NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

To navigate the links in the worksheet: Click to go to the link. ALT + LEFT ARROW to return

Purple text represents the responses from measure developers.

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

Brief Measure Information

NQF #: 0089e

Corresponding Measures: 0089

De.2. Measure Title: Diabetic Retinopathy: Communication with the Physician Managing Ongoing Diabetes Care

Co.1.1. Measure Steward: PCPI Foundation

De.3. Brief Description of Measure: Percentage of patients aged 18 years and older with a diagnosis of diabetic retinopathy who had a dilated macular or fundus exam performed with documented communication to the physician who manages the ongoing care of the patient with diabetes mellitus regarding the findings of the macular or fundus exam at least once within 12 months

1b.1. Developer Rationale: Diabetic retinopathy is a prevalent complication of diabetes, estimated to affect 28.5% of diabetic patients in the US. (1) Diabetic Retinopathy is a key indicator of systemic complications of diabetes. (1) Coordination of care between the eye care specialist and the provider managing a patient's ongoing diabetes care is essential in stemming the progression of vision loss. Communication from the eye care specialist to a primary care physician facilitates the exchange of information about the severity and progression of a patient's diabetic retinopathy, adherence to recommended ocular care, need for follow-up visits, and treatment plans. (2) Data from the Diabetes Control and Complications Trial showed that diabetic treatment and maintenance of glucose control delays the onset and slows the progression of diabetic retinopathy. (3)

1. Zhang X, Saaddine JB, Chou CF, et al. Prevalence of diabetic retinopathy in the United States, 2005-2008. JAMA 2010;304: 649–656

2. Storey PP, Murchison AP, Pizzi LT, Hark LA, Dai Y, Leiby BE, Haller JA. Impact of physician communication on diabetic eye examination adherence: Results from a Retrospective Cohort Analysis. Retina. 2016 Jan;36(1):20-7.

3. Aiello LP; DCCT/EDIC Research Group. Diabetic retinopathy and other ocular findings in the diabetes control and complications trial/epidemiology of diabetes interventions and complications study. Diabetes Care. 2014;37(1):17-23.

S.4. Numerator Statement: Patients with documentation, at least once within 12 months, of the findings of the dilated macular or fundus exam via communication to the physician who manages the patient´s diabetic care

S.6. Denominator Statement: All patients aged 18 years and older with a diagnosis of diabetic retinopathy who had a dilated macular or fundus exam performed

S.8. Denominator Exclusions: Denominator Exceptions:

Documentation of medical reason(s) for not communicating the findings of the dilated macular or fundus exam to the physician or other qualified health care professional who manages the ongoing care of the patient with diabetes.

Documentation of patient reason(s) for not communicating the findings of the dilated macular or fundus exam to the physician who manages the ongoing care of the patient with diabetes.

De.1. Measure Type: Process

S.17. Data Source: Electronic Health Records

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

IF Endorsement Maintenance – Original Endorsement Date: Nov 04, 2015 **Most Recent Endorsement Date:** Nov 04, 2015

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?

Preliminary Analysis: Maintenance of Endorsement

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

Criteria 1: Importance to Measure and Report

1a. Evidence

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

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

The developer provides the following evidence for this measure:

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

Summary of prior review in 2014:

- Brief background: This is an eCQM of patients 18+ years of age with diabetic retinopathy that evaluates whether or not the results of a macular or fundus exam were communicated to the clinician managing the patient's diabetes within the last 12 months.

- This measure was reviewed in the Eye Care and Ear, Nose, and Throat Conditions 2014-2016 project. The evidence at that time citd guidelines from that state "Close partnership with the primary care physician is important to make sure that the care of the patient is optimized: Strong recommentation.
- Level III- non-analytic studies; Support for the recommendation is inferred from observational studies and RCTs supporting tight control of blood glucose in slowing the progression of Diabetic Retinopathy. However, there are no studies that directly support a close partnership with the primary care physician, so the recommendation is listed as being supported by Level III, non-analytic studies.
- The EENT committee noted that there was little evidence showing that communication with a primary physician will save vision; however, Committee members agreed that if the eye care specialist is aware that the diabetes is poorly controlled they can encourage the patient toward better glucose control, which will lead to less progression of the retinopathy.

Changes to evidence from last review

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

☒ **The developer provided updated evidence for this measure:**

Updates:

The developer provided updated guidelines from the AAO 2017 Preferred Practice Pattern® Guidelines. Diabetic Retinopathy.

- Recommendation: Close partnership with the primary care physician is important to make sure that the care of the patient is optimized: III; Good; Strong
- Recommendation: Ophthalmologists should communicate the ophthalmologic findings and level of retinopathy with the primary care physician as well as the need for optimizing metabolic control: III; Good; Strong

Exception to evidence

- The Committee may wish to consider an exception for the evidence for this measure. It may not be feasible to develop an outcome measure assessing this concept, expert opinion supports the measure, and there is no evidence of harm from them measure.

Questions for the Committee:

- The evidence provided by the developer is updated and directionally the same compared to that for the previous NQF review. Does the Committee agree there is no need for repeat discussion and vote on Evidence?
- *For possible exception to the evidence criteria:*
    - *Are there, or could there be, performance measures of a related health outcome, OR evidence-based intermediate clinical outcomes, intervention/treatment?*
    - *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*
    - *Does the SC agree that it is acceptable (or beneficial) to hold providers accountable without empiric evidence?*

Guidance from the Evidence Algorithm

Process measure; no SR; evidence about something other than what is being measured (Box 3) →Box 7 – no empirical evidence → (Box 10) If no, consider "insufficient evidence with exception" (Box 11, 12)

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

RATIONALE:

The evidence to support this measure is inferred from studies supporting glucose control in slowing the progression of diabetic retinopathy. Evidence for this process measure should demonstrate that

communication with other clinicians improves patient outcomes. However, given the evidence that communication could not be harmful and could be potentially beneficial, the Committee may with to consider the evidence exception for this measure.

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

Maintenance measures – increased emphasis on gap and variation

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

Data from the CMS QPP and PQRS programs:

2017 74.78%

2016 77.3%

2015 74.8%

2014 81.0%

Disparities

The developer noted this measure is used in a federal reporting program but disparities data is not available to analyze and report. The developer provided evidence from the literature noting evidence of disparities in the incidence, prevalence, assessment and treatment of diabetic retinopathy but noted they are unaware of studies identifying disparities in examined disparities in the communication between specialists and clinicians managing the ongoing care of patients with diabetes.

Questions for the Committee:

- Is there a gap in care that warrants a national performance measure?
- Are you aware of evidence that disparities exist in this area of healthcare?

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?

- This is a process measure. The evidence applies tangentially to slowing vision loss from diabetic retinopathy. The evidence to support this measure is inferred from studies supporting glucose control in slowing the progression of diabetic retinopathy. No new studies change the evidence base for the measure as written.

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?

- Based on the sample of 7,951 included providers, the mean performance rate is 0.65, the median performance rate is 0.69 and the mode is 1.00. The standard deviation is 0.26. The range of the performance rate is 1.00, with a minimum rate of 0.00 and a maximum rate of 1.00. The interquartile range is 0.39 (0.86–0.48). The range of performance from 0.00 to 1.00 and the fairly even spread of the data provided in the percentile chart suggests that there exists clinically meaningful variation across providers' performance. Disparity data was not provided directly, but it can be inferred that improvement

in performance will address underlying disparity that exists in the care of patients with diabetic retinopathy. Disparity data was not provided directly, but it can be inferred that improvement in performance will address underlying disparity that exists in the care of patients with diabetic retinopathy.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications **and** Testing

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

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

Reliability

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

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

Validity

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

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

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.

eCQM Technical Advisor(s) review:

Submitted measure is an HQMF compliant eCQM	The submitted eCQMspecifications follow the industry accepted format for eCQM (HL7 Health Quality Measures Format (HQMF)).
Documentation of HQMF,QDM, or CQL limitations	N/A – All components in the measure logic of the submitted eCQM are represented using the HQMF, QDM, or CQL standards;
Value Sets	The submitted eCQM specifications uses existing value sets when possible and uses new value sets that have been vetted through the VSAC
Measure logic is unambiguous	Submission includes test results [from a simulated data set] demonstrating the measure logic can be interpreted precisely and unambiguously.
Feasibility Testing	Number of data elements included in measure calculation: 17 Number of data elements scoring less than 3 on scorecard: 7 LevelofSeverityofRetinopathyFindings_Communication:FromProviderToProviderNotDone • low scoring domains: availability, standards

	• The developer also noted that there is a place to assign communication and mark any return information. ◦ comments on availability domain: Data element is indirectly obtained by purposefully not choosing a severity. ◦ comments on standards domain: Standards may or may not be available in EHR as data element is indirectly obtained OutpatientConsultation_EncounterPerformed • low scoring domains: availability, accuracy, standards, workflow • The developer noted that they know we can mark post op care that is outpatient. MacularExam_DiagnosticStudyPerformed • low scoring domains: availability, accuracy, standards, workflow • The developer noted that quality codes for AMD are again the same as above (A provider can set up a trigger to then select which quality code they would like) but not a specific way to mark this. CareServicesinLong-TermResidentialFacility_EncounterPerformed • low scoring domains: availability, accuracy, standards, workflow The developer noted that data element never performed based on criteria and that most OD's do not perform this service so as such the EHR does not have a way of tracking. MacularEdemaFindingsPresent_Communication:FromProviderToProviderNotDone • low scoring domains: availability, standards • The developer also noted that a provider can set up a trigger to then select which quality code they would like. ◦ comments on availability domain: Data element is indirectly obtained by purposefully not documenting macular edema findings present ◦ comments on standards domain: Standards may or may not be available in EHR as data element is indirectly obtained MacularEdemaFindingsAbsent_Communication:FromProviderToProviderNotDone • low scoring domains: availability, standards • The developer also noted that a provider can set up a trigger to then select which quality code they would like. ◦ comments on availability domain: Data element is indirectly obtained by purposefully not documenting macular edema findings absent ◦ comments on standards domain: Standards may or may not be available in EHR as data element is indirectly obtained NursingFacilityVisit_EncounterPerformed • low scoring domains: availability, accuracy, standards, workflow • The developer noted that data element never performed based on criteria and that most OD's do not perform this service so as such the EHR does not have a way of tracking.

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

Evaluators: Staff

Review A

Evaluation of Reliability and Validity (and composite construction, if applicable):

- Measure was testing using pooled data for individual and group clinicians, but is specified for both. NQF requires separate analysis for these two types of providers. Therefore, this measure is considered insufficient until developer provides separate analyses.

- Developer has specified the measure for outpatient, post acute care and domicillary settings, but these analyses were not conducted separately, nor is the care setting clearly articulated in the submission.
- The specifications of the measure are otherwise precise and unambiguous. Score level reliability testing was conducted using a signal to noise approach. Results of this testing demonstrated high reliability for the pooled data. To test validity, the developer conducted a correlation analysis with Diabetes: Eye Exam (PQRS #117). Results of that analysis showed a very weak correlation.

Questions for the Committee regarding reliability:

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

Questions for the Committee regarding validity:

- The Committee will need to discuss validity. Results of empiric valdidity testing were weak. Does the Committee believe the developer should be given an exception for face validity testing only?

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

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

RATIONALE:

- Measure was evaluated as insufficient by NQF staff
- Measures must be tested according to specification
    - Measure developer did not separate analyses by level of analysis and care setting

Evaluation A: Scientific Acceptability

Measure Number: 0089e

Measure Title: Diabetic Retinopathy: Communication with the Physician Managing Ongoing Diabetes Care

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

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

   - Measure exceptions include documentation of "medical reason(s)" or "patient reasons" for not communicating results; NQF prefers specific, auditable exclusions rather than broad, non-specific categories
   - Measure is specified for multiple levels of analysis, but not tested appropriately. Evaluations will be for a subset of specifications, described below:
     - Measure was testing using pooled data for individual and group clinicians, but is specified for both. NQF requires separate analysis for these two types of providers. Therefore, this measure is only under consideration for endorsement at the group/practice level. For consideration of individual providers, separate testing must be submitted.
     - Developer has specified the measure for outpatient, post acute care and domicillary settings, but these analyses were not conducted separately, nor is the care setting clearly articulated in the submission. The measure will therefore only be considered for outpatient specification. For additional care settings, separate testing must be submitted.

RELIABILITY: TESTING

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

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

   ☐ **Yes** ☐ **No**

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

   Submission document: Testing attachment, section 2a2.2

   - Developer must separate the submission into level of analysis and care setting.
   - Reliability of the computed measure score was measured as the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in provider performance and the noise is the total variability in measured performance.
   - Reliability was tested using a beta-binomial model. The beta-binomial model assumes the provider performance score is a binomial random variable conditional on the provider's true value that comes from the beta distribution. The beta distribution is usually defined by two parameters, alpha and beta. Alpha and beta can be thought of as intermediate calculations to get to the needed variance estimates.
   - A reliability equal to zero implies that all the variability in a measure is attributable to measurement error. A reliability equal to one implies that all the variability is attributable to real differences in provider performance. A reliability of 0.70 – 0.80 is generally considered the acceptable threshold for reliability, 0.80 – 0.90 is considered high reliability, and 0.90 – 1.0 is considered very high.

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

   Submission document: Testing attachment, section 2a2.3

   - The data set included data from from 8,832 providers reporting on this measure through the EHR reporting option for CMS's PQRS in 2016 and reflects a combination of individual provider data and group data. Of those 8,832 providers, 7,951 had all the required data elements and at least one quality reporting event for a total of 891,902 quality events. For this measure, 90 percent of providers are included in the analysis, and the average number of quality reporting events are 112 for the remaining 891,902 events. The range of quality reporting events for the 7,951 providers included is from 1 to 1,856.
   - The developer conducted analyses for providers with at least one quality reported event and an analysis limiting to physicians with 10 or more events.
     - The average reliability including providers with 10 ore more quality reporting events is 0.91.
     - The average reliability including providers with at least one event is 0.72.

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 <u>all</u> testing results):

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

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

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

    ☒ **Insufficient** (NOTE: Should rate <u>INSUFFICIENT</u> 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.**

    Precise specifications (Box 1) →empiric reliability testing (Box 2) →performance score testing (Box 4) →appropriate method of testing (Box 5) → Use Box 6 to determine rating
    Based on the reliability statistics and scope of testing we cannot determine reliability at this time.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

- During the previous review of this measure the Committee did not note any concerns with the measure exclusions and noted that the exclusion for patient reason is needed because some patients do not want their information sent to their primary care provider or there is no primary care provider.

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

    Submission document: Testing attachment, section 2b4.

    - Based on the sample of 7,951 included providers, the mean performance rate is 0.65, the median performance rate is 0.69 and the mode is 1.00. The standard deviation is 0.26. The range of the performance rate is 1.00, with a minimum rate of 0.00 and a maximum rate of 1.00. The interquartile range is 0.39 (0.86–0.48).

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.

    - No data on comparability of measure results from different data sources was provided.

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

    Submission document: Testing attachment, section 2b6.

    - The measure developer noted that the dataset provided by CMS did not contain missing data so this test was not performed. However, missing data may have been rejected when submitted to CMS in which case those values would not be counted towards measure performance. However, the developer had no indication that missing data was systematic.

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**

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

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

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

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

VALIDITY: TESTING

19. **Validity testing level: ☒ Measure score ☐ Data element ☐ Both**
20. **Method of establishing validity of the measure score:**

    ☐ **Face validity**

    ☒ **Empirical validity testing of the measure score**

    ☐ **N/A (score-level testing not conducted)**
21. **Assess the method(s) for establishing validity**

    Submission document: Testing attachment, section 2b2.2

    - The developer provided updated validity testing to meet the requirement that face validity testing is not acceptable for maintenance measures.
        - Previous testing: Validity of the measure score was assessed by systematic assessment of face validity by an expert panel of 16 members who strongly agreed that the measure could distinguish quality of care.
    - To test the measure empirically, the developer conducted a correlation analysis between this measure and Diabetes: Eye Exam (PQRS #117), and Diabetic Retinopathy: Documentation of Presence or Absence of Macular Edema and Level of Severity of Retinopathy (PQRS #018) due to similarities in the domain and patient population.
    - The developer hypothesized that there exists a positive association of scores between providers who performed a dilated macular or fundus exam with a documented communication to the physician managing ongoing care of the patient with diabetes mellitus regarding the findings at least once within 12 months and those who performed an eye exam (retinal) on patients with diabetes (type 1 and 2).
    - The developer also hypothesized that also hypothesize that there exists a positive association of scores between providers who performed a dilated macular or fundus exam with a documented communication to the physician managing ongoing care of the patient with diabetes mellitus regarding the findings at least once within 12 months and those providers who performed a dilated macular or fundus exam on those with a diagnosis of diabetic retinopathy and included the documentation of the level or severity of retinopathy and the presence or absence of macular edema during one or more office visits within 12 months.
    - Datasets were reviewed to identify shared providers based on NPI and TIN identifiers. Comparing performance scores of those shared providers, the empirical analysis uses regression with dataset 1 as the outcome and dataset 2 as the predictor. Results identify the multiple R value (the correlation coefficient) and P-value of the regression variables to assess the association between performance scores of these shared providers.
22. **Assess the results(s) for establishing validity**

    Submission document: Testing attachment, section 2b2.3

    - The developer used the following guidance for interpreting correlation:
        - 0.80 – 1.00 Very Strong
        - 0.60 – 0.79 Strong
        - 0.40 – 0.59 Moderate
        - 0.20 - 0.39 Weak
        - 0 – 0.19 Very Weak
    - PQRS #117

- Coefficient of correlation = 0.03
- Alpha level = 0.05
- P-value = 0.003
- Number of shared providers based on NPI and TIN identifiers = 10474

- PQRS #018
  - Coefficient of correlation = 0.53
  - Alpha level = 0.05
  - P-value < 0.001
  - Number of shared providers based on NPI and TIN identifiers = 7742
- The measure showed a very weak positive correlation with Diabetes: Eye Exam(PQRS #117) and the correlation is statistically significant at the 95% significance level. The measure showed a moderate correlatin with Diabetic Retinopathy: Documentation of Presence or Absence of Macular Edema and Level of Severity of Retinopathy (PQRS #018).

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 <u>are</u> threats to validity and/or relevant threats to validity were <u>not assessed OR</u> 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 <u>is required</u>; 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.**

    Guidance from the validity algorithm:

- All threats to validity addressed (Box 1) If yes, → Empiric validity testing conducted (Box 2), if yes, →Testing with measure score (Box 5) → Method appropriate (Box 6) → Use Box 7 to rate validity
- Results show moderate certainity the measure is a valid indicator of quality

ADDITIONAL RECOMMENDATIONS

27. **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.**

    N/A

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?

- none

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

- None. The average reliability including providers with at least one quality reporting event is 0.91.

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

- Low validity with diabetic retinopathy testing. Re; validity, PQRS #117 Coefficient of correlation = 0.03 and PQRS #018 Coefficient of correlation = 0.53.

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?

- see response to #4

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?

- n/a

Criterion 3. Feasibility

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

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

- The measure is specified for an EHR.
- Developer indicates all data elements are in defined fields in electronic health records
- However, per feasibility scorecard report some data elements are not currently captured in structured fields. Data elements for communication: from provider to provider, not done are currently unavailable in structured fields.

Questions for the Committee:

- Are the required data elements routinely generated and used during care delivery?
- Are the required data elements available in electronic form, e.g., EHR or other electronic sources?

- Is the data collection strategy ready to be put into operational use?
- Does the eCQM Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?

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?

- Normally documentation of the degree of diabetic retinopathy and macular edema is performed in outpatient charting, and is typically well-defined in an EHR. Exactly what constitutes communication with primary care is addressed in the measure.

Criterion 4: Usability and Use

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

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

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

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

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

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

OR

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

Accountability program details Merit-based Incentive Payment System (MIPS)

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

Feedback on the measure by those being measured or others

- The developer noted they maintain several pathways for feedback on the measure including the use of topic-specific technical expert panels during the measure development and during the annual maintenance process as well as feedback via an online public comment and an email-based process set up to receive measure inquiries from implementers.
- Feedback received by the developer focused on clarifications about the types of communication that would et the measure. To address this feedback the developer added a definition to the measure to

clarify that communication woul deed to be documented and must include the findings of the dilated macular or fundus exam. Communication be verbal or via a written or electronic communication.

- The developer also received feedback requesting clarification about the information that should be included in the communication. As a result, they added a definition stating that the communication would include level of severity of retinopathy (eg, mild nonproliferative, moderate nonproliferative, severe nonproliferative, very severe nonproliferative, proliferative) AND the presence or absence of macular edema.

Additional Feedback:

- The developer noted that other users provided similar feedback requesting clarification about the types of communication required and what information the communication should include in order to meet the measure.

***Questions for the Committee**:*

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

Preliminary rating for Use: ☒ Pass ☐ No Pass

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

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

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

Improvement results

- The measure developer noted that theintent of this measure is to help improve the care of patients with diabetic retinopathy.
- The developer found that CMS data showed a performance rate of 74.78% in 2017 which marked a decrease from the rate of 81.0% in 2014.

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

Unexpected findings (positive or negative) during implementation

- The developer noted they did not know of any unexpected findings during implementation.

Potential harms

- No harms have been identified from use of the measure.

Additional Feedback: N/A

***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?

- MIPS, IRIS registry. PCPI has a forum for this process and is detailed in the submission

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.

- It is unclear to me how the exact delineation of the level of retinopathy and presence or absence of macular edema in clinical communication with the provider would directly impact a primary care physician's management of diabetes, although small studies quoted in the proposal have suggested positive impacts:

Criterion 5: Related and Competing Measures

- Related measures:
  - 0055: Comprehensive Diabetes Care: Eye Exam (retinal) performed
- There are no competing measures.

Harmonization

- Measure #0055 evaluates the percentage of patients 18-75 years of age with diabetes who had an eye exam (retinal) performed. While the population is similar, this measure requires that a dilated macular or fundus exam be performed, and the results communicated to the physician who manages the ongoing care of the patient with diabetes.

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?

- Measure #0055 evaluates the percentage of patients 18-75 years of age with diabetes who had an eye exam (retinal) performed, which is similar to the denominator of this measure. Other eye related measures (for example, vision screening in the pediatrician's office) have blended screening rates with "referral if indicated" into the numerator. I would be interested to know if most other quality measures look at the screening/examination as one measured performance, and referral/communication as a separate measure, or whether it is more standard to combine the two.

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 06/12/2019

- **No NQF Members have submitted support/non-support choices as of this date.**

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

DR_evidence_attachment_2019_0089e-636912043969856581.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*)**:** 0089

Measure Title: Diabetic Retinopathy: Communication with the Physician Managing Ongoing Diabetes Care

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

Date of Submission: 4/9/2019

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

Outcome

☐ Outcome:

☐Patient-reported outcome (PRO):

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

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

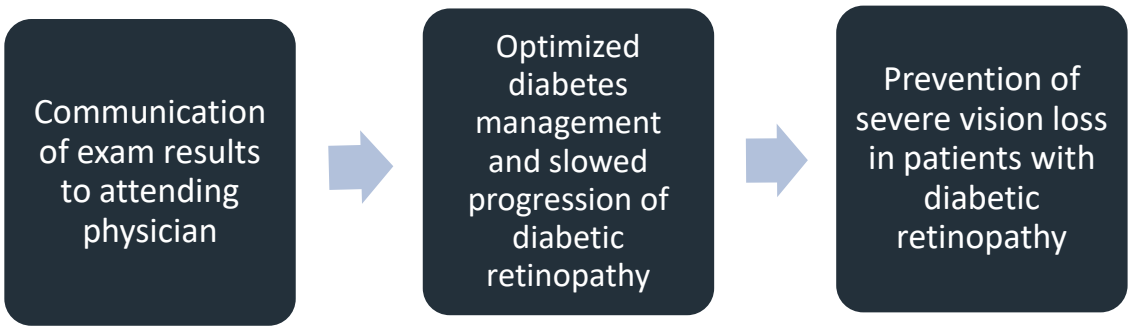
☒ Process: Communication of dilated macular or fundus exam results to physician managing diabetes care of patient

☐ Appropriate use measure:

☐ Structure:

☐ Composite:

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



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

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

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

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

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

☒ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Source of Systematic Review: • Title • Author • Date • Citation, including page number • URL	• Title: Preferred Practice Patterns Committee. Diabetic Retinopathy • Author: American Academy of Ophthalmology • Date: 2014 • Citation: American Academy of Ophthalmology. Preferred Practice Patterns Committee. Diabetic Retinopathy. San Francisco, CA: American Academy of Ophthalmology. 2014 (updated from 2003) • URL: http://one.aao.org/preferred-practice-pattern/diabetic-retinopathy-ppp--2014 • Title: Preferred Practice Pattern® Guidelines. Diabetic Retinopathy. • Author: American Academy of Ophthalmology Retina/Vitreous Panel • Date: Updated December 2017

	• Citation: American Academy of Ophthalmology Retina/Vitreous Panel. Preferred Practice Pattern® Guidelines. Diabetic Retinopathy. San Francisco, CA: American Academy of Ophthalmology; 2017, page 29. • URL: https://www.aao.org/ppp
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.	**2014 AAO Preferred Practice Patterns Committee Diabetic Retinopathy: Guideline** Recommendation: Close partnership with the primary care physician is important to make sure that the care of the patient is optimized **2017 Preferred Practice Pattern® Guidelines. Diabetic Retinopathy** Recommendation: Close partnership with the primary care physician is important to make sure that the care of the patient is optimized: III; Good; Strong Recommendation: Ophthalmologists should communicate the ophthalmologic findings and level of retinopathy with the primary care physician as well as the need for optimizing metabolic control: III; Good; Strong
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	**2014 AAO Preferred Practice Patterns Committee** **Diabetic Retinopathy** Level of evidence: III Definition: Nonanalytic studies (e.g., case reports, case series) 2017 Preferred Practice Pattern® Guidelines. Diabetic Retinopathy **NOTE:** grade assigned to evidence is the same as the 2014 version noted above Level of evidence: III Definition: Nonanalytic studies (e.g., case reports, case series)
Provide all other grades and definitions from the evidence grading system	**2014 AAO Preferred Practice Patterns Committee** **Diabetic Retinopathy: Evidence grading system** **I++** High-quality meta-analyses, systematic reviews of randomized controlled trials (RCTs), or RCTs with a very low risk of bias **I+** Well-conducted meta-analyses, systematic reviews of RCTs, or RCTs with a low risk of bias **I-** Meta-analyses, systematic reviews of RCTs, or RCTs with a high risk of bias **II++** High-quality systematic reviews of case-control or cohort studies High-quality case-control or cohort studies with a very low risk of confounding or bias and a high probability that the relationship is causal **II+** Well-conducted case-control or cohort studies with a low risk of confounding or bias and a moderate probability that the relationship is causal **II-** Case-control or cohort studies with a high risk of confounding or bias and a significant risk that the relationship is not causal **III** Nonanalytic studies (e.g., case reports, case series) 2017 Preferred Practice Pattern® Guidelines. Diabetic

	Retinopathy NOTE: Same as 2014 guideline evidence grading system outlined above
Grade assigned to the **recommendation** with definition of the grade	**2014 AAO Preferred Practice Patterns Committee Diabetic Retinopathy:** **Grade assigned to the evidence:** Good; Strong Definitions: Good quality - Further research is very unlikely to change our confidence in the estimate of effect Strong recommendation- Further research is very unlikely to change our confidence in the estimate of effect; Desirable effects of intervention clearly outweigh the undesirable effects 2017 Preferred Practice Pattern® Guidelines. Diabetic Retinopathy Good; Strong Definitions: Good quality - Further research is very unlikely to change our confidence in the estimate of effect Strong recommendation - Used when the desirable effects of an intervention clearly outweigh the undesirable effects or clearly do not
Provide all other grades and definitions from the recommendation grading system	**2014 AAO Preferred Practice Patterns Committee Diabetic Retinopathy** Recommendations for care are formed based on the body of the evidence. The body of evidence quality ratings are defined by GRADE as follows: Good quality - Further research is very unlikely to change our confidence in the estimate of effect Moderate quality - Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate Insufficient quality - Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate. Any estimate of effect is very uncertain Key recommendations for care are defined by GRADE as follows: Strong recommendation: Used when the desirable effects of an intervention clearly outweigh the undesirable effects or clearly do not Discretionary recommendation: Used when the trade-offs are less certain-either because of low-quality evidence or because evidence suggests that desirable and undesirable effects are closely balanced 2017 Preferred Practice Pattern® Guidelines. Diabetic Retinopathy NOTE: Same as 2014 guideline body of evidence ratings outlined above
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	**2014 AAO Preferred Practice Patterns Committee Diabetic Retinopathy** **Quantity:** 10 Randomized Controlled Trials, 5 observational studies **Quality:** Level III - non-analytic studies; Support for the recommendation is inferred from observational studies and RCTs supporting tight control of blood glucose in slowing the progression of Diabetic Retinopathy. However, there are no studies that directly support a close partnership with the primary care physician, so the recommendation is listed as being supported by Level III, non-analytic studies. 2017 Preferred Practice Pattern® Guidelines. Diabetic Retinopathy

	Within the context of prevention and early detection of diabetic retinopathy, the AAO 2017 guidelines reference several observational studies highlighting gaps in referrals to ophthalmic care, and the screening and early detection of diabetic retinopathy to facilitate early treatment and to prevent visual loss. The AAO guidelines cite a physician survey of practice patterns and a population-based cross-sectional study to highlight variations or noncompliance with guidelines for vision care. Additionally, the guidelines reference eight (8) studies with data supporting early detection of diabetic retinopathy as a process that may lead to improved patient outcomes in terms of the prevention of vision loss. Results of a systematic review, a prospective comparative observational case series study, 2 retrospective cross-sectional studies, and 4 prospective observational studies support the use of photography as a screening tool for diabetic retinopathy. The 2017 AAO guidelines cite 2 studies, including a prospective observational case series and a prospective cohort study on the use of digital cameras to determine level of diabetic retinopathy to optimize disease management. The 2 studies concluded that digital photographs as a suitable alternative for annual retinal examination. Also cited in the 2017 guidelines were 2 studies with data on the use of optical coherence tomography (OCT) for detecting macular edema. The studies included a literature review, and a comparative observational study of retinal thickness measurements in eyes with diabetic macular edema. Two studies also showed a positive association between screening programs and diabetic retinopathy assessment rates. A retrospective observational study showed that participation in a telehealth program was related to the use of recommended eye care. The study concluded that such programs may address aspects of eye care essential to reduce the complications related to diabetic retinopathy including vision loss. A prospective observational case control study showed that a teleretinal imaging group had improved adherence and improved rates of diabetic retinopathy assessment. A report on the Diabetes Control and Complications Trial highlighted the recommendations that intensive therapy intended to optimize blood glucose values reduces the risk of both the development and the progression of diabetic retinopathy.
Estimates of benefit and consistency across studies	**2014 AAO Preferred Practice Patterns Committee Diabetic Retinopathy** Quantitative estimate not provided; the benefit of a close partnership with the primary care physician is that the primary care physician can use the results of eye exam to convey the importance of tight glycemic control to the patient. 2017 Preferred Practice Pattern® Guidelines. Diabetic Retinopathy The referenced studies in support of the recommendations highlight the potential for and the benefits of the early detection and of slowing the progression of diabetic retinopathy. The benefits identified within the studies include optimizing blood glucose levels and adherence to eye care recommendations intended to prevent vision loss. These processes and improved visual outcomes would benefit from the communication between the clinicians providing eye care and the clinician managing the ongoing treatment of the patient with diabetes. The clinical recommendations within the AAO 2017 guidelines are designated as being based on good quality evidence

	meaning that research is very unlikely to change confidence in the estimate of effect.
What harms were identified?	**2014 AAO Preferred Practice Patterns Committee Diabetic Retinopathy** No harms were discussed related to having a close partnership with the primary care physician. 2017 Preferred Practice Pattern® Guidelines. Diabetic Retinopathy No harms were discussed related to having a close partnership with the primary care physician managing the ongoing care of patients with diabetes.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	**2014 AAO Preferred Practice Patterns Committee Diabetic Retinopathy** No updates to the body of evidence have been found for this report. There are studies showing that communication between the specialist and primary care physician results in greater patient satisfaction and better management of disease overall, but nothing specific to diabetic retinopathy. We conducted a search using the Medical Subject Headings (MeSH®) terms "Diabetic Retinopathy" [Mesh] AND "Communication"[Mesh], "Interprofessional Collaboration"[Mesh] AND "Diabetes"[Mesh] to identify articles published after 2017. None of the articles published in this timeframe were relevant to the body of evidence that supports this measure nor would they change the recommendation to maintain communication with the primary care physician to optimize the care of patients with diabetes and diabetic retinopathy.

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.

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)*

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

Diabetic retinopathy is a prevalent complication of diabetes, estimated to affect 28.5% of diabetic patients in the US. (1) Diabetic Retinopathy is a key indicator of systemic complications of diabetes. (1) Coordination of care between the eye care specialist and the provider managing a patient's ongoing diabetes care is essential in stemming the progression of vision loss. Communication from the eye care specialist to a primary care physician facilitates the exchange of information about the severity and progression of a patient's diabetic retinopathy, adherence to recommended ocular care, need for follow-up visits, and treatment plans. (2) Data from the Diabetes Control and Complications Trial showed that diabetic treatment and maintenance of glucose control delays the onset and slows the progression of diabetic retinopathy. (3)

1. Zhang X, Saaddine JB, Chou CF, et al. Prevalence of diabetic retinopathy in the United States, 2005-2008. JAMA 2010;304: 649–656

2. Storey PP, Murchison AP, Pizzi LT, Hark LA, Dai Y, Leiby BE, Haller JA. Impact of physician communication on diabetic eye examination adherence: Results from a Retrospective Cohort Analysis. Retina. 2016 Jan;36(1):20-7.

3. Aiello LP; DCCT/EDIC Research Group. Diabetic retinopathy and other ocular findings in the diabetes control and complications trial/epidemiology of diabetes interventions and complications study. Diabetes Care. 2014;37(1):17-23.

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.*

Diabetic Retinopathy 0089e: EHR

2016 EHR data from the PQRS program was provided to the PCPI by CMS for the purposes of testing the measure. The data are analyzed for the time period January 2016 through December 2016 and include 891,902 quality events. The mean performance rate is 0.65, the standard deviation is 0.26, the minimum is 0.00, the maximum is 1.00, and the interquartile range is 0.39 (0.86 – 0.48). Performance Scores by Decile: (1st,0.06; 2nd,0.29, 3rd,0.44; 4th,0.56; 5th,0.65; 6th, 0.73; 7th,0.80; 8th,0.88; 9th,0.96; 10th,1.00)

CMS published the following data in the 2017 QPP and 2016 PQRS Experience Reports. Experience report data does not differentiate average performance rates between EHR, registry, and claims. The average performance rates on the Diabetic Retinopathy: Communication with the Physician Managing Ongoing Diabetes Care measure between 2014-2017 are listed below. It is important to note that both QPP and PQRS are voluntary reporting programs in which eligible providers choose which measure(s) to report, which is reflected in the reporting rate among those eligible to report on this measure. The performance scores listed below are not consistently derived from a nationally representative sample. Nevertheless, the performance scores indicate a gap in care as the average performance rate for the last 4 years range from 74.7% to 81.0%

Year	Average Performance Rate	Reporting Rate
2017	74.78%	-
2016	77.3%	15.4%
2015	74.8%	21.9%
2014	81.0%	25.6%

2017 Quality Payment Program Experience Report. Available at

https://qpp-cm-prod-content.s3.amazonaws.com/uploads/492/2017%20QPP%20Experience%20Report%20Appendix.zip

2016 Reporting Experience Including Trends (2007-2016), Physician Quality Reporting System. Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/Downloads/2016-Appendix.xlsx

2018 Benchmark Report Data

Submission method	Claims	EHR	Registry/QCDR
Std Deviation	13.1	25	29.9
Average	96.6	66.8	76.5
Decile 3	-	46.15-56.85	52.00-72.40
Decile 4	-	56.86-65.13	72.41-81.47
Decile 5	-	65.14-72.38	81.48-90.76
Decile 6	-	72.39-78.21	90.77-96.54
Decile 7	-	78.22-84.27	96.55-99.99
Decile 8	-	84.28-89.93	-
Decile 9	-	89.94-95.41	-
Decile 10	-	>=95.42	100

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.

Recent studies have found less than optimal performance in communication between providers. A retrospective cohort study found that written communication from an ophthalmologist to a primary care physician is associated with improved adherence to recommended diabetic eye examinations. Despite these findings, communication from the ophthalmologist to the primary care provider only occurred for approximately 15% of patients. (1)

A survey of more than 4,000 physicians found similar gaps in communication between specialists and primary care providers, in general. The study found that approximately 81% of specialists stated that they send consultation reports with results to the referring provider, but only 62% of referring providers stated they received such information. The study also reported that physicians who did not receive timely consultation communications were more likely to report a threat to their ability to provide quality care to their patients. (2)

1. Storey PP, Murchison AP, Pizzi LT, Hark LA, Dai Y, Leiby BE, Haller JA. Impact of physician communication on diabetic eye examination adherence: Results from a Retrospective Cohort Analysis. Retina. 2016 Jan;36(1):20-7.

2. O´Malley AS, Reschovsky JD. Referral and consultation communication between primary care and specialist physicians: finding common ground. Arch Intern Med. 2011 Jan 10;171(1):56-65.

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.*

While this measure is included in a federal reporting program(s), the program does not provide disparities data to analyze and report. In Section 1b.5 below, we provide disparities data reported in the literature.

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

Various studies have examined disparities in the incidence, prevalence, assessment and treatment of diabetic retinopathy. (1-4) We are unaware of studies that have examined disparities in the communication between specialists and clinicians managing the ongoing care of patients with diabetes.

1. Varma R, Choudhury F, Klein R, Chung J, Torres M, Azen SP; Los Angeles Latino Eye Study Group. Four-year incidence and progression of diabetic retinopathy and macular edema: the Los Angeles Latino Eye Study. Am J Ophthalmol. 2010 May;149(5):752-61.e1-3.

2. Luo H, Bell RA, Garg S, Cummings DM, Patil SP, Jones K. Trends and racial/ethnic disparities in diabetic retinopathy among adults with diagnosed diabetes in North Carolina, 2000-2015. N C Med J. 2019 Mar-Apr;80(2):76-82.

3. Hwang J, Rudnisky C, Bowen S, Johnson JA. Income-related inequalities in visual impairment and eye screening services in patients with type 2 diabetes. J Public Health. 2016 Dec 2;38(4):e571-e579.

4. Fathy C, Patel S, Sternberg P Jr, Kohanim S. Disparities in adherence to screening guidelines for diabetic retinopathy in the United States: A Comprehensive Review and Guide for Future Directions. Semin Ophthalmol. 2016;31(4):364-77.

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):*

Ears, Nose, Throat (ENT), Endocrine, Endocrine : Diabetes, Eye Care

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

Health and Functional Status : Change

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

Elderly

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

The measure specifications are attached to this submission. Additional measure details may be found at: eCQI Resource Center https://ecqi.healthit.gov/eligible-professional-eligible-clinician-ecqms. Value set details at VSAC: https://vsac.nlm.nih.gov/.

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 an eMeasure **Attachment:** CMS142v7.zip

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)*

Attachment **Attachment:** CMS142_NQF0089_ValueSets_20180917.xlsx

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

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

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

Not an instrument-based measure

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

Yes

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.

Supporting guidelines and coding value sets included in the measure are reviewed on an annual basis. This annual review has resulted in minor changes to the value sets, to account for updates to the coding terminologies for existing data elements, as well as removal of coding related to 'unspecified eye,' as these codes were determined by clinical experts to potentially lead to poor documentation practices. Measure specifications are annually updated to align with any changes to the standards or tools used to support electronic measurement.

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

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

Patients with documentation, at least once within 12 months, of the findings of the dilated macular or fundus exam via communication to the physician who manages the patient´s diabetic care

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

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

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

DEFINITIONS:

Communication - May include documentation in the medical record indicating that the findings of the dilated macular or fundus exam were communicated (eg, verbally, by letter) with the clinician managing the patient´s diabetic care OR a copy of a letter in the medical record to the clinician managing the patient´s diabetic care outlining the findings of the dilated macular or fundus exam.

Findings - Includes level of severity of retinopathy (eg, mild nonproliferative, moderate nonproliferative, severe nonproliferative, very severe nonproliferative, proliferative) AND the presence or absence of macular edema.

GUIDANCE:

The measure, as written, does not specifically require documentation of laterality. Coding limitations in particular clinical terminologies do not currently allow for that level of specificity (ICD-10-CM includes laterality, but ICD-9-CM and SNOMED-CT do not uniformly include this distinction). Therefore, at this time, it is not a requirement of this measure to indicate laterality of the diagnoses, findings or procedures. Available

coding to capture the data elements specified in this measure has been provided. It is assumed that the eligible professional or eligible clinician will record laterality in the patient medical record, as quality care and clinical documentation should include laterality.

The communication of results to the primary care physician providing ongoing care of a patient´s diabetes should be completed soon after the dilated exam is performed. Eligible professionals or eligible clinicians reporting on this measure should note that all data for the reporting year is to be submitted by the deadline established by CMS. Therefore, eligible professionals or eligible clinicians who see patients towards the end of the reporting period (ie, December in particular), should communicate the results of the dilated macular exam as soon as possible in order for those patients to be counted in the measure numerator. Communicating the results as soon as possible after the date of the exam will ensure the data are included in the submission to CMS.

HQMF eCQM developed and is attached to this submission in fields S.2a and S. 2b.

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

All patients aged 18 years and older with a diagnosis of diabetic retinopathy who had a dilated macular or fundus exam performed

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

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

Time Period for Data Collection: 12 consecutive months

HQMF eCQM developed and is attached to this submission in fields S.2a and S. 2b.

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

Denominator Exceptions:

Documentation of medical reason(s) for not communicating the findings of the dilated macular or fundus exam to the physician or other qualified health care professional who manages the ongoing care of the patient with diabetes.

Documentation of patient reason(s) for not communicating the findings of the dilated macular or fundus exam to the physician who manages the ongoing care of the patient with diabetes.

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

Time Period for Data Collection: During the encounter within the 12-month period

Exceptions are used to remove a patient from the denominator of a performance measure when the patient does not receive a therapy or service AND that therapy or service would not be appropriate due to patient-specific reasons. The patient would otherwise meet the denominator criteria. Exceptions are not absolute, and are based on clinical judgment, individual patient characteristics, or patient preferences. The PCPI exception methodology uses three categories of reasons for which a patient may be removed from the denominator of an individual measure. These measure exception categories are not uniformly relevant across all measures; for each measure, there must be a clear rationale to permit an exception for a medical, patient, or system reason. For measure Diabetic Retinopathy: Communication with the Physician Managing Ongoing Diabetes Care, exceptions may include medical reason(s) for not communicating the findings of the dilated macular or fundus exam to the physician or other qualified health care professional who manages the ongoing care of the patient with diabetes, or patient reason(s) for not communicating the findings of the dilated macular or fundus

exam to the physician or other qualified health care professional who manages the ongoing care of the patient with diabetes. Although this methodology does not require the external reporting of more detailed exception data, the PCPI recommends that physicians document the specific reasons for exception in patients' medical records for purposes of optimal patient management and audit-readiness. The PCPI also advocates the systematic review and analysis of each physician's exceptions data to identify practice patterns and opportunities for quality improvement.

HQMF eCQM developed and is attached to this submission in fields S.2a and S. 2b.

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

Consistent with CMS' Measures Management System Blueprint and national recommendations put forth by the IOM (now NASEM) and NQF, the PCPI encourages collection of race and ethnicity data as well as the results of this measure to be stratified by race, ethnicity, administrative sex, and payer.

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

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Higher score

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

To calculate performance rates:

1. Find the patients who meet the initial population (ie, the general group of patients that a set of performance measures is designed to address).
2. From the patients within the initial population criteria, find the patients who qualify for the denominator (ie, the specific group of patients for inclusion in a specific performance measure based on defined criteria). Note: in some cases the initial population and denominator are identical.
3. From the patients within the denominator, find the patients who meet the numerator criteria (ie, the group of patients in the denominator for whom a process or outcome of care occurs). Validate that the number of patients in the numerator is less than or equal to the number of patients in the denominator.
4. From the patients who did not meet the numerator criteria, determine if the provider has documented that the patient meets any criteria for exception when denominator exceptions have been specified [for this measure: medical reason(s) for not communicating the findings of the dilated macular or fundus exam to the physician who manages the ongoing care of the patient with diabetes, or patient reason(s) for not communicating the findings of the dilated macular or fundus exam to the physician who manages the ongoing care of the patient with diabetes]. If the patient meets any exception criteria, they should be removed from the denominator for performance calculation. --Although the exception cases are removed from the denominator population for the performance calculation, the exception rate (ie, percentage with

valid exceptions) should be calculated and reported along with performance rates to track variations in care and highlight possible areas of focus for QI.

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

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

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

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

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

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

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

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

If other, please describe in S.18.

Electronic Health Records

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.

Not applicable.

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

No data collection instrument provided

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

Clinician : Group/Practice, Clinician : Individual

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

Other, Outpatient Services, Post-Acute Care

If other: Domiciliary

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

Not applicable. The measure is not a composite.

2. Validity – See attached Measure Testing Submission Form

v2_0089e_nqf_testing_attachment_7.1-636849654289864033.docx,0089e_MAR282019_nqf_testing_attachment_7.1.docx

2.1 For maintenance of endorsement

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

Yes

2.2 For maintenance of endorsement

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

Yes

2.3 For maintenance of endorsement

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

No - This measure is not risk-adjusted

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

Measure Number (*if previously endorsed*)**:** 0089e
Measure Title: Diabetic Retinopathy: Communication with the Physician Managing Ongoing Diabetes Care
Date of Submission: 3/28/2019

Type of Measure:

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

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

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

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

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

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

Previous 2015 Testing

The data source is EHR data.

Current Testing

The data source is 2016 EHR data from the PQRS program, provided by the Centers for Medicare & Medicaid Services (CMS), and includes data reported from a large number of certified EHR vendors. These vendors include several of the major EHR solutions used by inpatient and outpatient care practices. For example: Allscripts, Epic, MEDITECH, Cerner, GE Healthcare, Nextgen, eClinicalWorks, and other smaller EHR vendors.

In 2016 there were six participation options for submitting measure data to PQRS. Of those, the following can be used to submit EHR data:

- Eligible Professionals (EPs) could submit data directly through a qualified EHR product or through a qualified data submission vendor that is Certified EHR Technology.
- Group practices with 2 or more Eligible Professionals can participate through the group practice reporting option (GPRO) using an EHR direct submission or qualified data submission vendor that is Certified EHR Technology.

To participate in the PQRS program, Eligible Professionals and Group practices submit performance data such as number of eligible instances (denominator), instances of quality service performed (numerator), number of performance exclusions, reporting rates, and performance rates—in a file format specified by CMS. Data is then summarized at the practice level and includes both Eligible Professionals participating individually as well as group practices participating through GPRO.

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

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

Data collected from patients sampled between June 1, 2011 and July 6, 2012

Current Testing

The data are for the time period January 2016 through December 2016 and cover the entire United States. Given the required conversion to ICD-10 in late 2015, the testing was completed on the ICD-10 specified measure.

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*)

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

The data sample came from a physician-owned private practice with one ophthalmologist. On average this practice sees over five hundred patients per month.

Current Testing

We received data from 8,832 providers reporting on this measure through the EHR reporting option for CMS's PQRS in 2016. This data set reflects a combination of individual provider data and group data and our analysis of the data as a whole is reflected throughout this submission. Of those 8,832 providers, 7,951 had all the required data elements and at least one quality reporting event for a total of 891,902 quality events. For this measure, 90 percent of providers are included in the analysis, and the average number of quality reporting events are 112 for the remaining 891,902 events. The range of quality reporting events for the 7,951 providers included is from 1 to 1,856.

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*)

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

- The sample consisted of approximately 155 charts for a total of 155 eligible patients
- 2 trained abstractors reviewed the 155 patient charts
- Data abstraction performed from October 23, 2012 to October 30, 2012
- Patients were selected using random sampling

Current Testing

There were 891,902 quality events included in this reliability testing and analysis. These were the quality events that were associated with providers who had all the required data elements and at least one quality reporting event.

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

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

Not applicable

Current Testing

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

After conducting a thorough evaluation of available and relevant PQRS data we selected Diabetes: Eye Exam (PQRS #117) and Diabetic Retinopathy: Documentation of Presence or Absence of Macular Edema and Level of Severity of Retinopathy (PQRS #018) to use for empirical validity correlation 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.

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

Patient-level sociodemographic (SDS) variables were not analyzed in the data sample used.

Current Testing

Patient-level socio-demographic (SDS) variables were not captured as part of the testing as that information was not provided in the CMS data used for analysis.

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*)

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

See 2b2.2 for Validity Against the Gold Standard Results

Previous 2019 Submission

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

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

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

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

Reliability is evaluated by averaging over provider specific reliabilities for all providers that meet the minimum number of quality reporting events for the measure. Each provider must have at least 10 eligible reporting events to be included in this calculation.

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

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

Current Testing

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

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

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

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

Reliability is evaluated by averaging over provider specific reliabilities for all providers that have at least 1 quality reporting event for the measure.

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

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

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

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

See 2b2.3 for Validity Against the Gold Standard Results

Current Testing

The average reliability including providers with at least one quality reporting event is 0.91. We also evaluated the reliability at the 10th, 25th, 50th, 75th, and 90th cutpoints:

Percentile	Value
10th	0.75
25th	0.89
50th	0.96
75th	0.99
90th	1.00

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?*)

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

See 2b2.4 for Validity Against the Gold Standard Results

Current Testing

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

2b1. VALIDITY TESTING

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

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

☒ **Performance measure score**

☒ **Empirical validity testing**

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

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

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

Data abstracted from randomly sampled patient records were used to calculate parallel forms reliability for the measure. Charts for abstraction were selected for patients aged 18 years and older with a diagnosis of diabetic retinopathy.

Data analysis included:

- Percent agreement
- Kappa statistic to adjust for chance agreement

Face Validity

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

After the measure was fully specified, the expert panel was asked to rate their agreement with the following statement:

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

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

Current Testing

For this measure, the PCPI has conducted review and updates to the measure specifications, which satisfy the NQF's ICD-10 Conversion requirements. We are providing the information below to support the three requirements:

- NQF ICD-10-CM Requirement 1: Statement of intent related to ICD-10 CM
  Goal was to convert this measure to a new code set, fully consistent with the original intent of the measure.
- NQF ICD-10-CM Requirement 2: Coding Table
  See attachment in S.2b
- NQF ICD-10-CM Requirement 3: Description of the process used to identify ICD-10 codes
  The PCPI uses the General Equivalence Mappings (GEMs) as a first step in the identification of ICD-10 codes. We then review the ICD-10 codes to confirm their inclusion in the measure is consistent with the measure intent, making additions or deletions as needed. We have an RHIA-credentialed professional on our staff who reviews all ICD-10 coding. For measures included in CMS' Quality Payment Program (QPP), the ICD-10 codes have also been reviewed and vetted by the CMS contractor. Comments received from stakeholders related to ICD-10 coding are first reviewed internally. Depending on the nature of the comment received, we also engage clinical experts to advise us as to whether a change to the specifications is warranted.

Validity testing method

For this measure (Diabetic Retinopathy: Communication with the Physician Managing Ongoing Diabetes Care (PQRS #019)), Diabetes: Eye Exam (PQRS #117), and Diabetic Retinopathy: Documentation of Presence or Absence of Macular Edema and Level of Severity of Retinopathy (PQRS #018) were chosen as suitable candidates for correlation analysis due to the similarities in patient population and domain. We hypothesize that there exists a positive association of scores between providers who performed a dilated macular or

fundus exam with a documented communication to the physician managing ongoing care of the patient with diabetes mellitus regarding the findings at least once within 12 months and those who performed an eye exam (retinal) on patients with diabetes (type 1 and 2). We also hypothesize that there exists a positive association of scores between providers who performed a dilated macular or fundus exam with a documented communication to the physician managing ongoing care of the patient with diabetes mellitus regarding the findings at least once within 12 months and those providers who performed a dilated macular or fundus exam on those with a diagnosis of diabetic retinopathy and included the documentation of the level or severity of retinopathy and the presence or absence of macular edema during one or more office visits within 12 months.

Providers included in the analysis had at least one quality reporting event and were cleaned in the same process as the PQRS dataset.

Datasets were reviewed to identify shared providers based on NPI and TIN identifiers. Comparing performance scores of those shared providers, the empirical analysis uses regression with dataset 1 as the outcome and dataset 2 as the predictor. Results identify the multiple R value (the correlation coefficient) and P-value of the regression variables to assess the association between performance scores of these shared providers.

We use the following guidance to describe correlation[1]:

Correlation	Interpretation	
0.80 – 1.00	Very Strong	
0.60 – 0.79	Strong	
0.40 – 0.59	Moderate	
0.20 - 0.39	Weak	
0 – 0.19	Very Weak	

1. "11. Correlation and Regression." *The BMJ*, 21 March 2019, https://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one/11-correlation-and-regression/ .

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

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

Reliability: N, % Agreement, Kappa (95% CI)
Numerator: 155, 89.7%, 0.52 (0.32, 0.73)
Denominator: 155, 100.0%, NC (NC, NC)*
Exceptions: 155, 100.0%, NC* (NC, NC)**
Overall: 155, 89.7%, 0.52 (0.32, 0.73)
*Cannot calculate Kappa statistics when only one of the four possible categories is represented (Yes/Yes), as this causes a divide-by-zero error in the computational formula

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

Face Validity
Our expert panel included 16 members. The list of expert panel members is as follows:
Gregory Kwasny, MD
David J. Forster, MD
John McAllister, MD
David B. Glasser, MD
Michael Repka, MD
Trexler M. Topping, MD
Jeffrey P. Edelstein, MD
Sonya Shah, MD

John M. Haley, MD
George Williams, MD
Joseph LoCascio, MD
Cynthia Mattox, MD
Daniel Briceland, MD
Kristin Carter, MD
Craig Kliger, MD
Bradley Fouraker, MD

Current Testing

Data from the PQRS program were used to perform the correlation analysis for this measure. Data comes from the EHR versions of Diabetic Retinopathy: Communication with the Physician Managing Ongoing Diabetes Care (PQRS #019), Diabetes: Eye Exam (PQRS #117), and Diabetic Retinopathy: Documentation of Presence or Absence of Macular Edema and Level of Severity of Retinopathy (PQRS #018).

Diabetic Retinopathy: Communication with the Physician Managing Ongoing Diabetes Care (PQRS #019) was positively correlated with Diabetes: Eye Exam (PQRS #117) and Diabetic Retinopathy: Documentation of Presence or Absence of Macular Edema and Level of Severity of Retinopathy (PQRS #018).

PQRS #117

Coefficient of correlation = 0.03

Alpha level = 0.05
P-value = 0.003
Number of shared providers based on NPI and TIN identifiers = 10474

PQRS #018

Coefficient of correlation = 0.53
Alpha level = 0.05
P-value < 0.001
Number of shared providers based on NPI and TIN identifiers = 7742

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?*)

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

This measure demonstrates moderate agreement with a kappa score of 0.52. 89.7 percent agreement was found between the abstractors and the electronic measure implemented in the EHR.

Scale for interpreting kappa:

Kappa	Strength of Agreement
0.00	Poor
0.01 – 0.20	Slight
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Substantial
0.81 – 0.99	Almost perfect

Landis, J.R. and Koch, G.G. (1977) "The measurement of observer agreement for categorical data" in Biometrics. Vol. 33, pp. 159-174

Face Validity

The results of the expert panel rating of the validity statement were as follows: N = 16; Mean rating = 4.63 and 93.75% of respondents either agree or strongly agree that this measure can accurately distinguish good and poor quality.

Frequency Distribution of Ratings

1 – 0 responses (Strongly Disagree)

2 – 0 responses

3 – 1 responses (Neither Agree nor Disagree)

4 – 4 responses

5 – 11 responses (Strongly Agree)

Current Testing

Diabetic Retinopathy: Communication with the Physician Managing Ongoing Diabetes Care has a very weak positive correlation with Diabetes: Eye Exam and a moderate positive correlation with Diabetic Retinopathy: Documentation of Presence or Absence of Macular Edema and Level of Severity of Retinopathy. Both correlations are statistically significant at the 95% significance level. For Diabetes: Eye Exam with a coefficient of correlation of 0.08, the correlation is very weak but an association exists. For Diabetic Retinopathy: Documentation of Presence or Absence of Macular Edema and Level of Severity of Retinopathy with a coefficient of correlation of 0.59, the correlation is moderate, significant, and aligns with our hypothesis. The moderate positive correlation with Diabetic Retinopathy: Documentation of Presence or Absence of Macular Edema and Level of Severity of Retinopathy demonstrates the criterion validity of the measure.

2b2. EXCLUSIONS ANALYSIS

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

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

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

Exceptions included documentation of medical and patient reason(s) for not performing a dilated macular or fundus examination. Exceptions were analyzed for frequency and variability across providers.

Current Testing

Exceptions include:

- Documentation of medical reason(s) for not communicating the findings of the dilated macular or fundus exam to the physician who manages the ongoing care of the patient with diabetes.
- Documentation of patient reason(s) for not communicating the findings of the dilated macular or fundus exam to the physician who manages the ongoing care of the patient with diabetes.

Exceptions were analyzed for frequency across providers.

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*)

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

- Exception rate for this measure was 0.00%
- The exceptions demonstrated 100.00 % agreement. The kappa was non-calculable since only one of the four possible categories was represented which causes a divide-by-zero error in the computational formula.

Current Testing

Amongst the 7,951 included providers, there were a total of 193 exceptions reported. The average number of exceptions per provider in this sample is 0.02. The proportion of exceptions to quality events is 0.00.

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*)

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

Exceptions are necessary to account for those situations when it is not appropriate from a medical or patient reason standpoint to communicate the findings of a dilated macular or fundus examination to the physician who manages the ongoing care of the patient with diabetes. Exceptions are discretionary and the methodology used for measure exception categories are not uniformly relevant across all measures; for this measure, there is a clear rationale to permit an exception for a medical or patient reason. Rather than specifying an exhaustive list of explicit medical and patient reasons for exception for each measure, the measure developer relies on clinicians to link the exception with a specific reason for the decision not to communicate the findings of the dilated macular or fundus exam to the physician who manages the ongoing care of the patient with diabetes.

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

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

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

References:

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

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

Current Testing

See previous 2015 testing response above

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

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

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

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with risk factors**

☐ **Stratification by risk categories**

☐ **Other,**

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

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

Not applicable

Current Testing

Not applicable

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

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

Not applicable

Current Testing

Not applicable

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

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

Not applicable

Current Testing

Not applicable

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

☐ **Published literature**

☐ **Internal data analysis**

☐ **Other (please describe)**

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

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

Not applicable

Current Testing

Not applicable

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

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

Not applicable

Current Testing

Not applicable

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

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

Not applicable

Current Testing

Not applicable

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

If stratified, skip to 2b3.9

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

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

Not applicable

Current Testing

Not applicable

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

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

Not applicable

Current Testing

Not applicable

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

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

Not applicable

Current Testing

Not applicable

2b3.9. Results of Risk Stratification Analysis:

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

Not applicable

Current Testing

Not applicable

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

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

Not applicable

Current Testing

Not applicable

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

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

Not applicable

Current Testing

Not applicable

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

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

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

This data sample was not used to test for meaningful differences in performance across providers or practice sites

Current Testing

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

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

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

This data sample was not used to test for meaningful differences in performance across providers or practice sites

Current Testing

Based on the sample of 7,951 included providers, the mean performance rate is 0.65, the median performance rate is 0.69 and the mode is 1.00. The standard deviation is 0.26. The range of the performance rate is 1.00, with a minimum rate of 0.00 and a maximum rate of 1.00. The interquartile range is 0.39 (0.86–0.48). Percentiles are provided below:

Percentile	Value
10th	0.25
25th	0.48
50th	0.69
75th	0.86
90th	0.96

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?*)

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

This data sample was not used to test for meaningful differences in performance across providers or practice sites

Current Testing

The range of performance from 0.00 to 1.00 and the fairly even spread of the data provided in the percentile chart suggests that there exists clinically meaningful variation across providers' performance. Outliers are

considered to be values less than quartile 1 (0.48) or greater than quartile 3 (0.86) by more than 1.5 the IQR (0.39) and there were no outliers found within this dataset.

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

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

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

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

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

This test was not performed for this measure

Current Testing

This test was not performed for this measure.

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*)

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

This test was not performed for this measure

Current Testing

This test was not performed for this measure.

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*)

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

This test was not performed for this measure

Current Testing

This test was not performed for this measure.

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

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

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

Data were not available to complete this testing

Current Testing

The PQRS dataset provided to us by CMS did not contain missing data so this test was not performed. Nevertheless, missing data may have been rejected when submitted to CMS in which case those values would not be counted towards measure performance. There is no indication that this missing data was systematic, thus their omission would lead to unbiased performance results.

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

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

Data were not available to complete this testing

Current Testing

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

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

Previous 2015 Testing: AMA-PCPI Testing Project (EHR - Validity Against the Gold Standard)

Data were not available to complete this testing

Current Testing

The PQRS dataset provided to us by CMS did not contain missing data so this test was not performed. Nevertheless, missing data may have been rejected when submitted to CMS in which case those values would not be counted towards measure performance. There is no indication that this missing data was systematic, thus their omission would lead to unbiased performance results.

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)

If other:

3b. Electronic Sources

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

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

ALL data elements are in defined fields in electronic health records (EHRs)

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

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

Attachment:

3c. Data Collection Strategy

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

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

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

We have not identified an areas of concern or made any modifications as a result of testing and operational use of the measure in relation to data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, and other feasibility issues unless otherwise noted.

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

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

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 Merit-based Incentive Payment System (MIPS) https://qpp.cms.gov/mips/quality-measures Merit-based Incentive Payment System (MIPS) https://qpp.cms.gov/mips/quality-measures Payment Program Merit-based Incentive Payment System (MIPS) https://qpp.cms.gov/mips/quality-measures Intelligent Research in Sight (IRIS) http://www.aao.org/iris-registry/ AOA MORE Registry https://www.aoa.org/more Merit-based Incentive Payment System (MIPS) https://qpp.cms.gov/mips/quality-measures Intelligent Research in Sight (IRIS) http://www.aao.org/iris-registry/ AOA MORE Registry https://www.aoa.org/more

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

Merit-based Incentive Payment System (MIPS)-Sponsored by the Centers for Medicare and Medicaid Services (CMS)

Prior to 2016, this measure was used for Eligible Providers (EPs) in the Physician Quality Reporting System (PQRS). As of 2017, PQRS has been replaced by the MIPS program. MIPS is a national performance-based payment program that uses performance scores across several categories to determine payment rates for EPs. MIPS takes a comprehensive approach to payment by basing consideration of quality on a set of evidence-based measures that were primarily developed by clinicians, thus encouraging improvement in clinical practice and supporting advances in technology that allow for easy exchange of information

According to the CY 2019 Quality Payment Program final rule, Physician Compare has continued to pursue a phased approach to public reporting under MACRA. CMS intends to make all measures under MIPS quality performance category available for public reporting on Physician Compare. These measures include those reported via all available submission methods for MIPS-eligible clinicians and groups. This measure has now been included in Physician Compare and performance rates will be available sometime in 2019.

Measures and Outcomes Registry for Eyecare (MORE)

The MORE registry is a qualified clinical data registry (QCDR) sponsored by the American Optometric Association. AOA MORE is the nations' first optometric-focused registry. The primary initial goals of the registry are to assist eye-care practices in improving the quality of care, and to submit quality measures to the MIPS program. As of April 2018, AOA MORE was able to attest and submit data for over 600 optometrists and more than 7,492 AOA members are registered with MORE.

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?*)

The PCPI strongly encourages the use of its measures in quality improvement and accountability initiatives and promotes their use in public reporting programs. Measures developed by the PCPI, while copyrighted, can be

reproduced and distributed, without modification, for noncommercial purposes, e.g., use by health care providers in connection with their practices. As a measure developer, we work with measure implementers as opportunities arise to encourage and facilitate the integration of PCPI measures in their programs.

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.*)

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

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

The PCPI measure development and maintenance process is a rigorous, evidence-based process that has been refined and standardized since the PCPI's inception in 2000. Throughout its tenure, the PCPI has conducted its measure development and maintenance process with strict adherence to several key principles, including the following which underscore the role those being measured have played in the development and maintenance process and in providing feedback based on measure implementation:

Collaborative Approach to Measure Development

PCPI measures are developed and maintained through cross-specialty, multi-disciplinary technical expert panels. Representatives of relevant clinical specialties are invited to participate in our expert panels to advise us throughout the measure development process and as questions arise during measure implementation. Additionally, other health care providers and stakeholders participate in our panels as equal contributors to the measure development process. The PCPI also strives to include on its panels individuals representing the perspectives of patients, consumers, private health plans, and employers. Liaisons from key measure development organizations, including The Joint Commission and NCQA, at times participate in the PCPI's measure development process to ensure measure harmonization. Measure methodologists and coding and informatics experts are also considered important members of the expert panel. This broad-based approach to measure development maximizes the input from those being measured and other stakeholders to develop evidence-based, feasible and clinically meaningful measures.

Public Comment Period

Input from a wide range of stakeholders is integral to the measure development process. To invite other perspectives and expertise beyond the expert panels and particularly from those providers and facilities that will implement these measures, the PCPI submits the measures for public comment. All measures are released for a 30-day public and PCPI member comment period. All comments are reviewed by the technical expert panel to determine whether measure modifications are needed based on comments received.

Feedback Mechanisms

The PCPI has a dedicated mechanism set up to receive measure-related comments and questions from implementers. As comments and questions are received, they are shared with appropriate staff for follow up.

Feasibility Assessments

The PCPI solicits feedback on measure feasibility in the following domains: data availability, data accuracy, data standards, and workflow to guide future modifications to the measure.

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

Feedback Mechanisms

The PCPI has a dedicated mechanism set up to receive measure-related comments and questions from implementers. As comments and questions are received, they are shared with appropriate staff for follow up. If

comments or questions require expert input, these are shared with the PCPI's technical expert panels to determine if measure modifications may be warranted. Additionally, for PCPI measures included in federal reporting programs, there is a system that has been set up to elicit timely feedback and responses from PCPI staff in consultation with technical expert panel members, as appropriate.

Feasibility Assessments

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

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

Describe how feedback was obtained.

As described in Section 4a2.1.1, the PCPI invites feedback through various mechanisms. We obtain input from our topic-specific technical expert panels during the measure development and during the annual maintenance process. Additionally, the PCPI obtains feedback via an online public comment and an email-based process set up to receive measure inquiries from implementers.

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

Feedback obtained for this measure includes a request for clarification about the types of communication that would meet the measure. We added a definition to the measure to clarify that the communication would need to be documented and must include findings of the dilated macular or fundus exam. Communication may be verbal or via a written or electronic communication.

We have also obtained feedback requesting clarification about the information that should be included in the communication. As a result, we added a definition stating that the communication would include level of severity of retinopathy (eg, mild nonproliferative, moderate nonproliferative, severe nonproliferative, very severe nonproliferative, proliferative) AND the presence or absence of macular edema.

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

Feedback obtained from other users similarly requested clarification about the types of communication that would meet the measure.

We have also obtained feedback requesting clarification about the information that should be included in the communication.

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.

In response to the feedback we received, we added a definition to the measure to clarify that the communication would need to be documented and must include findings of the dilated macular or fundus exam. Communication may be verbal or via a written or electronic communication. We also added a definition stating that the communication would include level of severity of retinopathy (eg, mild nonproliferative, moderate nonproliferative, severe nonproliferative, very severe nonproliferative, proliferative) AND the presence or absence of macular edema.

Improvement

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

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

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

The intent of this measure is to help improve the care of patients with diabetic retinopathy. CMS data report a clear gap in care evidenced with a performance rate of 74.78% in 2017 which marked a decrease from the rate of 81.0% in 2014. However, performance rates represent but one facet of the quality improvement process.

While the PCPI creates measures with the ultimate goal of improving the quality of care, measurement is a mechanism to drive improvement but does not equate with improvement. Measurement can help identify opportunities for improvement with actual improvement requiring making changes to health care processes and/or structure. In order to promote improvement, quality measurement systems need to provide feedback to front-line clinical staff in as close to real time as possible and at the point of care whenever possible. (1)

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

4b2. Unintended Consequences

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

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

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

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

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

Yes

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

0055 : Comprehensive Diabetes Care: Eye Exam (retinal) performed

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.

Measure #0055 evaluates the percentage of patients 18-75 years of age with diabetes who had an eye exam (retinal) performed. While the population is similar, the PCPI measure requires that a dilated macular or fundus exam be performed, and the results communicated to the physician who manages the ongoing care of the patient with diabetes so as to facilitate the coordination of care.

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.)

not applicable

Appendix

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

No appendix **Attachment:**

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): PCPI Foundation

Co.2 Point of Contact: Samantha, Tierney, samantha.tierney@thepcpi.org, 312-224-6071-

Co.3 Measure Developer if different from Measure Steward: PCPI Foundation

Co.4 Point of Contact: Elvia, Chavarria, elvia.chavarria@thepcpi.org, 312-224-6064-

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.

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

Technical Expert Panel:

John T. Thompson, MD – TEP Co-Chair

Parag Parekh, MD – TEP Co-Chair

TEP members:

Murray Fingeret, OD

David Glasser, MD

Richard Hellman, MD,

Mathew W. MacCumber, MD, PhD

Zachary McCarty, OD

Marc Piccolo, OD

Thomas Wong, OD

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2006

Ad.3 Month and Year of most recent revision: 12, 2018

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

Ad.5 When is the next scheduled review/update for this measure? 12, 2019

Ad.7 Disclaimers: Physician Performance Measures (Measures) and related data specifications have been developed by the American Medical Association (AMA) - convened Physician Consortium for Performance Improvement® (PCPI®). These Measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications. The Measures, while copyrighted, can be reproduced and distributed, without modification, for noncommercial purposes, eg, use by health care providers in connection with their practices. Commercial use is defined as the sale, license, or distribution of the Measures for commercial gain, or incorporation of the Measures into a product or service that is sold, licensed or distributed for commercial gain. Commercial uses of the Measures require a license agreement between the user and the AMA, (on behalf of the PCPI). Neither the AMA, PCPI nor its members shall be responsible for any use of the Measures.

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

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

Ad.8 Additional Information/Comments: