Fam Med. 2015;13(5):456-465.
3. Kringos DS, Boerma WG, Hutchinson A, van der Zee J, Groenewegen PP. The breadth of primary care: a systematic literature review of its core dimensions. BMC Health Serv Res. 2010;10:65.
4. Starfield B. Primary care and equity in health: the importance to effectiveness and equity of responsiveness to people´s needs. Humanity and Society. 2009;33(1/2):56-73.
5. Starfield B. Equity in health. J Epidemiol Community Health. 2002;56(7):483-484.
6. Macinko JA, Starfield B. Annotated bibliography on equity in health, 1980-2001. Int J Equity Health. 2002;1(1):1.
7. Starfield B, Shi L, Grover A, Macinko J. The effects of specialist supply on populations´ health: assessing the evidence. Health Aff (Millwood). 2005;Suppl Web Exclusives:W5-97-W95-107.
8. Starfield B, Shi L, Macinko J. Contribution of primary care to health systems and health. The Milbank Quarterly. 2005;83(3):457-502.
9. Williams RL, Flocke SA, Stange KC. Race and preventive services delivery among black patients and white patients seen in primary care. Med Care. 2001;39(11):1260-1267.
10. Haggerty JL, Reid RJ, Freeman GK, Starfield BH, Adair CE, McKendry R. Continuity of care: a multidisciplinary review. Vol 3272003.
11. Starfield B. Primary Care: Balancing Health Needs, Services, and Technology. Rev. ed. ed. New York, NY: Oxford University Press; 1998.
12. Starfield B. Primary Care: Concept, Evaluation, and Policy. New York, NY: Oxford University Press; 1992.

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.)*

This measure does not yet have a measure-specific web page yet.

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.

Attachment**:** PCPCM_Instrument.docx

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.

Patient

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

No

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

N/A

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

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

The PCPCM PRO-PM allows all patients to report their assessment of the quality of primary care received through responses to PCPCM PRO instrument.

The target population is all active patients in a practice during the performance reporting period. A patient is defined as active if the patient has had a documented interaction with the practice within 12 months of the patient's birth month. The PCPCM PRO is the same for all patients, regardless of age. Because the PCPCM PRO applies to all patients and is not particular to a clinical encounter, it is administered once a year to each patient during their birth month.

The target population is defined the same, regardless of unit of analysis (clinician or practice).

The numerator is the sum of all PCPCM PRO scores for active patients.

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).

All patients receive the PCPCM PRO instrument once a year during their birth month. In any given reporting period, any returned PCPCM PRO instruments that do not have at least 8 of the 11 PCPCM PRO items completed are not included in calculations.

Before calculating the PCPCM PRO total scores, it is necessary to calculate the PCPCM PRO item scores. For PCPCM PRO item scores, the numerator is the sum of all received patient responses eligible for calculation. The value for patient responses is based on the scale of 4 (Definitely) to 1 (Not At All), as described above.

The time frame for PCPCM PRO-PM scores is 12 months.

This process is same, regardless of unit of analysis (clinician or practice).

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

The target population for the denominator is the same as for the numerator.

The denominator is the total number of complete PCPCM PRO instruments received in the reporting period. A completed PRO instrument is defined as a PRO instrument for which the patient has responded to at least 8 of 11 items.

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).

The target population is all active patients in a practice during the performance reporting period. A patient is defined as active if the patient has had a documented interaction with the practice within 12 months of their birth month. The PCPCM PRO is the same for all patients, regardless of age. Because the PCPCM PRO applies to all patients and is not particular to a clinical encounter, it is administered once a year to each patient during their birth month.

The target population is defined the same, regardless of unit of analysis (clinician or practice).

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

None.

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

N/A

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

No stratification of measure results is required.

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:

Continuous variable, e.g. average

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.)*

Scoring for the PCPCM PRO-PM is completed through a simple 4 step process using the PCPCM PRO to assess the broad scope of primary care from a patient's perspective.

Step One: Exclude incomplete patient responses.

Any PCPCM PRO instrument for which a patient failed to answer at least 8 of the 11 items is excluded from calculations.

Step Two: Calculate PCPCM PRO item specific mean scores.

Patients choose one of four response options for each item in the PCPCM PRO instrument. In scoring the PCPCM PRO, the first step requires determining an item mean score for each of the 11 items. Since the instrument scale is word based – Definitely, Mostly, Somewhat, Not At All – each response option must be assigned a value. Values are assigned as follows: Definitely = 4, Mostly = 3, Somewhat = 2, Not At All = 1.

Calculating the mean score for each item then requires looking across all PCPCM PRO instruments received for the entity being assessed during the analysis period. For example, if the entity is a clinician, then all completed (see Step One) PCPCM PRO instruments collected for that clinician are included in the calculation. If the entity is a practice, then all PCPCM PRO instruments collected for that practice are included in the analysis.

An entity's score for each PCPCM PRO item is calculated as a mean, i.e., the summary of all responses across PCPCM PRO instruments received for the entity, divided by the number of instruments received. This process leads to 11 item specific PCPCM PRO scores. Means should be reported to two decimal points.

Step Three: Calculate the PCPCM PRO total score.

The PCPCM PRO total score for the entity is calculated by determining the mean of the 11 scored PRO items. This is done by adding the mean scores of all 11 PRO items and then dividing by 11. PRO means should be reported to two decimal points.

Step Four: Converting PCPCM PRO total scores and to PCPCM PRO-PM performance score.

In order to use the PCPCM PRO as a performance measure for reporting, the 4 point PCPCM PRO scale must be converted to a 0-100 performance scale. To do this, the PCPCM PRO total score for an entity, as calculated in Step Three, is divided by 4 and then multiplied by 100.

Thus, a PCPCM PRO total score of 2.78 (based on a scale of 1-4) becomes a PCPCM PRO-PM performance score of 69.5 (on a scale of 0-100).

The monthly data collection allows for assessed entities to receive regular feedback during the course of the year. However, PCPCM PRO-PM performance scores are calculated based on quality reporting program requirements or a 12-month time frame.

There is no stratification required with the PCPCM.

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.

The PCPCM PRO-PM is calculated on all PCPCM PRO responses received.

Patients unable to respond to the PCPCM PRO because of cognitive impairment may have their responses turned in by proxy. Such proxy would be a non-practice-based caregiver with knowledge of the patient's interaction with the practice.

Patients under the age of 13 should have their PCPCM PRO responses turned in by proxy. Such proxy would be their parent, guardian, or other adult responsible for their care and with knowledge of the patient's interaction with the practice.

Patients 13 years of age through 17 years of age may be given the option to respond to the PCPCM PRO on their own, or to have their responses filled by proxy by the adult responsible for their care.

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.

To inform quality improvement activities, a minimum of 10 completed PCPCM PRO instruments must be received (8 of 11 items must have a response).

To provide a PCPCM PRO-PM score, a minimum of 30 completed PCPCM PRO instruments must be received (8 of 11 items must have a response).

If a practice or system has more four or more full time clinicians for whom PCPCM PROs are being collected, quality or performance scores for the practice or system will require a minimum of 120 completed responses.

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

If other, please describe in S.18.

Instrument-Based 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 PCPCM PRO-PM performance data are collected using the PCPCM PRO instrument. The PCPCM PRO is an 11-item patient reported instrument. The measure has been tested and validated using the following methods for administration:

- Paper-based delivery, point of care. The paper instrument can be mailed to active patients (defined as having a documented encounter with the practices within 12 months prior to the patient's birth month). Data entry will then be required. Data may be entered into a simple Excel-type document for data management and scoring. Point of care instrument use should not be used for performance measure purposes as these responses will skew positive.
- Asynchronous delivery, electronic administration and submission. Patients active in a practice (defined as having a documented encounter with the practices within 12 months prior to the patient's birth month) can receive the PCPCM PRO via email, patient portal, or email invitation with a unique link, during the month of their birth. Triggering an invitation to complete the PCPCM PRO immediately following a clinical encounter should not be used for performance measure purposes as these responses will skew positive.

The PCPCM PRO instrument is available and validated in the following languages: simple Chinese, Czech, Danish, Dutch, English (British), English (American), Estonian, Finnish, French (European), German, German (Swiss), Greek, Hebrew, Hungarian, Icelandic, Italian, Japanese, Korean, Latvian, Lithuanian, Luxembourgian,

Norwegian, Polish, Portuguese (European), Slovakian, Slovenian, Spanish (European), Spanish (Latin American), Swedish, and Turkish. The manuscript supporting the validation of the PCPCM PRO in these languages has been accepted by the Annals of Family Medicine but is not yet been published.

Table 1: The Person-Centered Primary Care Measure (PCPCM) Patient Reported Outcome (PRO) Instrument

HOW WOULD YOU ASSESS YOUR PRIMARY CARE EXPERIENCE?

The practice makes it easy for me to get care.

Definitely Mostly Somewhat Not at all

This practice is able to provide most of my care.

Definitely Mostly Somewhat Not at all

In caring for me, my doctor considers all of the factors that affect my health.

Definitely Mostly Somewhat Not at all

My practice coordinates the care I get from multiple places.

Definitely Mostly Somewhat Not at all

My doctor or practice knows me as a person.

Definitely Mostly Somewhat Not at all

My doctor and I have been through a lot together.

Definitely Mostly Somewhat Not at all

My doctor or practice stands up for me.

Definitely Mostly Somewhat Not at all

The care I get takes into account knowledge of my family.

Definitely Mostly Somewhat Not at all

The care I get in this practice is informed by knowledge of my community.

Definitely Mostly Somewhat Not at all

Over time, this practice helps me to meet my goals.

Definitely Mostly Somewhat Not at all

Over time, my practice helps me to stay healthy.

Definitely Mostly Somewhat Not at all

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)*

Available in attached appendix at A.1

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

Clinician : Group/Practice, Clinician : Individual

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

Outpatient Services

If other:

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

N/A

2. Validity – See attached Measure Testing Submission Form

PCPCM_testing_nqf_revised__11_7_2020.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.

No

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.

No

2.3 For maintenance of endorsement

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

No - This measure is not risk-adjusted

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

Measure Number (*if previously endorsed*)**:** N/A
Measure Title: Person-Centered Primary Care Measure PRO-PM
Date of Submission: **8/3/2020**

Type of Measure:

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

*cell intentionally left blank

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

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

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

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

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*).

Two existing datasets were used to test the Person-Centered Primary Care Measure (PCPCM) PRO-PM: 1) the patient dataset and 2) the clinician and group practice dataset.

Patient dataset. The patient dataset included 2,552 patient responses to the PCPCM PRO instrument used to calculate the PCPCM PRO-PM Performance Score. There are two subgroups in the patient dataset: the patient online group and the patient point of care group. These are described in more detail in section 1.6 below.

Results from the patient dataset were used to validate the PCPCM PRO. That work was published in the *Annals of Family Medicine* in May 2019. The article and full description of all data samples used in that validation can be found as supplemental information: http://www.annfammed.org/content/17/3/221/suppl/DC1

Clinician and group practice dataset. The clinician and practice dataset includes 1,205 patient responses to the PCPCM PRO. Each patient response is attributable to one of seven practices and one of 52 clinicians.

1.3. What are the dates of the data used in testing? The patient dataset was collected January 2018 through April 2018. The clinician and practice dataset was collected April 2020 through May 2020.

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

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

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

Both group practices and individual clinicians were measured entities in the test and analysis of the PCPCM PRO-PM.

Group practices. Data were collected among seven practices. These practices are associated with the Department of Family Medicine and the Department of Internal Medicine at Virginia Commonwealth University in Richmond, Virginia. All family medicine and internal medicine practices at VCU were offered participation. None declined. Five of the seven practices are located in the community. Number of patient responses to the PCPCM PRO per practice were sufficient during the limited data collection period (four weeks) to allow for practice level measurement at six of the seven practices. Sufficient sample means at least 30 responses per clinician in the practice or at least a minimum practice sample of 50 patient responses, if the practice has more than 1 full-time clinician. This is explained in further detail later in this application. Below is a table of the participating group practice characteristics.

Table 1.5.1 Group Practice Characteristics

ID	Size	Specialty	Rural (Y/N)
P1	Small, 2-5 clinicians	Pediatrics	N
P2	Small, 2-5 clinicians	Family Medicine	N
P3	Small, 2-5 clinicians	Family Medicine	Y
P4	Medium, 6-20 clinicians	Family Medicine	N
P5	Medium, 6-20 clinicians	Internal Medicine	N
P6	Medium, 6-20 clinicians	Family Medicine	N
P7	Large, > 20 clinicians	Internal Medicine	N

Individual clinicians. Data were collected among 52 individual clinicians working within the seven practices participating in data collection. All patients with active email addresses on file with their practice were sent an email invitation from the research group to participate in use of a potential new primary care measure. Participation was opt-in and voluntary. Number of patient responses to the PCPCM PRO per clinician were sufficient during the limited data collection period (four weeks) to allow for clinician level measurement of 16 clinicians. Sufficient sample means at least 30 responses per clinician. Below is a table of clinician characteristics among the five practices that encompassed the 16 clinicians used for measure testing. Note, two of the practices did not have enough data for clinician level analysis and have been excluded from Table 1.5.2.

Table 1.5.2 Clinician Characteristics

ID	Gender	Degree	Discipline	Percent Minority Clinicians	Included in Practice Score Testing	Included in Clinician Score Testing
P1	1 Female 2 Male	3 MDs	Pediatrics	0%	Yes	Yes
P2	1 Female 2 Male	3 MDs	Family Med	33%	Yes	Yes
P3	2 Female 1 Male	2 NPs 1 MD	Family Med	0%	Yes	Yes

ID	Gender	Degree	Discipline	Percent Minority Clinicians	Included in Practice Score Testing	Included in Clinician Score Testing
P4	6 Female 4 Male	1 NP 1 DO 8 MDs	Family Med	10%	Yes	Yes
P5	8 Female	1 NP 7 MDs	Internal Med	50%	Yes	Yes

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

Three patient groups were included in the testing and analysis for the PCPCM PRO-PM. Patients were included if they submitted completed PCPCM PRO instruments. The PCPCM PRO instrument is used to collect the data necessary for calculating the PCPCM PRO-PM Performance Score. The PCPCM PRO is an 11-item patient reported instrument and is described in more detail in a later section.

Patient groups 1 and 2 – Participants in instrument validation (n=2,552).

We conducted testing and analysis on the PCPCM PRO instrument among 2,552 patients. This is referred to as the patient dataset and includes two subgroups: the patient online group and the patient point of care group.

Patient online group. Electronic responses to the PCPCM PRO were collected among 2,229 patients. Characteristics of patients in the online patient group are found in Table 1.6.1 below. Responses were collected through SurveyMonkey and a service they offer called SurveyMonkey Audience. This is a national service. People who self-identify as willing to take a survey are screened by SurveyMonkey Audience to obtain demographic information. We asked SurveyMonkey Audience to field the PCPCM PRO among adults, 18 years old or older. We requested two fieldings of the PCPCM PRO among at least 1,000 responses each time. We requested that each sample include a mix of patient characteristics that mirrored census-based population demographics. As part of the survey process, respondents were asked if they had completed a visit with primary care within the last year, as a screening question for participation.

Table 1.6.1 Patient Characteristics – Online patient group

Characteristic	N = 2,229
Age: 18-29	371
Age: 30-44	578
Age: 45-60	442
Age: >60	685
Gender: Female	1109
Gender: Male	967

Characteristic	N = 2,229
Self-identify as minority: Yes	439
Self-identify as minority: No	1718
Years knowing physician: <1 yr.	155
Years knowing physician: 1-3 yrs.	702
Years knowing physician: 4-10 yrs.	697
Years knowing physician: >10 yrs.	611
Income: <25K	297
Income: 25-49.9K	383
Income: >50K	1071
Region: Northeast	414
Region: Midwest	524
Region: South	634
Region: West	482

Patient point of care group. Paper responses to the PCPCM PRO were collected among 323 patients at 4 practice locations in Richmond, Virginia. The practices used for this data collection included a community health center, a pediatric practice, a hospital owned practice, and a privately-owned practice. Practices were selected from the Ambulatory Care Outcomes Research Network (ACORN). All Richmond-based practices in ACORN were offered participation and those selected were the first four to volunteer. Characteristics of patients in the point of care dataset are found in Table 1.6.2 below.

Data collection for the patient point of care group was paper-based. A research assistant was located in each practice's waiting room for a one-week period. At check-out, the practice staff would identify the research assistant to patients, mention the collection of information, and that participation was voluntary and for research purposes. Patients opted-in to participation when leaving the practice. If

they opted-in, the research assistant would offer the PCPCM PRO on a clipboard for the patient to fill out.

Table 1.6.2 Patient Characteristics – Point of care patient group

Participant Characteristic	2 Private Family Medicine Practices N= 126	Community Health Center N = 97	Pediatric Practice N = 100
Age	N (%)		
18-29	15 (22)	N (%)	N (%)
30-44	31 (41)		
45-60	52 (25)	***	***
> 60	28 (12)		
Gender (Female)	88 (70)	***	69 (69)
Years knowing physician			
<1	28 (22)	6 (7)	
1-3	52 (41)	21 (23)	
4-10	31 (25)	33 (37)	***
>10	15 (12)	30 (33)	

***** Item not included on this version of the survey**

Patient group 3 – participants in performance measure validation (n=1,205).

We conducted testing and analysis on the PCPCM PRO-PM among 1,205 patients. This is referred to as the clinician and group practice dataset. The clinicians and group practices are described above. Invitations to submit a PCPCM PRO instrument response were sent to all patients with known email addresses among the seven participating practices. Patient responses included in our testing and analysis were any responses submitted within a four-week period. Characteristics of patients in patient group 3 are found in Table 1.6.3 below.

Table 1.6.3 Patient Characteristics – Clinician and group practice

Characteristic	N = 1,205
Age: 18-29	2.4%
Age: 30-44	15.5%
Age: 45-60	30%
Age: >60	52.1%
Gender: Female	71.5%
Gender: Male	28.5%

Characteristic	N = 1,205
Self-identify as minority: Yes	29.3%
Self-identify as minority: No	70.7%
Years knowing physician: <1 yr.	2%
Years knowing physician: 1-3 yrs.	29%
Years knowing physician: 4-10 yrs.	48%
Years knowing physician: >10 yrs.	21%
Did you find this survey meaningful?: Yes	73%

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.

Not applicable; same data used for each aspect of testing.

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.*

For patient groups 1 and 2, social risk factors available and analyzed included the following: age, gender, self-identified minority status, income, and self-identified health status. See Tables 1.6.1 and 1.6.2. For patient group 3, social risk factors are listed in Table 1.6.3 but were not analyzed. Both English and Spanish versions of the PCPCM PRO were offered to patients. All responding patients responded in English.

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*)

The PCPCM PRO-PM Performance Score is calculated using the PCPCM PRO instrument. The PCPCM PRO instrument includes 11-items that are critical data elements used in this measure.

Critical Data Elements – Reliability Testing

For testing the critical data elements, we used two methods of reliability testing. These methods are discussed in detail in our peer reviewed publication. That publication is paraphrased and quoted below.

Reliability test one of critical data elements: exploratory principal axes factor analysis

> In order to identify the number of constructs represented among the 11 PCPCM PRO items, we conducted exploratory principal axes factor analysis within the patient online group (n=2,229). We also examined the scree plot and Eigenvalues.

Reliability test two of critical data elements: Rasch item fit statistics

> Rasch item fit statistics were computed for each PCPCM PRO item. In Rasch modeling, when all items in a measure are a good fit, evidence of construct validity of the measure is provided. Findings from this analysis can reveal variation in level of question (item) difficulty. Item sets of varying difficulty allow greater ability to see variation among responding populations. In addition, Rasch item reliability statistics were computed to assess the level of confidence that items would have the same level of difficulty in another sample of participants. All Rasch analyses were computed using WINSTEPS 4.10 software and based on the Rasch partial credit model.

Reliability test three of critical data elements score: Cronbach's α reliability coefficient

> We used Cronbach's internal consistency alpha statistic to test score reliability. Conceptually, this estimate of reliability best fits what we expect the total score to provide, i.e., we expect every item to contribute independently to the total score. Moreover, we expect each item response to be consistent, in the same direction, with the other 10 items on the scale with no inversions or inconsistencies. The only way to get a high Cronbach reliability coefficient is if all items contribute to the total score and in the same direction. One or more poorly performing items would pull down the total reliability.

Performance Measure Score – Reliability Testing

For testing the performance measure score, we used two methods of reliability testing.

Reliability test one of performance measure score: intra-class correlation coefficient (ICC)

> To assess performance measure score reliability, we calculated the intra-class correlation coefficient using a one-way random model in which patients were randomized and the PCPCM PRO items were fixed. This was conducted at both the clinician – individual and the clinician – group/practice levels.

Reliability test two of performance measure score: split sample reliability

> Another aspect of reliability is to understand the degree to which 'repeated measurements of the same entity agree with each other.' In this case, we consider whether randomly selected subsets of submitted PCPCM PRO instruments will produce similar findings with regarding to performance measure score. To assess this, we used a Guttman lamda 4 split half reliability findings.

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 results of reliability testing are presented here by level of test and then by name of testing method. These descriptions are followed by tables that support these findings.

Critical Data Elements – Results of Reliability Testing

Reliability test one of critical data elements: Results – exploratory principal axes factor analysis

This analysis resulted in the identification of a single factor: person centered primary care. This was further confirmed through calculated Eigenvalues of 6.9 for the patient online group and 4.7 for the patient point of care group.

Reliability test two of critical data elements: Results – Rasch item fit statistics

Rasch item fit statistics ranged from 0.62 to 1.44 for the patient online group. Rasch item reliability was 0.99 for this group.

Rasch item fit statistics ranged from 0.55 to 1.49 for the patient point of care group. Rasch item reliability was 0.98 for this group.

Table 2a2.3.1 PCPCM PRO Items, Means, Factor Loading and Goodness of Fit Statistics

Item	Mean	S.D.	Factor Loading	Item-Total Correlation
My practice makes it easy for me to get care.	3.13	.85	.70	.67
My practice is able to provide most of my care.	3.11	.84	.70	.66
In caring for me, my doctor considers all of the factors that affect my health.	3.24	.85	.80	.76
My practice coordinates the care I get from multiple places.	2.83	1.01	.64	.62
My doctor or practice knows me as a person.	2.93	1.06	.83	.81
My doctor and I have been through a lot together.	2.26	1.16	.66	.64
My doctor or practice stands up for me.	2.74	1.01	.85	.83
The care I get takes into account knowledge of my family.	2.69	1.09	.80	.78
The care I get in this practice is informed by knowledge of my community.	2.30	1.06	.71	.70
Over time, this practice helps me to meet my goals.	3.03	.91	.85	.82
Over time, my practice helps me stay healthy.	2.84	.96	.85	.81

Table 2a2.3.2 Items, Rasch Item Measures and S.E., Item-Total Correlations, Outfit Mean Squares, and Standardized Fit Statistics

Item	Item Measure	S.E.	Corrected Item Total Correlation	Outfit Mean Square	ZSTD
My practice makes it easy for me to get care.	-.86	.05	.71	1.18	2.6
My practice is able to provide most of my care.	-.73	.05	.71	1.33	4.8
In caring for me, my doctor considers all of the factors that affect my health.	-1.05	.05	.82	0.65	5.7
My practice coordinates the care I get from multiple places.	-.12	.05	.74	1.15	2.7
My doctor or practice knows me as a person.	-.06	.05	.82	0.86	2.7
My doctor and I have been through a lot together.	1.74	.05	.66	1.44	7.0
My doctor or practice stands up for me.	.16	.05	.83	0.73	5.9
The care I get takes into account knowledge of my family.	.40	.05	.77	1.03	0.7
The care I get in this practice is informed by knowledge of my community.	1.08	.05	.70	1.36	6.6
Over time, this practice helps me to meet my goals.	-.54	.05	.84	0.62	7.5
Over time, my practice helps me stay healthy.	-.02	.05	.82	0.79	4.3

Reliability test three of critical data elements: Results – Cronbach's α reliability coefficient

The Cronbach alpha internal consistency reliability coefficient was computed to be 0.95 for the patient online group and 0.91 for the patient point of care group. The corrected item-total correlations for the patient online group ranged from 0.66 to 0.84 for the 11 items in the scale. The corrected item-total correlations for the patient point of care group ranged from 0.27 to 0.86 for the 11 items in the scale.

The Cronbach alpha internal consistency reliability coefficient was computed to be 0.91 for the clinician and group practice dataset. The corrected item-total correlations for this group ranged from 0.58 to 0.79 for the 11 items in the scale.

Performance Measure Score – Results of Reliability Testing

PCPCM PRO-PM performance scores varied by clinician and by practice. See Tables below.

Table 2b4.2.1 PCPCM PRO-PM Performance Measure Scores by Practice

Practice	Mean PCPCM PRO Score	PCPCM PRO-PM Performance Score	Standard Deviation	F-Statistic	P-Value
1	3.45	86%	0.56	3.45	0.004
2	3.44	86%	0.58	3.45	0.004
3	3.48	87%	0.55	3.45	0.004
4	3.60	90%	0.50	3.45	0.004
5	3.45	86%	0.64	3.45	0.004
6	3.36	84%	0.61	3.45	0.004

Table 2b4.2.2 Mean PCPCM PRO-PM Performance Measure Scores by Clinicians

Clinician	Mean PCPCM PRO Score*	PCPCM PRO-PM Performance Score*	Standard Deviation	F-Statistic	P-Value
1	3.37	84%	0.58	4.14	0.0001
2	3.50	88%	0.42	4.14	0.0001
3	3.50	88%	0.54	4.14	0.0001
4	2.98	75%	0.91	4.14	0.0001
5	3.37	84%	0.61	4.14	0.0001
6	3.59	90%	0.50	4.14	0.0001
7	3.45	86%	0.55	4.14	0.0001
8	3.39	85%	0.57	4.14	0.0001
9	3.57	89%	0.49	4.14	0.0001
10	3.66	92%	0.39	4.14	0.0001
11	3.27	82%	0.69	4.14	0.0001
12	3.34	84%	0.73	4.14	0.0001
13	3.41	85%	0.68	4.14	0.0001
14	3.65	91%	0.46	4.14	0.0001
15	3.25	81%	0.70	4.14	0.0001
16	3.49	87%	0.55	4.14	0.0001

Reliability test one of performance measure score: Results – ICC and split sample reliability

Clinician-level Findings

Clinician	N	Alpha	ICC	Guttman Reliability
1	35	0.9	0.87	0.82
2	83	0.82	0.76	0.79

Clinician	N	Alpha	ICC	Guttman Reliability
3	58	0.89	0.86	0.86
4	34	0.95	0.94	0.95
5	37	0.9	0.87	0.92
6	96	0.9	0.87	0.91
7	31	0.88	0.86	0.86
8	46	0.89	0.87	0.89
9	41	0.89	0.87	0.9
10	32	0.81	0.8	0.82
11	45	0.93	0.91	0.92
12	39	0.94	0.93	0.94
13	31	0.95	0.94	0.92
14	86	0.89	0.88	0.89
15	59	0.93	0.91	0.9
16	74	0.92	0.89	0.88

Practice-level Findings for ICC

Practice	N	Alpha	ICC
1	170	0.91	0.88
2	168	0.91	0.88
3	220	0.9	0.88
4	149	0.9	0.89
5	188	0.92	0.91
6	264	0.91	0.89

Practice-level Findings for Split half reliability and Guttman Reliability

Practice	N*	Split half reliability	Split half reliability	Guttman reliability	Guttman reliability
- -	- -	*Random sample 1*	*Random sample 2*	*Random sample 1*	*Random sample 2*
1	85	0.9	0.86	0.94	0.9
2	79	0.9	0.88	0.93	0.91
3	109	0.88	0.92	0.92	0.9
4	64	0.87	0.91	0.93	0.91

Practice	N*	Split half reliability	Split half reliability	Guttman reliability	Guttman reliability
5	106	0.92	0.92	0.93	0.95
6	125	0.88	0.88	0.91	0.93

* Sample size is for each of the split half samples

- - cell intentionally left blank

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?*)

Our interpretation of the results are presented here by level of test and then by name of testing method.

Critical Data Elements – Interpretation of Results of Reliability Testing

Reliability test one of critical data elements: Interpretation – exploratory principal axes factor analysis

Eigenvalues calculated support reliability. We used factor loading, scree plots and Eigenvalues to determine that all PCPCM PRO items represented a single factor – person centered primary care. Following accepted standards, we interpreted an Eigenvalue of >1.0 to be significant. The Eigenvalues resulting in the use of the PCPCM PRO that is the basis of the PCPCM Performance measure far exceed values of >1.0.

Factor loading calculated supports reliability. Following accepted standards, we interpreted factor loadings of >0.4 as a good association between the item and the construct being assessed, loadings of >0.6 to demonstrate a strong association, and those >0.8 to demonstrate a very strong association. All factor loadings for the PCPCM PRO items exceeded 0.6 with the majority exceeding 0.7, indicating strong and very strong associations.

Reliability test two of critical data elements: Interpretation – Rasch item fit statistics

Following accepted standards, we interpreted Rasch item reliability statistics of >0.9 to represent excellent internal reliability, and Rasch item fit statistics of acceptable to be 0.5 to 1.5. Within our data, the lowest Rasch item was 0.98 and the Rasch item fit statistics ranged from 0.62 to 1.44.

Reliability test three of critical data elements: Interpretation – Cronbach's α reliability coefficient

Cronbach α reliability statistics support reliability. Following accepted standards, we interpreted Cronbach α reliability statistics of >0.8 to represent excellent internal consistency and reliability. All samples had Cronbach α scores exceeding 0.9.

Performance Measure Score – Interpretation of Results of Reliability Testing

To interpret the ICC findings, we follow the guidance of Koo and Li.[1] They have found the following:

- < 0.50 is poor
- Between 0.50 and 0.75 is moderate
- Between 0.75 and 0.90 is good
- > 0.90 is excellent

Clinician level ICC findings ranged from 0.76 to 0.94 indicating good to excellent reliability.

Practice level ICC findings ranged from 0.88 to 0.91 indicating largely excellent reliability.

Clinician level Guttman reliability findings ranged from 0.79 to 0.95.

Practice level Guttman reliability findings ranged from 0.9 to 0.95

Both are indications of good to excellent reliability.

1. Koo TK, Li MY (June 2016). "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research". *Journal of Chiropractic Medicine*. **15** (2): 155–63.

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)*

Critical Data Elements – Validity Testing

The PCPCM PRO-PM is calculated using the PCPCM PRO instrument. The 11 items of the PCPCM PRO are the critical data elements of the PCPCM PRO-PM

Background. The PCPCM PRO was created through a multi-step approach. First, we conducted a survey to identify a preliminary set of quality indicator areas of greatest importance to primary care clinicians (n=525), patients (n=412), and employers (n=85). Qualitative content analysis was used to identify 18 quality indicator areas among the ~10,000 written comments received.[1] Next, we hosted a national conference (http://www.starfieldsummit.com/starfield3) in which we engaged a diverse range of 70 national leaders in primary care quality and measurement, including patients and patient advocates.[2] Qualitative analyses of conference data, combined with previously conducted literature reviews, resulted in a reduction of quality areas from 18 to 11. A single item was created to reflect each quality area as identified. These 11 items resulted in the first draft of the PCPCM PRO.

3. Etz RS, Gonzalez MM, Brooks EM, Stange KC. Less AND more are needed to assess primary care. J Am Board Fam Med. 2017; 30(1): 13–15.
4. Etz, RS and the Starfield Writing Team. Conference Brief: Framework of PC Measure Domains and Key Elements. 2017. Starfield Summit III: Washington, DC. Accessed January 6, 2020, http://www.starfieldsummit.com/resources3

We used four different analyses to test the validity and meaningfulness of our critical data elements.

Validity test one of critical data elements: Cognitive interviews among patients

We conducted cognitive interviews among 10 patients regarding the 11 items of the PCPCM PRO. For these interviews, we asked participants to read each item aloud and then to explain in their own words what the item is asking. We then asked participants to pick one of the four response options for each item, followed by asking them to share in their own words why they chose that answer. This process allowed us to understand if the items were likely to be interpreted as intended by participants, and whether the answers they chose were providing the type of information we expected.

Validity test two of critical data elements: Pilot testing of the PCPCM PRO

We fielded the PCPCM PRO to a pilot set of online participants. The invitation to respond to the PCPCM PRO was given to an existing set of patient respondents that had participated in previous ACORN studies. We accepted all responses, stopping data collection at 30. This online pilot fielding included each of the 11 items, as well as questions following each of the 11 items that asked whether the question was hard to understand, whether the question was hard to answer, and whether the question seemed relevant to their primary care experience.

Validity test three of critical data elements: Survey for meaningfulness

Two groups were surveyed to determine if the 11 items of the PCPCM PRO were meaningful to participants: clinicians and patients.

Drawing on a database of clinicians that have participated in previous ACORN studies, we sent an invitation to participate in a quick validation survey of the PCPCM PRO. We accepted all responses within a one-week time period. We received 285 responses. The clinician validation survey presented the 11 items to clinicians and then asked 1) if the item appeared meaningful to them, 2) if they currently collected that information in their practice, and 3) if they would find it meaningful to receive patient responses to these items.

Among 1,140 of the participants in the patient dataset described above, we fielded the PCPCM PRO along with an additional question: If your doctor had these survey results, would it help your care?

Validity test four of critical data elements: Concurrent validity testing

We used concurrent validity to test the validity of the PCPCM critical data elements. Concurrent validity is used to determine how well a new measure assesses an area in comparison to an established measure that assesses a similar area. We conducted this test using all participants in the patient dataset described above (n=2,552). The PCPCM PRO was fielded alongside two previously validated patient reported PRO-PMs in a single survey of patient participants: the Patient Enablement Instrument (PEI),[1] and the What Matters Index (WMI).[2,3] The PEI is a patient reported measure of a patient's ability to understand and cope with their health issues as a result of the care received. The WMI is a patient reported measure that has a demonstrated positive association with cost and utilization of health services. For comparative analyses, we used *t* tests and analysis of variance.

1. Howie JG, Heaney DJ, Maxwell M, Walker JJ. A comparison of a Patient Enablement Instrument (PEI) against two established satisfaction scales as an outcome measure of primary care consultations. Fam Pract. 1998; 15(2): 165–171.
2. Wasson JH, Ho L, Soloway L, Moore LG. Validation of the What Matters Index: A brief, patient-reported index that guides care for chronic conditions and can substitute for computer-generated risk models. PLoS One. 2018; 13(2): e0192475.
3. Wasson JH, Soloway L, Moore LG, Labrec P, Ho L. Development of a care guidance index based on what matters to patients. Qual Life Res. 2018; 27(1): 51–58.

Performance Measure Score – Validity Testing

We also used concurrent validity to test the validity of the PCPCM PRO-PM performance measure score. We tested the PCPCM PRO-PM performance score alongside the PEI score at both the clinician and practice levels.

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

Critical Data Elements – Results from Validity Testing

Validity test one of critical data elements: Results – Cognitive interviews

Responses from patients during the cognitive interviews were compared with the previously written detailed descriptions of each PCPCM PRO item. Nine of 10 patients successfully identified the intended quality indicator area for each item. The language patients used in their description of indicator areas resulted in minor revisions to the language of 3 items. Patient descriptions of their intention when selecting response options for each item were consistent with the development team's intentions regarding the established scale.

Validity test two of critical data elements: Results – Pilot testing

Among the 30 patient respondents, responses to the two questions fielded with the piloting of the PCPCM PRO:

- Was this item hard to answer: 27 said no.
- Was this item relevant to their primary care experience: 24 said yes.

Validity test three of critical data elements: Results – Survey for meaningfulness

Among the 285 clinicians surveyed, responses to the three questions asked after sharing the 11 PCPCM PRO items:

- Is this item important to your practice: 225 (79%) said yes
- Is this item personally important to you: 242 (85%) said yes
- Would it be meaningful to you to receive your patients' responses to these items: 239 (84%) said yes

Among the 1,140 patients surveyed, 60% said yes – if their doctor received their responses to the PCPCM PRO, it would help their care.

Validity test four of critical data elements: Results – Concurrent validity testing

Results of the comparison of the PCPCM PRO with the WMI and the PEI are shared in the table below. As a reminder, the patient dataset includes 2,552 PCPCM PRO responses: 2,229 from the patient online group and 323 from the patient point of care group.

Validation Item	**Measurement**	**Patient Online Group** (N=2,229): **N**	**Patient Online Group** (N=2,229): **Mean Scores**	**Patient Online Group** (N=2,229): **S.D.**	**Patient Online Group** (N=2,229): **P**	**Patient Point of Care Group** (N=323): **N**	**Patient Point of Care Group** (N=323): **Mean Scores**	**Patient Point of Care Group** (N=323): **S.D.**	**Patient Point of Care Group** (N=323): **P**
What Matters Index (1)	0	315	3.23	.56	- -	106	3.62	.39	- -
What Matters Index (2)	1	348	2.73	.87	0.0001	57	3.47	.49	0.08
What Matters Index (3)	≥2	396	2.59	.82	- -	37	3.63	.52	- -
Patient Enablement Index (1)	0 1-5	488 543	1.94 2.56	.61 .57	0.0001	23	3.04 3.25	.58	0.0001
Patient Enablement Index (2)	6-11	764	3.15	.49	- -	81	3.50	.44	- -
Patient Enablement Index (3)	12	386	3.67	.39	- -	81	3.78	.30	- -

- - cell intentionally left blank

Performance Measure Score – Results of Validity Testing

Results of the comparison of the PCPCM PRO-PM with the PEI are shared in the tables below. At the clinician level, this test included only 4 clinicians as it required each clinician have at least 25 patients complete both the PCPCM and the PEI.

Clinician level concurrent validity results

Clinician	N	Alpha	ICC*	PEI Pearson correlation**	PEI (n,%)***
1	83	0.82	0.76	0.39	29 (35)
2	96	0.9	0.87	0.64	37 (38)
3	86	0.89	0.88	0.55	28 (33)
4	74	0.92	0.89	0.65	26 (35)

* ICC one-way random: patients are random, items fixed

** All correlations, p<0.05

*** percentages are percent of the patient respondents that completed both the PCPCM and the PEI

Practice level concurrent validity results

Practice	N	Alpha	ICC*	PEI Pearson correlation**	PEI (n,%)***
1	170	0.91	0.91	0.58	60 (35)
2	168	0.91	0.91	0.6	52 (31)
3	220	0.9	0.9	0.54	89 (40)
4	149	0.9	0.9	0.56	53 (36)
5	188	0.92	0.92	0.67	70 (37)
6	264	0.91	0.91	0.5	82 (31)

* ICC one-way random: patients are random, items fixed

** All correlations, p<0.001

*** percentages are percent of the patient respondents that completed both the PCPCM and the PEI

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?*)

Critical Data Elements – Interpretation of Validity Testing

Validity test one of critical data elements: Interpretation – Cognitive interviews

The match of patient descriptions regarding the meaning of each item with the development team's intended meaning of each item indicated the items were appropriately articulated.

Validity test two of critical data elements: Interpretation – Pilot testing

Items were determined to be well articulated among a larger sample as 90% of patients had no trouble answering the PCPCM PRO and 80% indicated the PCPCM PRO was relevant to their care.

Validity test three of critical data elements: Interpretation – Survey for meaningfulness

Items were determined to be valid as perceived by clinicians with 79% feeling that the items were relevant to their practice and a larger percentage (85%) stating that the items were personally important to them. A large majority (84%) indicated the responses would be meaningful to them if received from their patients.

Among the 1,140 patients surveyed, 60% said yes – if their doctor received their responses to the PCPCM PRO, it would help their care.

Validity test four of critical data elements: Interpretation – Concurrent validity testing

PCPCM performance measure scores were strongly and positively associated with the WMI (online sample, p = .0001; clinical sample weaker at p = .08) and PEI (online and clinical samples, both p = .0001). We interpreted this to mean the PCPCM PRO-PM performance measure score had strong concurrent validity.

Performance Measure Score – Interpretation of Validity Testing

Pearson correlations between the PEI and the PCPCM for each practice are consistent in the upper .50s, and in agreement with the correlation for the total sample as well. This indicates strong evidence for concurrent validity of the PCPCM at the practice level.

3 of the 4 Pearson correlations between the PEI and the PCPCM at the clinician level are consistent in the upper .50s and lower .60s. One clinician's comparison yielded a Pearson correlation of 0.39. The sample size here is lower due to the lower number of clinicians for whom sufficient PEI responses were recorded to allow for analysis. This indicates moderate concurrent validity for the PCPCM at the clinician level. More analysis needs to be done and will be done over the next 12 months.

Also note the last column in the table above shows the sample size for each practice that completed a PEI, i.e., the percentage of patients who first completed the PCPCM and then also completed the PEI. Percentages of respondents also completing the PEI are displayed to show that the sample sizes varied from 31% to 40% or generally only about a third of each practice administered the PEI and yet we still got strong evidence of concurrent validity for the PEI at the practice level.

2b2. EXCLUSIONS ANALYSIS

NA ☒ no exclusions — *skip to section 2b4*

We have no exclusion analysis. This section is skipped in line with instructions.

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

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

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

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

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

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.

This question is skipped as we used no risk adjustment or stratification.

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.

In our published study using the patient dataset,[1] the PCPCM showed no association with patient's age, sex, or number of years knowing the physician, and in online samples that asked about race, the PCPCM scores showed no association with race. Therefore, to foster ease of scoring and interpretation, we do not recommend adjusting for these factors. Because the recommended use of the PCPCM is for individual clinicians or practices to conduct a self-assessment, and to use their total and individual item scores to improve the patient-centeredness of their care, in comparison to themselves rather than to others, we do not recommend any further risk adjustment or stratification.

1. Etz RS, Zyzanski SJ, Gonzalez MM, Reves SR, O'Neal JP, Stange KC. A New Comprehensive Measure of High-Value Aspects of Primary Care. Ann Fam Med. 2019 May;17(3):221-230.

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?

This question is skipped as we used no risk adjustment or stratification.

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply: N/A as we used no risk adjustment or stratification.

- ☐ **Published literature**
- ☐ **Internal data analysis**
- ☐ **Other (please describe)**

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

This question is skipped as we used no risk adjustment or stratification.

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.**

This question is skipped as we used no risk adjustment or stratification.

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*)

This question is skipped as we used no risk adjustment or stratification.

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*)**:**

This question is skipped as we used no risk adjustment or stratification.

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

This question is skipped as we used no risk adjustment or stratification.

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

This question is skipped as we used no risk adjustment or stratification.

2b3.9. Results of Risk Stratification Analysis:

This question is skipped as we used no risk adjustment or stratification.

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*)

This question is skipped as we used no risk adjustment or stratification.

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*)

This question is skipped as we used no risk adjustment or stratification.

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)*

Our statistical approach to identifying practice and clinician differences in the PCPCM PRO-PM relies on a two-step evaluation. In step one, an Analysis of Variance (ANOVA) is computed as an omnibus test statistic to determine if any of the pairs of practices or clinicians being compared show evidence of statistical significance. Once a significant F-statistic is obtained, the second step is to test pairs of practice or clinician means for statistical significance using t-tests for independent groups. The selection of which pairs to test will be guided by the larger differences observed but also by their clinical significance as reflected in Cohen's d statistic which is a measure of effect size. Statistically significant pairs that have clinically meaningful effect size differences are the pairwise difference of most relevance in reporting the results.

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*)

In testing for practice differences among our six practice sites, in step 1, the F-statistic in the Analysis of Variance was 3.45. Based on 5 and 1153 degrees of freedom, this F-statistic was significant at the p=0.004 level. In step 2, independent t-test were computed for the larger differences observed and Cohen's d statistic was also computed to estimate the effect size difference which has clinical relevance. With a total scale standard deviation of 0.59, a small effect size of d=0.2 standard deviations translate into a difference of 0.12 between practice or clinician mean scores. Any two practice means that differed by at least 0.12 were considered clinically different. These pairs of means were then tested for statistical significance by means of a t-test. We observed four of the six practices to have mean performance scores to be equal within sampling variation. For the data in the table, these are the means for Practices 1, 2, 3, and 5. Practices 4 and 6 had the largest difference in means (0.24) which is a small to medium effect size. Practice 4 also differs from Practices 1, 2, 3, and 5 as well. These observed differences are followed by testing the differences for statistical significance. For example, the t-test between the means for Practice 4 and Practice 6 was t=4.09, p=0001. In this way we are able to determine which practices have similar mean score performance and which practices have very different score performances. The same logic and step testing procedure was applied to the 16 clinicians with at least 30 patients. In testing for clinician differences, the ANOVA F-statistic was 4.14. Based on 15 and 811 degrees of freedom, this F-statistic was significant at the p=0.0001 level. In the subsequent pairwise analyses, we observed much larger effect sizes for many clinician differences in the moderate range as defined by Cohen's d statistic. For example, the difference in score performance between Clinician 10 and Clinician 4 was 0.68, which is in the medium to large effect size range. The t-test for this difference was 3.90, p=0.0002.

Table 2b4.2.1 PCPCM PRO-PM Performance Measure Scores by Practice

Practice	Mean PCPCM PRO Score	PCPCM PRO-PM Performance Score	Standard Deviation	F-Statistic	P-Value
1	3.45	86%	0.56	3.45	0.004
2	3.44	86%	0.58	3.45	0.004
3	3.48	87%	0.55	3.45	0.004
4	3.60	90%	0.50	3.45	0.004
5	3.45	86%	0.64	3.45	0.004
6	3.36	84%	0.61	3.45	0.004

Table 2b4.2.2 Mean PCPCM PRO-PM Performance Measure Scores by Clinicians

Clinician	Mean PCPCM PRO Score*	PCPCM PRO-PM Performance Score*	Standard Deviation	F-Statistic	P-Value
1	3.37	84%	0.58	4.14	0.0001
2	3.50	88%	0.42	4.14	0.0001
3	3.50	88%	0.54	4.14	0.0001
4	2.98	75%	0.91	4.14	0.0001
5	3.37	84%	0.61	4.14	0.0001
6	3.59	90%	0.50	4.14	0.0001
7	3.45	86%	0.55	4.14	0.0001
8	3.39	85%	0.57	4.14	0.0001
9	3.57	89%	0.49	4.14	0.0001
10	3.66	92%	0.39	4.14	0.0001
11	3.27	82%	0.69	4.14	0.0001
12	3.34	84%	0.73	4.14	0.0001
13	3.41	85%	0.68	4.14	0.0001
14	3.65	91%	0.46	4.14	0.0001
15	3.25	81%	0.70	4.14	0.0001
16	3.49	87%	0.55	4.14	0.0001

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 variation of PCPCM PRO-PM performance scores, as found in our clinician and practice dataset, illustrates a difference of moderate (0.5) to large effect size (0.8) among clinicians. This is evidence of a performance gap, existence of measurable differences among practices, and the evidence of opportunities for improvement. With a greater than 0.5 standard deviation as our benchmark, our analyses indicate the PCPCM PRO and PCPCM PRO-PM performance scores show clinically meaningful differences.

The results support the contention that PCPCM PRO-PM performance scores are able to distinguish between practices and between individual clinicians, providing that individual clinicians have an adequate sample size (at least 30) of patients and the practices have an adequate sample size (at least 50). Even with the lower statistical power provided by the clinician samples, the effect size differences observed were larger than for practices and thus statistical significance attained. Differences in mean performance scores were much greater for the clinicians than for the practices and this is logical, as the practice is an aggregate measure of clinicians. Since the PCPCM PRO-PM can differentiate among practices and clinicians, it could be used as an outcome measure to evaluate clinician or practice level performance.

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

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

Following instructions, we have only one set of specifications and therefore skipped this section.

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

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

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

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

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

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

The crowd-sourced nature of our methodology for online samples did not allow us to record response rates for the PCPCM PRO. Among those who answered at least 1 item on the PCPCM PRO, >99% answered at least 8 items as required for the instrument responses to be included in analyses. PCPCM PRO instruments with fewer than 8 items answered, were considered to be incomplete, and were omitted from calculations. Since <1% of our sample of 2,552 respondents were omitted, it is statistically impossible for the exclusion of incomplete instruments to bias the data.

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*)

Less than 1% of respondents, among a total sample of 2,552 patients, had incomplete instruments. Received instruments that had fewer than eight (8) responses were considered to have missing data and therefore were deemed incomplete. The amount of missing data (<1%) were insufficient to examine trends across providers.

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

Negligible when fielded as specified.

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.

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**.

Patient/family reported information (may be electronic or paper)

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).

Data elements used are collected directly from patients. Patients are invited to fill out the PCPCM PRO instrument electronically. In almost all cases, patients are sent an email with an embedded link either to an electronic survey platform, or to an electronic Patient Reported Outcomes (PRO) module as part of the PRIME registry. The most likely format will be electronic sources, however paper-based instruments can be used. In all implementation of the PCPCM PRO-PM to date, performance scores and feedback are provided electronically to practices and clinicians. PCPCM PRO-PM scores are calculated at the point of data collection and then shared with the measured entity.

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.

Collecting patient reported information has known challenges. Some practices are able to collect patient reported data using a patient portal in their EHR, however many practices use third party vendors to assist with this process. Based on electronically collected PCPCM PRO instruments from over 4,000 patients, the average time to complete the PCPCM is 60-90 seconds.

We have previously implemented the PCPCM PRO-PM in the PRIME qualified clinical data registry, consisting of a potential pool of over 2400 clinicians, nationally. In this case, PRIME has a Patient Reported Outcomes (PRO) module through which patient data is collected. The QCDR takes on the responsibility of contacting patients as specified by the measure and calculating performance scores, as specified. The PRIME registry includes a clinician dashboard for its members. On the dashboard, PRIME members are able to receive information on their PCPCM PRO-PM score. In this case, practices and clinicians are able to see their data as collected, comparing it to local and national benchmarks. They are also able to see the individual item scores within the PCPCM PRO instrument, thereby enabling targeted quality improvement work. All practices and clinicians involved in this implementation have indicated very low burden related to the PCPCM PRO-PM.

Additionally, we have piloted the PCPCM PRO-PM among 10 practices that do not share an EHR, patient portal, or registry. In this case, our research team created a HIPAA compliant electronic platform that we then made freely available to participating practices. Onboarding for PCPCM PRO-PM use with the help of our platform required that practices fill out basic demographic information for the practice and for participating clinicians. They were then able to securely upload known patient email addresses, patient birth month, and clinician to which the patient is attributed within the practice.

We were able to onboard practices within 1 week and were able to collect PCPCM PRO-PM scores for 16 clinicians and 6 practices within the following 2 weeks. All of this was done in the Spring 2020, during the height of the COVID-19 pandemic in the US. That practices and clinicians were willing to participate, onboarded quickly, and met the threshold of responses required to compute a PCPCM PRO-PM score during the pandemic is evidence of the low practice/clinician burden associated with use of the PCPCM PRO-PM, and successful strategies to mitigate difficulty with patient generated data collection.

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)*.

There are no fees or other requirements to use any aspect of the measure as specified.

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 Quality Improvement (Internal to the specific organization)	Payment Program https://qpp.cms.gov/mips/explore-measures?tab=qualityMeasures&py=2020#measures CMS QPP MIPS

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: CMS Quality Payment Program – Merit-based Incentive Payment System
Purpose: As stated on the CMS website, "To improve the care received by Medicare beneficiaries. To lower costs to the Medicare program through improvement of care and health. To advance the use of healthcare information between allied providers and patients. To educate, engage and empower patients as members of their care team."
Geographic area and number and percentage of accountable entities and patients included: This is a national program for CMS applying to all clinicians and practices who receive CMS payments.
Level of measurement and setting: Primary care Clinician – Individual and Clinician – group/practice; the PCPCM PRO-PM has been endorsed for use as a PRIME QCDR measure in the CMS QPP MIPS program as QPP# ABFM10

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?*)
As described above, the PCPCM PRO-PM has been endorsed for use as a PRIME QCDR measure in the CMS QPP MIPS program as QPP# ABFM10. The PCPCM PRO-PM is on the draft 2020 CMS MUC list as MUC20-0042. At least one insurer is experimenting with the measure, as are two other health systems. The PCPCM PRO-PM is a new measure and is newly being piloted. There are no policies or actions that restrict access or impede implementation. We expect to be able to report current use in our maintenance for endorsement application.

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.*)

As described above, the PCPCM PRO-PM has been adopted and implemented by the PRIME registry. Beta testing was conducted in 2019 and full use of the PCPCM PRO-PM began in 2020. As PRIME members are able to report PCPCM PRO-PM for the 2020 CMS QPP cycle, we expect to have data reporting this implementation and adoption within 12 months.

The University of Colorado, in conjunction with Anthem, did their first pilot testing of the PCPCM PRO-PM in July and August of 2020. Approximately 1,550 PCPCM PRO instruments were collected and analysis is ongoing.

UC/Anthem intend a second fielding of the PCPCM PRO-PM in January 2021. If the measure continues to perform well, the intention is for it to be used as part of a new primary care payment program.

The University of Missouri, in partnership with their patient survey vendor NRC, is programming the PCPCM PRO-PM to be used among all of their primary care clinicians and locations (family medicine, internal medicine, and pediatrics). Work on this has proceeded according to schedule during the pandemic. Pilot testing of the PCPCM PRO-PM in this setting will take place within the next 12 months. If successful, the intention is for use as part of the payment system.

Virginia Commonwealth University practices were able to use our research-based platform to field the PCPCM PRO-PM among their 6 primary care settings. This work is expected to expand during the next 12 months.

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.

Virginia practices that fielded the PCPCM PRO-PM using our research based electronic platform were provided with their performance scores in relation to other Virginia practices and in relation to national benchmarks as established through validation testing. Clinicians and practices were provided with summary performance scores, as well as PCPCM PRO item level scores (as compared with benchmarks) to enable easy identification of areas in need of improvement. No practices reported negative experience or consequences. Some practices did not have many email addresses for their patients. Among those that did, response rates varied from 14-31%. Future implementation will need to include ways to help practices collect patient email addresses and to help practices know how best to engage patients. In addition, the low maintenance platform created by our research team does not yet have a clinician dashboard. We are in the process of applying for funding to create that functionality. Some tools are available on our website (https://www.green-center.org/pcpcm) to assist with quality improvement suggestions however more resources could, and will, be made available.

The PRIME registry has a clinician dashboard and has successfully implemented the PCPCM PRO-PM. No practices have reported burden or difficulty related to PCPCM PRO-PM implementation. Ten practices were involved in the beta testing of the PCPCM and over 34 have currently signed up to use the PCPCM PRO-PM for the 2020 QPP reporting. The PCPCM PRO-PM results are displayed in a variety of ways on the clinician dashboard (see Appendix A.1 pages 6-9). Patient response rate to the PCPCM PRO within PRIME was low. The email address used to distribute the PCPCM PRO-PM may not have been well identified by patients or email providers and may have been screened out as spam. A new address for PCPCM PRO delivery to patients is now being considered. Practices and clinicians are also being provided with more information regarding how best to engage patients in the use of the PCPCM PRO.

In both cases – in Virginia and among the PRIME members, all active patients were eligible to receive the PCPCM PRO. Among the Virginian practices, the all patients with known email addresses were targeted in order to establish practice baselines for use of this new measure. In PRIME, all active patients receive the PCPCM PRO instrument in the month of their birth.

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 mentioned above, clinicians have access to their data on their PRIME dashboard on an on-going basis and dashboards are updated as new data becomes available. Educational and explanatory efforts include the following:

1) The lead PCPCM PROM developer and researcher, Dr. Rebecca Etz with the Larry A Green Center, held a live webinar for all PRIME practices. Dr. Etz presented the measure, discussed each step of measure use, measure benefits, and measure results leading to practice improvement activities. Dr. Etz ended with time for questions and dialogue. The webinar was recorded and is available on the PRIME website: https://primeregistry.org/measures-that-matter/.

2) The PRIME Registry team distributes patient education materials to the practices to encourage patient participation. These materials can be found here: https://www.green-center.org/pcpcm.
3) Our QCDR vendor presents the measure to each practice in which they are working for measure review.
4) PRIME sends email announcements to PRIME practices and promotes the measure during conversations with practices during PRIME webinars and during individual demonstrations.

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.

While implementing and testing this measure, we did not receive any comments regarding burden or unexpected negative consequences related to adoption. A more thorough analysis of use experience by clinicians and patients is planned over the next 12 months.

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

This step has not yet been conducted but will be conducted within the next 12 months.

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

This step has not yet been conducted but will be conducted within the next 12 months.

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.

This step has not yet been conducted but will be conducted within the next 12 months.

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.

Each item of the PCPCM is actionable. For example, if a clinician scores poorly on the item "Over time, this practice helps me to meet my goals", clinicians can use that feedback to make sure discussion of patient goals is part of the clinical encounter. The PCPCM is freely available online, as is some advice regarding quality improvement activities and action reflection items to assist with quality improvement efforts.[12] Examples related to each of the PCPCM items include:

PCPCM Item: The practice makes it easy for me to get care

Example Actions: Altering scheduling options, availability, or who does the scheduling; Providing options for asynchronous communication or telehealth visits

PCPCM Item: This practice is able to provide most of my care

Example Actions: Schedule longer visits for more complicated problems or patients so that you provide more of the care rather than referring out; Refer in-house to staff or clinicians with specialized expertise or interests

PCPCM Item: In caring for me, my doctor considers all of the factors that affect my health

Example Actions: Consider starting visits by asking patients what matters to them for this visit; Ask patients, "What one thing would you like someone taking care of you to know?" and add this to the medical record in a consistent and easy-to-see place

PCPCM Item: My practice coordinates the care I get from multiple places

Example Actions: The medical assistant asks and documents any care received elsewhere since my last visit; When doing medication reconciliation, ask about care received elsewhere.

PCPCM Item: My doctor or practice knows me as a person.

Example Actions: Try to talk about at least one non-medical item during each visit; Ask the patients what matters to them; Link recommended treatments to what gives meaning in the patient's life

PCPCM Item: My doctor and I have been through a lot together.

Example Actions: Do phone follow up after hospital discharges; Consider other ways you might connect with patients' important health and life events.

PCPCM Item: My doctor or practice stands up for me.

Example Actions: Let patients know when you spend time doing prior authorizations; Discuss options regarding medications with patients to show them you are aware of patient costs and taking that into account.

PCPCM Item: The care I get takes into account knowledge of my family.

Example Actions: Do a quick and dirty family tree and update it periodically – try to find a consistent place in the EHR to keep this information; More routinely ask about the family as a resource or the impact of the patient's illness on the family.

PCPCM Item: The care I get in this practice is informed by knowledge of my community.

Example Actions: Participate in community events and include that in posters or on the practice website; Ask about the patient's neighborhood.

PCPCM Item: Over time, this practice helps me to meet my goals.

Example Actions: Frame care plans around patients' goals or what matters to them; Do HOPE notes: https://drwaynejonas.com/wp-content/uploads/2018/01/HOPENoteQuestions_WEB.pdf.

PCPCM Item: Over time, my practice helps me to stay healthy.

Example Actions: Look for teachable moments when the patient is open to a health behavior change; Standing orders for immunizations.

Each clinician or practice can create quality improvement activities best suited to their context.

There is a clear and large body of evidence that demonstrates the strong connection between patient experience of care and traditional health care outcomes, such as improved intermediate outcomes, greater adherence to recommended treatment, and reduced use of health care services. Two systematic reviews – one in the US and one in the UK – provide clear evidence to that effect.[1-3]

The items within the PCPCM PRO instrument used to calculate the PCPCM PRO-PM performance measure are also each individually supported by empirical resource as having a strong effect on desirable health outcomes. For instance, continuity of care is associated with improved intermediate outcomes and reductions in cost of care.[4,5] Comprehensiveness has been shown to be associated with lower hospitalization rates, greater use of preventive services, greater adherence to recommended treatment and reduction of burnout among clinicians.[3,6-10]

The ability to assess those aspects of primary care that uniquely contribute to primary care's proven ability to improve patient health outcomes and experience while reducing health burden and costs warrants a national measure. A 2014 review of measures used to assess primary care shows many aspects of care remain unassessed by current measures. The PCPCM allows for patient reported assessment of those aspects of primary care identified by patients and clinicians as most important.[11]

3 Stange KC, Etz RS, Gullett H, Sweeney SA, Miller WL, Jaén CR, Crabtree BF, Nutting PA, Glasgow RE. Metrics for assessing improvements in primary health care. Annu Rev Public Health. 2014;35:423-42.

4 Starfield B, Shi L, Macinko J. Contribution of primary care to health systems and health. Milbank Q. 2005;3:457-502.

5 Institute of Medicine (U.S.). Division of Health Care Services. Committee on the Future of Primary Care., Donaldson MS. Primary care : America´s health in a new era. Washington, D.C.: National Academy Press; 1996.

6 McWhinney IR, Freeman T. Textbook of family medicine. Oxford ; New York: Oxford University Press; 2009.

7 Institute of Medicine (U.S.). Committee on Core Metrics for Better Health at Lower Cost, Blumenthal D, Malphrus E, et al. Vital signs : core metrics for health and health care progress. Washington D.C.: National Academies Press; 2015.

8 Stange KC. The paradox of the parts and the whole in understanding and improving general practice. Int J Qual Health Care. 2002 Aug;4:267-8.

9 Starfield B. Is patient-centered care the same as person-focused care? Perm J. 2011 Spring;2:63-9.

10 Soler JK, Okkes I, Wood M, et al. The coming of age of ICPC: celebrating the 21st birthday of the International Classification of Primary Care. Fam Pract. 2008 Aug;4:312-7.

11 Etz RS, Zyzanski SJ, Gonzalez MM, Reves SR, O´Neal JP, Stange KC. A New Comprehensive Measure of High-Value Aspects of Primary Care. Ann Fam Med. 2019 May;17(3):221-230.

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 had any unexpected findings to date. We will be looking for these as implementation of the measure increases.

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

We have not had any unexpected benefits to date. We will be looking for these as implementation of the measure increases.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

No

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

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

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):

Are the measure specifications harmonized to the extent possible?

Yes

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

Conceptually, the PCPCM PRO PM and CAHPS Clinician & Group Surveys (CG-CAHPS) Version 3.0 -Adult, Child (NQF #0005) are different measures, though we have heard some people wonder if there was some overlap. The PCPCM PRO PM has some overlap with population targeted by CG CAHPS: The PCPCM PRO PM targets all patients who have been to a primary care practice, and the CG CAHPS targets all patients who have been seen in ambulatory care settings. Other than this overlap in potential population, this measure is not conceptually similar to CG CAHPS. The CG CAHPS measure is a consumer-based measure that focuses on patient experience of care delivery and is limited to domains such as communication and access. The PCPCM PRO PM is a patient assessment of primary care (not a reporting of patient satisfaction or experience) that covers the full scope of primary care as identified by both clinicians and patients. The PCPCM PRO PM is not encounter specific and the CG CAHPS is based on clinical encounters. For these reasons, there is no need to harmonize these two measures. See Appendix A.1 (pages 10-11) for comparison chart.

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.)

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.

Attachment Attachment: Appendix_A.1-637413747614416086.docx

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): American Board of Family Medicine

Co.2 Point of Contact: Jill, Shuemaker, jshuemaker@theabfm.org, 202-524-8313-

Co.3 Measure Developer if different from Measure Steward: Virginia Commonwealth University, The Larry A. Green Center

Co.4 Point of Contact: Rebecca, Etz, rebecca.etz@vcuhealth.org, 804-827-4995-

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.

Here is the list of panel members involved in the measure development:

Rebecca S. Etz PhD

- Developer of the Person-Centered Primary Care Measure
- Co-Director, The Larry A. Green Center for the Advancement of Primary Health Care for the Public Good
- Associate Professor, Department of Family Medicine and Population Health, Virginia Commonwealth University, Richmond, Virginia
- Adjunct Associate Professor, Center for Community Health Integration, Case Western Reserve University, Cleveland, Ohio
- Adjunct Associate Professor, University of Colorado, Department of Family Medicine
- Visiting Visionary, The Institute for Integrative Health

Kurt C. Stange MD, PhD

- Developer and Subject Matter Expert, Person-Centered Primary Care Measure
- Co-Director, The Larry A. Green Center for the Advancement of Primary Health Care for the Public Good
- Director, Center for Community Health Integration, Case Western Reserve University, Cleveland, Ohio
- Professor, Departments of Family Medicine & Community Health, Population & Quantitative Health Sciences, General Medical Sciences and Sociology, and Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, Ohio
- Practicing Family Physician

Stephen J. Zyzanski PhD

- Analyst and Psychometrician, Person-Centered Primary Care Measure
- Center for Community Health Integration
- Professor Emeritus, Departments of Family Medicine & Community Health, Population & Quantitative Health Sciences, Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, Ohio

Martha M. Gonzalez BA

- Data Manager and Analyst, Person-Centered Primary Care Measure
- Data Manager, The Larry A. Green Center for the Advancement of Primary Health Care for the Public Good
- Data Analyst, Department of Family Medicine and Population Health, Virginia Commonwealth University, Richmond, Virginia

Sarah R. Reves MSN, FNP-C, MBA

- Data Analyst and Subject Matter Expert, Person-Centered Primary Care Measure
- Deputy Director, The Larry A. Green Center for the Advancement of Primary Health Care for the Public Good

- Nurse Researcher, Department of Family Medicine and Population Health, Virginia Commonwealth University, Richmond, Virginia
- Assistant Clinical Faculty, Virginia Commonwealth University Department of Family Medicine
- Practicing Family Nurse Practitioner

Jonathan P. O'Neal BA

- Data Analyst, Person-Centered Primary Care Measure
- Program Director, The Larry A. Green Center for the Advancement of Primary Health Care for the Public Good
- Department of Family Medicine and Population Health, Virginia Commonwealth University, Richmond, Virginia

We cross-checked this list with the roster of Primary Care and Chronic Illness committee members and none of them participated in the development of this measure.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released:

Ad.3 Month and Year of most recent revision:

Ad.4 What is your frequency for review/update of this measure? Current plans are to review/update annually, however this will be adjusted as we learn more

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

Ad.6 Copyright statement: This measure has been copyrighted through the Creative Commons and is freely available for use with no fee.

Ad.7 Disclaimers:

Ad.8 Additional Information/Comments: