NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 3571e

Corresponding Measures:

De.2. Measure Title: Retesting of Abnormal Blood Glucose in Patients with Prediabetes

Co.1.1. Measure Steward: American Medical Association

De.3. Brief Description of Measure: Percentage of patients aged 18 years and older who had an abnormal fasting plasma glucose, oral glucose tolerance test, or hemoglobin A1c result in the range of prediabetes in the previous year who have a blood glucose test performed in the one-year measurement period

1b.1. Developer Rationale: At least annual glucose testing in patients who were previously found to have lab results in the range of prediabetes is an important aspect of care so that patients can be monitored for improvement or potential transition to Type 2 diabetes. While there are no current studies that show patients with prediabetes do not have follow-up testing completed, the TEP felt that this is a key area in which to have a measure. Preventing the onset of type 2 Diabetes by screening for prediabetes, with the goal of referring for treatment and prevention, will help to reduce the 84 million patients with prediabetes, and the overall number that eventually are diagnosed with type 2 diabetes. Furthermore, cost savings for prevention of type 2 diabetes have the potential to be significant. For every beneficiary, at 15 months, costs savings are at least $ 2,650.00 per person.

S.4. Numerator Statement: Patients who had a blood glucose test performed

*Retesting for abnormal blood glucose may include using a fasting plasma glucose, 2-h plasma glucose during a 75g oral glucose tolerance test, or A1C.

S.6. Denominator Statement: All patients aged 18 years and older who had an abnormal fasting plasma glucose, oral glucose tolerance test, or hemoglobin A1c result in the range of prediabetes in the year prior to the one-year measurement period

**Abnormal lab result in the range of prediabetes includes a fasting plasma glucose level between 100 mg/dL (5.6 mmol/L) to 125 mg/dL (6.9 mmol/L) OR a 2-hour glucose during a 75g oral glucose tolerance test between 140 mg/dL (7.8 mmol/L) to 199 mg/dL (11.0 mmol/L) OR and A1C between 5.7-6.4% (39-47 mmol/mol).

S.8. Denominator Exclusions: Denominator Exclusions:

Exclude patients who are pregnant.

Exclude patients who have any existing diagnosis of diabetes (Type 1, Type 2, latent autoimmune diabetes of adults [LADA], monogenic diabetes [MODY]).

Exclude patients in palliative care/hospice

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: Most Recent Endorsement Date:

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

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

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

Preliminary Analysis: New Measure

Criteria 1: Importance to Measure and Report

1a. Evidence

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

The developer provides the following evidence for this measure:

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

Evidence Summary

- This process measure eCQM at the clinician: individual and group level is the percentage of patients aged 18 years and older who had an abnormal fasting plasma glucose, oral glucose tolerance test, or hemoglobin A1c result in the range of prediabetes in the previous year who have a blood glucose test performed in the one-year measurement period.
- Developer provided a logic model connecting the measure focus with positive patient outcomes, namely the testing of glucose, the evaluation of the results to determine the presence of diabetes onset, and appropriate therapeutic steps to reduce the incidence of diabetes related sequelae.
- Developer cites evidence found in guidelines from the United States Preventive Services Task Force (USPSTF) and from the American Diabetes Association (ADA).
    - At least annual monitoring for the development of diabetes in those with prediabetes is suggested. (ADA, 2018) (E Recommendation)
    - Developer provides evidence of disease prevalence and systematic misses of opportunities to intervene by clinicians.
    - Developer does not provide studies that offer evidence that annual monitoring is associated with positive outcomes. However, NQF staff recommend an exception to evidence.

Questions for the Committee:

- What is the relationship of this measure to patient outcomes? How strong is the evidence for this relationship? Is the evidence directly applicable to the process of care being measured?
- Does the Committee agree with the staff evaluation of the evidence presented by the developer?

- Is the Committee aware of studies that directly test the hypothesis of the measure focus that annual monitoring leads to better patient outcomes?
- Does the Committee wish to grant an exception to evidence based on the expert opinion of the ADA guideline writers?

Guidance from the Evidence Algorithm

Process measure based on guidelines (Box 3)→QQC presented, but based on expert opinion only (Box 7)→ No existing outcome measures (Box 10)→ Systematic assessment of expert opinion→ **Insufficient** (NQF Measure Evaluation Criteria Sept 2019, Algorithm 1 pg. 15)

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

RATIONALE:

- Developer does not identify literature that presents evidence of positive outcomes based on annual follow up testing, but this is recommended by experts and is a logical course of action for a care provider. NQF staff are suggesting an exception to evidence.

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.

- For this new measure, developer did not provide an analysis of provider performance as the data was not available. Rather, they summarized the literature indicating a performance gap.
- There are substantial numbers of patients with prediabetes; this is an opportunity to intervene when patients present for care.
    - United States has 84 million adults with prediabetes.
    - 9 out of 10 patients who have prediabetes are not aware.
    - Missed opportunities among primary care providers in diagnosing and managing patients with prediabetes represent a gap in care.
- Early detection and management of pre-diabetes is fundamental to preventing diabetes.
    - Despite established national screening guidelines in U.S., suboptimal screening rates are reported, with 45% of those meeting screening criteria being screened
    - Additionally, survey data show that while primary care physicians are aware of the guidelines that support screening for prediabetes, there is a disconnect between this knowledge and actual practice.

Disparities

- Developer did not provide a summary of the literature related to disparities.

Questions for the Committee:

- Does the Committee agree with the staff assessment that there is 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

Comments:

- Yes this part of their DPP measure set.
- How frequently is an 'exception to evidence' given? What other approved measures? Under what circumstances? Why this measure?
- Most concerning is the evidence to support the measure. This is especially true given the recent NQF statements about results. Why worry about retesting, when the better measure would be referral for treatment.
- The evidence relates directly to the specific process being measured.
- This process measure aligns directly to the ADA E Recommendations: At least annual monitoring for the development of diabetes in those with prediabetes is suggested. In addition, the USPSTF recommends screening for abnormal blood glucose as part of cardiovascular risk assessment in adults aged 40 to 70 years of age who are overweight or obese. Both recommendations applying directly to this process measure.
- This is a newly proposed process measure. It is proposed that annual tests for diabetes in those people defined as being Pre-DM would reduce the progression into DM and, thus, hold down DM complication rates and health care costs. The USPSTF guideline for overweight and obese people 40 to 70 y/o to be tested for DM was cited. This USPSTF guideline does not mention repeat testing recommendations. The ADA recommends annual DM testing for people diagnosed as Pre-DM but this is based on expert opinion only.
- Process Measure. Insufficient evidence provided.

1b. Performance Gap

Comments:

- Supported by study and evidence in literature. There is great room for improvement, especially in underserved populations.
- Evidence sited re gap in screening, but screening is different than retesting
- The gap in care is on the possible impact to reduce diabetes, not necessarily on the "testing" of the prediabetic status.
- Literature were cited, but not direct evidence of gaps in care were provided.
- No performance data submitted as this is a new measure. However, based on the literature review a performance gap was indicted. Supporting the need to monitor prediabetes retesting: 84 million adults with prediabetes, 9 out of 10 are unaware they have prediabetes, and rates of screening for ideal populations are suboptimal (45% of those meeting screenig criteria are being screened)
- " There are no current studies that show patients with prediabetes do not have follow-up testing completed, ...". There are no studies cited to support the benefits of annual testing of those diagnosed with pre-DM, although this is an expert opinion suggestion of the ADA. Is this sufficient to warrant a national performance measure?
- Performance gap exists. Is it failure to initially test high risk individuals or failure to retest those with an abnormal test result?

1b. Disparities

Comments:

- Pointed to studies. Again there are more evidence provided in the Quality Reports to MDPP and CDC DPP programs, as well as diabetes and endocrine journals with review provided by diabetes experts and education/care specialists.
- NA
- No literature provided, although clearly some racial and ethnic groups appear to have different rates of diabetes.
- Data was not provided that demonstrated disparities in care, though it was discused in some of the literature referenced.
- Data was not provided based on disparities.
- No published studies were found in the literature search regarding disparities by population subgroups.
- No data was provided

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

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

Reliability

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

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

Validity

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

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

Composite measures only:

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

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

Evaluators: NQF Primary Care and Chronic Illness Committee Staff

NQF Staff Full Evaluation

Evaluation Summary:
Specifications

- Submitted measure specification follows eCQM industry specs as indicated Sub-criterion 2a1
- Submitted measure specification is fully represented and is not hindered by any limitations in the eCQM industry specs

Reliability/Validity

- Developer used same testing for both data element reliability and validity.
- Developer performed data element reliability/validity testing at two facilities on two common EHR systems.
    - Test Site #1: An ambulatory facility in South Carolina, part of a larger health system comprised of 8 inpatient hospitals and more than 100 outpatient facilities. This facility uses Epic EHR in their facility.
    - Test Site #2: An ambulatory facility in South Carolina, part of a larger system comprised of a 1,600+ bed comprehensive integrated health system, serving 1 million patients. This facility uses Cerner EHR in their facility.
- Submission includes simulated data set results demonstrating unit testing covering 100% of the measure logic.
- The feasibility assessment indicated the following data elements had issues in the accuracy domain indicating that these data elements may not be correct:
    - Laboratory Test, Performed: Fasting Plasma Glucose Lab Test Mass Per Volume" (*measure developer noted that fasting status of glucose testing is not captured in discrete fields in either EHR, however capturing A1C testing is feasible. To test for prediabetes, fasting plasma glucose, 2-h plasma glucose during 75-g oral glucose tolerance test, and A1C are equally appropriate. (in Cerner and Epic)*
    - "Laboratory Test, Performed: Fasting glucose [Moles/volume] in Serum or Plasma" (in Cerner and Epic)
    - "Laboratory Test, Not Performed: Fasting glucose [Moles/volume] in Serum or Plasma" (in Cerner and Epic)
    - "Laboratory Test, Not Performed: Glucose [Moles/volume] in Serum or Plasma --2 hours post 75 g glucose PO" (in Cerner)
    - "Laboratory Test, Not Performed: Fasting Plasma Glucose Lab Test Mass Per Volume" (in Cerner and Epic)
    - "Laboratory Test, Not Performed: Glucose [Mass/volume] in Serum or Plasma --2 hours post 75 g glucose PO" (in Cerner)
    - "Intervention, Order: Comfort Measures" using "Comfort Measures (2.16.840.1.113883.17.4077.3.2030)" (*measure developer noted that Comfort Care as an exclusion is standard in NQF endocrine registry measures and it is expected that EMR developers to create a distinct field to collect this data in the future) (in Cerner)*
- Data element reliability/validity testing was conducted utilizing Parallel Forms Reliability Testing methodology to determine if data elements found through electronic data pulls could be confirmed by manual abstraction of the same data elements.
    - Verification of the data elements was obtained through automated data search strategies against a reference strategy (considered the gold standard) for obtaining the data elements.
    - Manual review of the data elements was used as the reference strategy against which automated data search and extraction strategies were evaluated.
    - Interrater reliability (crude agreement and Cohen's Kappa) was used to assess the reliability of the measure based on results from two independent reviewers trained in the same way reviewing the same patient record.
- Measure demonstrates strong kappa scores at the two testing sites
- NQF staff consider this an appropriate testing methodology according to current NQF criteria.

- NQF staff were not certain from the submission that the developer had tested all data elements, or minimally numerator, denominator and all exclusions, as required by NQF criteria.

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- Does the Committee agree with the NQF staff's assessment and rating the reliability testing?

Questions for the Committee regarding validity:

- Does the Committee still consider the measure valid given the accuracy issues noted in the feasibility assessment?
- Do you have any other concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?
- Does the Committee agree with the NQF staff's assessment and rating the validity testing?

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

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

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

2a1. Reliability – Specifications

Comments:

- Again this is the follow up to DPP set. So I same concerns as in 3569e
- Seems ok
- Although the variable chosen for inclusion are reliable, the question is are they the correct variables to measure to impact the onset of diabetes?
- Only concern is with the validation of the testing that was performed.
- Align with the eCQM industry specs (no concerns)
- With only two semesters of statistics, I am hardly proficient. But I need an explanation of why, with Site 1 having a preliminary denominator count of 53 patients and a calculated sample size of 344, the actual sample size selected is 112.
- No concerns

2a2. Reliability - Testing

Comments:

- Agree with NQF staff's rating of Moderate for testing specifically in SC's Epic and Cerner EMRs
- While process seems to be appropriate, there is some question regarding available data.
- If you agree with the measurement developer's logic, then it would be reliable.
- No, I do not have concerns.
- The measure has strong kappa scores at both sites and Parellel Forms Reliability Testing supports the reliability of the testing as it is repeatable (manual and automated)
- The data elements missing from the EMR search were all found by manual searching, but no frequency of missing data elements is provided. Interrater reliability was tested using only two raters. Gives a comparison but how generalizable are the conclusions from this?.
- No concerns

2b1. Validity – Testing

Comments:

- Agree with NQF Staff and have concerns about the accuracy issues.
- There is some question if all NQF criteria had been met
- Real concern about the logic of being able to minimize complications through management by knowing the history of glucose levels. I don't doubt the validity of the testing results.
- No, I do not have any concerns.
- No concerns
- Do the Kappa scores for interrater reliability have bearing on content validity?
- No concerns

2b4-7. Threats to Validity

Comments:

- I agree with the NQF staff concern: NQF staff are concerned that not all exclusions were tested as it is not clear from the submission whether this was in fact the case.
- Differences between fullness of available data (electronic and abstraction)
- From a statistical perspective there are minimal concerns about threats to validity. However, overarching concerns about validity of measure.
- It was not clear from the information provided by the developer what specific elements were tested. The process used appears to be reasonable and valid. More specific information should have been made available.
- No concerns beyond data capabilities, more data can likely be discovered via hybrid methods (i.e. chart review)
- "We do not have the number of the overall frequency of missing data."
- No concerns with validity

2b2-3. Other Threats to Validity

Comments:

- No Risk Adjustment applied.
- NA
- No risk adjustment and exlusions seem reasonable.
- Exclusions are appropriate, pregnant, diagnosed with diabetes and palliative/end of life care/hospice.
- Exclusions align to similar measures and no concerns noted. This measure doesn't use risk adjustments
- Why are ambulatory patients excluded (2b2.3)? There was no risk adjustment.
- NA

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 feasibility assessment indicated the following data elements have feasibility issues in the *availability*, *workflow, and standards* domains indicating the data elements are not routinely generated during care or available in electronic sources:
    - Laboratory Test, Performed: Fasting Plasma Glucose Lab Test Mass Per Volume" (*measure developer noted that fasting status of glucose testing is not captured in discrete fields in either EHR, however capturing A1C testing is feasible. To test for prediabetes, fasting plasma glucose, 2-h plasma glucose during 75-g oral glucose tolerance test, and A1C are equally appropriate. (in Cerner and Epic)*
    - "Laboratory Test, Performed: Fasting glucose [Moles/volume] in Serum or Plasma" (in Cerner and Epic)
    - "Laboratory Test, Not Performed: Fasting glucose [Moles/volume] in Serum or Plasma" (in Cerner and Epic)
    - "Laboratory Test, Not Performed: Glucose [Moles/volume] in Serum or Plasma --2 hours post 75 g glucose PO" (in Cerner)
    - "Laboratory Test, Not Performed: Fasting Plasma Glucose Lab Test Mass Per Volume" (in Cerner and Epic)
    - "Laboratory Test, Not Performed: Glucose [Mass/volume] in Serum or Plasma --2 hours post 75 g glucose PO" (in Cerner)
    - "Intervention, Order: Comfort Measures" using "Comfort Measures (2.16.840.1.113883.17.4077.3.2030)" (*measure developer noted that Comfort Care as an exclusion is standard in NQF endocrine registry measures and it is expected that EMR developers to create a distinct field to collect this data in the future) (in Cerner)*
- All value sets used in measure submission are accessible via the VSAC
- Submission includes simulated data set results demonstrating unit testing covering 100% of the measure logic.
- Measure developer notes that the data for the measure are collected as part of the routine provision of care.
- Developer asserts that all data elements reside in defined fields inside the EHR.
- No licensing agreement or fees required for use of the measure.

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?
- Does the eCQM Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?

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

RATIONALE: eCQM feasibility assessment demonstrated that key data elements have feasibility issues in the *availability*, *workflow, and standards* domains indicating the data elements are not routinely generated during care or available in electronic sources. The submission did not include a plan for addressing the feasibility issues with each data element.

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. Feasibility

Comments:

- have the same concern I had with 3569e as a set for DPP
- Seems a number of data elements have potential collection and/or accuracy concerns
- The study would be feasible as defined.

- The measure is feasible using EHRs. EHR data though is not consistently available, and, not all providers use EHR systems. This will remain a challenge in reporting this measure.
- There are some opportunities related to various laboratory test (low feasibility)
- Missing data elements in the EMR search were found by manual searching. However, we are given no numbers as to the frequency of missing data from the two sites with two different EMRs. This may have a bearing on the feasibility of depending on EMR data only.
- Challenges exist with feasibility. Data elements are not routinely generated in the routine delivery of care/normal work flow.

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

- This measure has not yet been implemented but the developer provides an implementation plan that includes:
    - MIPS QPP to complement the prediabetes improvement activities
    - Maryland Primary Care Program´s public reporting program for 2021
    - Prediabetes MIPS Value Pathway (MVP) for the 2021 performance period

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

- Developer notes that the measure has not been implemented and therefore they have no feedback from end users.

Additional Feedback: N/A

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

- This measure has not yet been implemented

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

- None identified by developer

Potential harms

- None identified by developer

Additional Feedback:

- None identified by developer

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

Comments:

- No feedback as it is new.
- None, but some planned for 2021
- Assuming that the measure would be reported publicly once in use.
- Planned use in other programs. Not yet in use.
- Measure not being reported to date but plan to use in programs in the future
- This measure is proposed and has not been implemented. There is no mention that the study results at the two hospital systems were used internally or that feedback from providers and others was requested.
- Planned use

4b1. Usability – Improvement

Comments:

- Developer did not identify. Some unintended consequences are over or under reporting that would impact the MIPS calculation. Benefit is increased and earlier intervention for persons at risk for diabetes.
- measure not currently in use
- No clear evaluation of benefits versus harms.
- No harm anticipated from use of the measure. Use of the measure, could improve outcomes for patients, depending on how results are discussed with patient.
- support the benefits vs harm this measure would provide
- The hypothesis that annual DM clinical tests will lead to fewer cases of people with Pre-DM progressing to DM, with subsequent decreased rates of DM complications and lower medical costs is not really explored in the evidence for this submission. The current (6/8/2020) USPSTF Web pages list that a Final Research Plan was developed in 2018 to explore "Abnormal Blood Glucose and Type 2 DM Screening". That study is not yet available. In the USPSTF recommendations under the same title dated 2015, "The USPSTF found inadequate direct evidence that measuring blood glucose leads to improvements in mortality or cardiovascular disease but concluded that there is a moderate net benefit to measuring blood glucose to detect [impaired glucose metabolism] or DM. Lifestyle modifications should be implemented for those with abnormal blood glucose. There is no mention of follow-up testing recommendations.
- NA

Criterion 5: Related and Competing Measures

Related or competing measures

- No related or competing measures were identified by the developer or NQF staff.

Harmonization

- N/A

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

5. Related and Competing

Comments:

- None in NQF, however have we looked at the CDC DPP and CMS MDPP quality data bases?
- NA
- None suggested.
- None.
- No competing measures
- None.
- None

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 06/22/20

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

Combined Methods Panel Scientific Acceptability Evaluation

Evaluating Scientific Acceptability: Instructions

Scientific Acceptability: 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 subcriteria for both reliability and validity to pass this criterion.

Instructions for filling out this form:

- Please complete this form for each measure you are evaluating. Relevant measure documents are at the bottom of the SharePoint site.
- If you are unable to check a box, please highlight or shade the box for your response.
- You must answer the "overall rating" item for both Reliability and Validity. Also, be sure to answer the composite measure question at the end of the form if your measure is a composite.
- For several questions, we have noted which sections of the submission documents you should ***REFERENCE*** and provided ***INSTRUCTON BOXES*** in comment bubbles to help you answer them.
- ***Please refer to the 2017*** Measure Evaluation Criteria and Guidance document ***(pages 18-24) and the 2-page*** Key Points document ***when evaluating your measures***.
- ***Please base your evaluations solely on the submission materials provided by developers.*** NQF strongly discourages the use of outside articles or other resources, even if they are cited in the submission materials. If you require further information or clarification to conduct your evaluation, please communicate with NQF staff as soon as possible (methodspanel@qualityforum.org). Is it possible that we can obtain the needed information, but only if requested in a timely manner.
- ***Remember*** that testing at either the data element level **OR** the measure score level is accepted for some types of measures, but not all (e.g., instrument-based measures, composite measures), and therefore, the embedded rating instructions may not be appropriate for all measures. Please review the box below to guide your rating.
- If a measure you are evaluating includes multiple measures (e.g., the Hopsital CAHPS measure submsission acutally includes 11 performance measures), all included measures must be rated. You may decide that one rating applies to all included measures, or you may need to provide separate ratings (e.g., if results are substantially better for one measure than for another).

Measure type	Requirements for Reliability testing	Requirements for Validity testing
Instrument-based measures	BOTH data element and score-level testing	BOTH data element and score-level testing
Composite measures	Score-level testing of the composite measure score; testing of the components is not sufficient	Score-level testing of the composite measure score is desired. At initial endorsement only, empirical or face validity testing of the components OR face validity of the composite is acceptable.

Measure type	Requirements for Reliability testing	Requirements for Validity testing
eCQMs	All eCQMs must be tested using the Health Quality Measure Format (HQMF) specifications, which should also use the QDM and value sets published through VSAC Reliance on data from structured data fields is expected; otherwise, unstructured data must be shown to be both reliable and valid. Thus, testing for elements that are not included in structured data fields should be tested at the data element level.	All eCQMs must be tested using the Health Quality Measure Format (HQMF) specifications, which should also use the QDM and value sets published through VSAC Reliance on data from structured data fields is expected; otherwise, unstructured data must be shown to be both reliable and valid. Thus, testing for elements that are not included in structured data fields should be tested at the data element level. Empirical testing is expected, and as of August 2019, data element validation will be required unless justification is provided/accepted. Face validity alone will not be sufficient. Use of a simulated data set (e.g. BONNIE) is no longer accepted for testing validity of data elements
Cost and Resource Use Cost and Resource Use Measure Evaluation Criteria	EITHER data element or score-level testing	Validity is considered in the context of measure intent and threats to validity based on these cost measure-specific components: • Attribution approach • Cost categories • Approach to outliers • Impact of Carve Outs EITHER data element or score-level testing; face validity not accepted for maintenance measures unless justification provided/accepted
All others (Process; Appropriate Use; Structure; Efficiency; Outcome; Intermediate Clinical Outcome; Access)	EITHER data element or score-level testing	EITHER data element or score-level testing; face validity not accepted for maintenance measures unless justification provided/accepted; if data element validity is demonstrated, additional reliability testing is not required

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: Measure submission, items S.1-S.22

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

2. **Briefly summarize any concerns about the measure specifications.**
   - None identified by staff

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 VALIDITY testing** of patient-level data conducted?

   ☒ **Yes** ☐ **No**

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

   Submission document: Testing attachment, section 2a2.2

   - Developer used same testing for both data element reliability and validity.
   - Developer performed data element reliability/validity testing at two facilities on two common EHR systems.
     - Test Site #1: An ambulatory facility in South Carolina, part of a larger health system comprised of 8 inpatient hospitals and more than 100 outpatient facilities. This facility uses Epic EHR in their facility.
     - Test Site #2: An ambulatory facility in South Carolina, part of a larger system comprised of a 1,600+ bed comprehensive integrated health system, serving 1 million patients. This facility uses Cerner EHR in their facility.

- Data element reliability/validity testing was conducted utilizing Parallel Forms Reliability Testing methodology to determine if data elements found through electronic data pulls could be confirmed by manual abstraction of the same data elements.
    - Verification of the data elements was obtained through automated data search strategies against a reference strategy (considered the gold standard) for obtaining the data elements.
    - Manual review of the data elements was used as the reference strategy against which automated data search and extraction strategies were evaluated.
    - Interrater reliability (crude agreement and Cohen's Kappa) was used to assess the reliability of the measure based on results from two independent reviewers trained in the same way reviewing the same patient record.
- NQF staff consider this an appropriate testing methodology according to current NQF criteria.

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

   Submission document: Testing attachment, section 2a2.3

   - The developer reported Kappa scores and crude agreement by site:

		% Agree	Kappa	N
Site 1	**Denominator**	98	0.927	112
	Numerator	94	0.850	17
Site 2	**Denominator**	100	1	75
	Exclusions	93	0.836	70
	Numerator	100	**	23

   **Kappa scores not calculable

   - Measure demonstrates strong kappa scores at the two testing sites

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

   Submission document: Testing attachment, section 2a2.2

   ☐ **Yes**

   ☐ **No**

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

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

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

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

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

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

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

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

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

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**
    - Developer used an appropriate testing methodology for the data element testing.
    - Staff were concerned that the developer notes a few instances across the measure where more full and accurate information could be found in the manual abstraction process than through electronic reporting.
        - "Numerator – Referrals to diabetes prevention program or dietician are often automated messages. These can be seen in manual abstraction and depending on level of access to the EHR system, not all medical staff can see these messages."

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.
    - Measure exclusions appear appropriate.
    - NQF staff are concerned that not all exclusions were tested as it is not clear from the submission whether this was in fact the case.

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

    Submission document: Testing attachment, section 2b4.
    - Developer notes that "Differences in performance were not tested, however during testing, performance was calculated with performance rates of 0.292 and 0.483 for sites 1 and 2, respectively."
    - NQF staff do not consider there to be significant threats to the measure's ability to detect meaningful differences in provider performance.

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
    Submission document: Testing attachment, section 2b5.
    - N/A

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

    Submission document: Testing attachment, section 2b6.
    - Developer notes that, "As part of the reliability testing, sites completed data element tables that assessed for missing elements. All elements that were missing on the sample cases were tracked. Since we found instances across the measures where more full and accurate information could be found in the manual abstraction process than through electronic reporting, this seems to be a consistent issue across all types of measures, not just this particular measure, given the nature of EHR capabilities and limitations."
    - No additional concerns from NQF staff.

16. **Risk Adjustment**

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

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

    ☐ Yes ☐ No ☒ Not applicable

16c. **Social risk adjustment:**

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

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

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

16d.**Risk adjustment summary:**

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

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

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

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

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

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

- N/A

VALIDITY: TESTING

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

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

☐ Face validity

☐ Empirical validity testing of the measure score

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

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

Submission document: Testing attachment, section 2b2.2

- Developer used same testing for both data element reliability and validity.
- Developer performed data element reliability/validity testing at two facilities on two common EHR systems.
    - Test Site #1: An ambulatory facility in South Carolina, part of a larger health system comprised of 8 inpatient hospitals and more than 100 outpatient facilities. This facility uses Epic EHR in their facility.
    - Test Site #2: An ambulatory facility in South Carolina, part of a larger system comprised of a 1,600+ bed comprehensive integrated health system, serving 1 million patients. This facility uses Cerner EHR in their facility.
- Data element reliability/validity testing was conducted utilizing Parallel Forms Reliability Testing methodology to determine if data elements found through electronic data pulls could be confirmed by manual abstraction of the same data elements.
    - Verification of the data elements was obtained through automated data search strategies against a reference strategy (considered the gold standard) for obtaining the data elements.
    - Manual review of the data elements was used as the reference strategy against which automated data search and extraction strategies were evaluated.
    - Interrater reliability (crude agreement and Cohen's Kappa) was used to assess the reliability of the measure based on results from two independent reviewers trained in the same way reviewing the same patient record.
- NQF staff consider this an appropriate testing methodology according to current NQF criteria.

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

Submission document: Testing attachment, section 2b2.3

- The developer reported Kappa scores and crude agreement by site:

		% Agree	Kappa	N
Site 1	Denominator	98	0.927	112
	Numerator	94	0.850	17
Site 2	Denominator	100	1	75
	Exclusions	93	0.836	70
	Numerator	100	**	23

**Kappa scores not calculable

- Measure demonstrates strong kappa scores at the two testing sites

21. **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)

22. **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)

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

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

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

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

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

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

- Developer used an appropriate testing methodology for the data element testing.

ADDITIONAL RECOMMENDATIONS

25. **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.**

- None identified by NQF staff.

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

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

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

NQF_evidence_attachment_Retesting_of_Abnormal_Glucose_.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*)**:** Click here to enter NQF number

Measure Title Retesting of Abnormal Blood Glucose in Patients with Prediabetes

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 4/9/2020

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

Outcome

☐ Outcome: Click here to name the health outcome

☐Patient-reported outcome (PRO): Click here to name the 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*): Click here to name the intermediate outcome

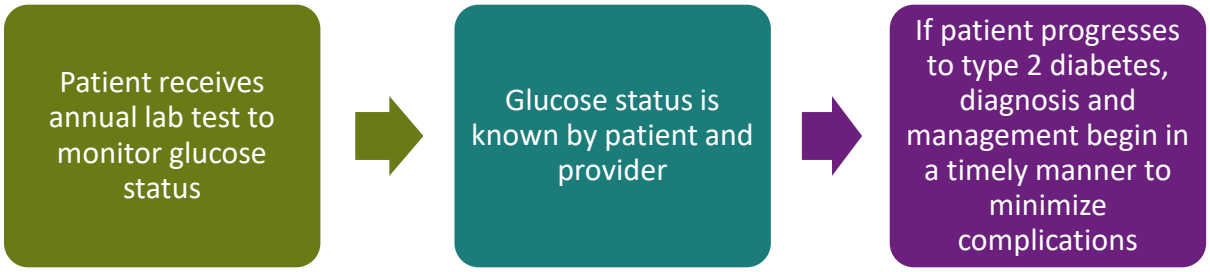
☒ Process: Glucose retesting in patients with prediabetes

☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

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.)

Implementing this measure to increase follow-up screening for patients with prediabetes can improve health outcomes for patients by preventing the progression to type 2 diabetes. Cost savings associated with preventing diabetes are significant. In the Medicare Diabetes Prevention Program (Medicare DPP) model test conducted through the Center for Medicare and Medicaid Innovation, implementation of the MDPP preventive service resulted in an estimated cost savings of $ 2,650.00 per participating Medicare beneficiary over 15 months. Individuals with diabetes typically have medical expenses 2.3 times higher than those without it. The longitudinal impact of this measure would be substantial in terms of cost savings and disease prevention.

****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.

N/A

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

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

x Clinical Practice Guideline recommendation (with evidence review)

x 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**	**Siu L on behalf of the U. S. Preventive Services Taskforce. Screening for abnormal blood glucose and type 2 diabetes mellitus: U. S. Preventive Services Task Force recommendation. Ann Intern Med. 2015;163: 861-868.** **American Diabetes Association. Standards of medical care in diabetes—2018. Diabetes Care. 2018. (41) Supplement 1. Available at: http://care.diabetesjournals.org.**
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	The following evidence statements are quoted **verbatim** from the referenced clinical guidelines and other sources, where applicable: At least annual monitoring for the development of diabetes in those with prediabetes is suggested. (ADA, 2018**Error! Bookmark not defined.**) (E Recommendation) To test for prediabetes, fasting plasma glucose, 2-h plasma glucose during 75-g oral glucose tolerance test, and A1C are equally appropriate. (ADA, 2018**Error! Bookmark not defined.**) (B Recommendation)
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	See above evidence statements with grades
Provide all other grades and definitions from the evidence grading system	ADA Grading: Grade A • Clear evidence from well-conducted, generalizable randomized controlled trials that are adequately powered, including • Evidence from a well-conducted multicenter trial • Evidence from a meta-analysis that incorporated quality ratings in the analysis • Compelling nonexperimental evidence, i.e., “all or none” rule developed by the Centre for Evidence-Based Medicine at the University of Oxford • Supportive evidence from well-conducted randomized controlled trials that are adequately powered, including • Evidence from a well-conducted trial at one or more institutions • Evidence from a meta-analysis that incorporated quality ratings in the analysis Grade B • Supportive evidence from well-conducted cohort studies

	• Evidence from a well-conducted prospective cohort study or registry • Evidence from a well-conducted meta-analysis of cohort studies • Supportive evidence from a well-conducted case-control study Grade C • Supportive evidence from poorly controlled or uncontrolled studies • Evidence from randomized clinical trials with one or more major or three or more minor methodological flaws that could invalidate the results • Evidence from observational studies with high potential for bias (such as case series with comparison with historical controls) • Evidence from case series or case reports • Conflicting evidence with the weight of evidence supporting the recommendation E Expert consensus or clinical experience
Grade assigned to the **recommendation** with definition of the grade	Grade B • Supportive evidence from poorly controlled or uncontrolled studies • Evidence from randomized clinical trials with one or more major or three or more minor methodological flaws that could invalidate the results • Evidence from observational studies with high potential for bias (such as case series with comparison with historical controls) • Evidence from case series or case reports Grade E Expert consensus or clinical experience
Provide all other grades and definitions from the recommendation grading system	See above
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	In addition to the USPSTF and ADA guidelines, we reviewed over five evidence-based peer reviewed journal articles that confirmed the gap in care around prediabetes screening.
Estimates of benefit and consistency across studies	Strong evidence exits that rates of screening patients for prediabetes and undiagnosed diabetes are suboptimal in clinical care, especially in patients who are at high risk for developing type 2 diabetes. Approximately 1/3 of physicians

	reported screening patients for prediabetes according to guidelines (ADA and USPSTF). In a nationally representative sample of patients from the National Health and Nutrition Examination Survey (NHANES) from 2005-2012, only 45% of those who met screening criteria were screened.[i] Additionally, survey data show that while primary care physicians are aware of the guidelines that support screening for prediabetes, there is a disconnect between this knowledge and actual practice[ii,iii]
What harms were identified?	None
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	None

---

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.

This measure is based on evidence-based guidelines from the American Diabetes Association (ADA). The ADA recommends screening patients who are overweight or obese with one risk factor, regardless of age, and follow-up screening at one year for those diagnosed with prediabetes.

To test for prediabetes, fasting plasma glucose, 2-h plasma glucose during 75-g oral glucose tolerance test, and A1C are equally appropriate. (ADA, 20181) (B Recommendation)

The USPSTF recommends screening for abnormal blood glucose as part of cardiovascular risk assessment in adults aged 40 to 70 years of age who are overweight or obese. Clinicians should offer or refer patients with abnormal blood glucose to intensive behavioral counseling interventions to promote a healthful diet and physical activity. (USPSTF, 2015) (B recommendation)

<u>In addition to the above referenced guidelines, other evidence-based studies also support this measure:</u>

Geiss L. et al. Diabetes risk reduction behaviors among US adults with prediabetes. Am J Prev Med. 2010. 38(4): 403-409.

- Based on data from 1402 adults without diabetes (with preDM) who participated in the 2005-2006 National Health and Nutrition Examination Survey (NHNES) and who had valid fasting glucose and OGTTs.
- Almost 30% of the US adult population had preDM in 2005-2006 but only 7.3% were aware they had it.

- About half of adults with preDM reported performing DM risk reduction behaviors in the past year but only one third of adults with preDM received healthcare provider advice about these behaviors in the past year.

Kiefer M, et al. National patterns in diabetes screening: Data from the National Health and Nutrition Examination Survey (NHANES) 2005-2012. J Gen Intern Med. 2014*;30(5)*: 612-618.

- In a nationally representative sample (NHANES), only 45% of those who met ADA criteria (thought to be approximately 76.6% of the US population) for screening were actually screened.

Mehta S, Mocarski M, Wisniewski T, Gillepsie K, Narayan Venkat KM, Lang K. Primary care physician's utilization of type 2 diabetes screening guidelines and referrals to behavioral interventions: a survey linked retrospective study. BMJ Open Diab Res Care. 2017;5:e000406. Doi:10.1136/bmjdrc-2017-000406.

- Online survey of 305 primary care physicians regarding use of guidelines in screening for type 2 guidelines and referral to DPP and DSME for newly diagnosed patients with prediabetes or type 2 diabetes.
- Findings show a disconnect between physician perception of following guidelines and actual practice when physician survey responses are compared to EMR data.
- 38% of physicians reported using guidelines (33% used ADA only, 25% use ADA only)

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

A thorough literature review was conducted to identify evidence-based guidelines and other evidence, gaps in care with supportive evidence, and gaps in measurement to support the identification of measure concepts.

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

---

[i] Kiefer M, et al. National patterns in diabetes screening: Data from the National Health and Nutrition Examination Survey (NHANES) 2005-2012. J Gen Intern Med. 2014*;30(5)*: 612-618

[ii] Mehta S, Mocarski M, Wisniewski T, Gillepsie K, Narayan Venkat KM, Lang K. Primary care physician's utilization of type 2 diabetes screening guidelines and referrals to behavioral interventions: a survey-linked retrospective study. BMJ Open Diab Res Care. 2017;5:e000406. Doi:10.1136/bmjdrc-2017-000406.

[iii] Tseng E, Greer R C, O'Rourke, P, Yeh, H-C, McGuire, M M, Clark, J M, & Maruthur, N M. Survey of primary care providers' knowledge of screening for, diagnosing and managing prediabetes. Journal of General Internal Medicine, 32(11), 1172–1178.

1b. Performance Gap

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

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

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

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

At least annual glucose testing in patients who were previously found to have lab results in the range of prediabetes is an important aspect of care so that patients can be monitored for improvement or potential transition to Type 2 diabetes. While there are no current studies that show patients with prediabetes do not have follow-up testing completed, the TEP felt that this is a key area in which to have a measure. Preventing the onset of type 2 Diabetes by screening for prediabetes, with the goal of referring for treatment and prevention, will help to reduce the 84 million patients with prediabetes, and the overall number that eventually are diagnosed with type 2 diabetes. Furthermore, cost savings for prevention of type 2 diabetes have the potential to be significant. For every beneficiary, at 15 months, costs savings are at least $ 2,650.00 per person.

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.*

This measure has not yet been implemented

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.

Strong evidence exits that rates of screening patients for prediabetes and undiagnosed diabetes are suboptimal in clinical care, especially in patients who are at high risk for developing type 2 diabetes. Approximately 1/3 of physicians reported screening patients for prediabetes according to guidelines (ADA and USPSTF).

In a nationally representative sample of patients from the National Health and Nutrition Examination Survey (NHANES) from 2005-2012, only 45% of those who met screening criteria were screened. Furtherfore, follow-up screening of this sub-set of patients who are initially screened is even less prevalent.

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.*

This measure has not yet been implemented

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

There are no published studies to address this disparity at one year follow-up, but we would point to the studies that address the lack of screening for prediabetes as an initial reason why this measure is critical.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

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

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

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

n/a

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:** RetestGlucose_v5_8_Artifacts_20200106.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:** Copy_of_Retest_Abnormal_Blood_Glucose_Value_Sets_20200106.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.

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.

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

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

Patients who had a blood glucose test performed

*Retesting for abnormal blood glucose may include using a fasting plasma glucose, 2-h plasma glucose during a 75g oral glucose tolerance test, or A1C.

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).

See attached file in S.2a and S.2b for information to calculate the numerator

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

All patients aged 18 years and older who had an abnormal fasting plasma glucose, oral glucose tolerance test, or hemoglobin A1c result in the range of prediabetes in the year prior to the one-year measurement period

**Abnormal lab result in the range of prediabetes includes a fasting plasma glucose level between 100 mg/dL (5.6 mmol/L) to 125 mg/dL (6.9 mmol/L) OR a 2-hour glucose during a 75g oral glucose tolerance test between 140 mg/dL (7.8 mmol/L) to 199 mg/dL (11.0 mmol/L) OR and A1C between 5.7-6.4% (39-47 mmol/mol).

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).

See attached file in S.2a and S.2b for information to calculate the denominator

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

Denominator Exclusions:

Exclude patients who are pregnant.

Exclude patients who have any existing diagnosis of diabetes (Type 1, Type 2, latent autoimmune diabetes of adults [LADA], monogenic diabetes [MODY]).

Exclude patients in palliative care/hospice

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.)*

See attached file in S.2a and S.2b for information to calculate the exclusions

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

n/a

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

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Higher score

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

See attached file in S.2a for information to calculate the measure logic

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

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

n/a

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

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

n/a

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

If other, please describe in S.18.

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.

Measure data elements will be collected through health care organization electronic health record query, electronic health data queries.

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)*

Outpatient Services

If other:

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

n/a

2. Validity – See attached Measure Testing Submission Form

NQF_testing_attachment_Retesting_of_Glucose_for_Prediabetes.docx

2.1 For maintenance of endorsement

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

2.2 For maintenance of endorsement

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

2.3 For maintenance of endorsement

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

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

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Retesting of Abnormal Blood Glucose in Patients with Prediabetes
Date of Submission: 1/6/2020

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: Click here to describe	☐ other: Click here to describe

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*).

1.3. What are the dates of the data used in testing? Click here to enter date range

The measurement period (data collected from patients seen) was 8/1/2018 through 09/30/2019.

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: Click here to describe	☐ other: Click here to describe

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*)

Two sites were recruited and identified to collect data for testing and analysis at the data element level, reviewing individual patient records from the EHR and comparing to a manual review of the same cases. Testing was completed using a convenience sample, whereas sample size requirements were calculated based on estimated rates for each measure by site using a calculator based on the calculation defined by Donner-Eliasziw (see section 1.6 below).

Using specifications defined by the measure developer, both testing sites were able to access and test the critical data elements that included all components of the numerator, all components of the denominator, and all components of the exclusions. Testing was completed at the data element level and was completed on all patient cases in the sample. It should be noted that although the measure is specified at the physician

and physician group level, testing was completed at the individual data element level (as opposed to signal to noise), so therefore there would not be counts of physicians included in the analysis.

- Test Site #1: An ambulatory facility in South Carolina, part of a larger health system comprised of 8 inpatient hospitals and more than 100 outpatient facilities This facility uses Epic EHR in their facility.

- Test Site #2: An ambulatory facility in South Carolina, part of a larger system comprised of a 1,600+ bed comprehensive integrated health system, serving 1 million patients. This facility uses Cerner EHR in their facility.

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

As described in 1.5, we recruited and subcontracted with two sites to collect patient-level EHR data for the measure. Patient cases were included in the testing and analysis that met the following criteria:

- Adults ages 18 and over
- All races
- All genders

Due to the large size of the data available at each site, a sample of the patients at each site were identified through a validated process.

Sample size requirements were calculated based on estimated rates for each measure by site using a calculator based on the calculation defined by Donner-Eliasziw[iii]. A discussion and application of the use of the kappa statistic in reliability studies is available in Sim and Wright, 2005.[iii] These methods were instituted in order to ensure that reliability testing and analyses occur on data sets that have a large enough sample size to detect statistically significant differences, thus minimizing variation due to the play of chance.

The important variables for the kappa sample size calculation are as follows:

- The value for the expected proportion of positive ratings for the measure being tested could be based on available data on the average performance of clinicians on the measure. If the average performance is 90%, the proportion of positive ratings is 0.90.

- The standard assumptions for testing projects are to specify the 2-tailed test at 80% power required to detect a difference between the value of the calculated kappa statistic and the null value for kappa, for example a kappa of .090 versus the null value of kappa of 0.60. This tests whether the difference in the kappa values of 0.30 (0.6 versus 0.9) is significant.

Each site provided us with preliminary counts of patients meeting the numerator and denominator to be used in sample size calculations. Following is a table that displays the data reported from the sites, the recommended sample size from the sample calculator, and the actual sample size for which the site was asked to collect data. Due to the low counts provided by Site 1 for this measure, the recommended sample size was

larger than what was feasible due to time constraints. We decided that a sample of 112 would be enough for this analysis.

	Site 1	**Site 2**
Numerator (preliminary counts)	51	171
Denominator (preliminary counts)	53	516
Calculated Sample Size	344	62
Actual Sample Used for Testing	**112**	**75**

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.

Reliability and validity of the data elements and exclusion testing utilized the same data from the practice site's respective EHR systems of Epic and Cerner. Risk adjustment and stratification were not applied and not applicable for these measures.

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.

There were no social risk factors accounted for as elements in the measures. However, the Supplemental Data Elements in the measure specifications include language, race, ethnicity, and payor as elements that can be collected for each measure to allow for the stratification of measure results by these variables to assess disparities and initiate subsequent quality improvement activities.

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*)

By current NQF standards, data element validity testing results may be reported for reliability results. Testing description, results, and interpretation of results are given here and in 2b1. Validity Testing.

Data element validity testing was conducted utilizing Parallel Forms Reliability Testing methodology. Parallel forms reliability testing considers the analysis of agreement, through assessing the extent to which multiple formats or versions of data abstraction yield the same results. Verification of the data elements was obtained through automated data search strategies against a reference strategy (considered the gold standard) for obtaining the data elements. Manual review of the data elements was used as the reference strategy against which automated data search and extraction strategies were evaluated.

For this electronic clinical quality measure (eCQM), testing was used to determine if data elements found through electronic data pulls could be confirmed by manual abstraction of the same data elements. Testing at the level of the data elements allows for the analysis of each individual required data element included in the performance measure.

Interrater reliability (Cohen's Kappa coefficient) is used to assess the reliability of the measure based on results from two independent reviewers trained in the same way reviewing the same patient record. To perform inter-rater reliability testing we created an electronic data collection tool and trained the reviewers (raters) on its' use. The reviewers separately reviewed every sampled patient and collected all data elements necessary for computation of the performance measure (contained on the electronic data collection tool). Data received was analyzed using SAS to calculate frequencies, level of agreement, and agreement statistics (Cohen's Kappa). Cohen's Kappa coefficient measures inter-rater agreement for qualitative items and takes into account any agreement occurring by chance.

The following table displays the interpretation of the kappa statistic:

Kappa	**Strength of Agreement**[iii]
0.00	Poor
0.01 - 0.20	Slight
0.21 - 0.40	Fair
0.41 - 0.60	Moderate
0.61 - 0.80	Substantial
0.81 - 0.99	Almost Perfect

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

The data element validity testing results including Kappa scores are presented below

		% Agree	Kappa	N
Site 1	**Denominator**	98	0.927	112
	Numerator	94	0.850	17

Site 2	**Denominator**	100	1	75
	Exclusions	93	0.836	70
	Numerator	100	**	23

**Kappa scores not calculable with multiple non-responses by raters (i.e., all No/No or all Yes/Yes)

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?*)
For this measure, we find almost perfect levels of agreement for all data elements. Kappa scores ranged from .83 to 1.0, which is considered almost perfect.

_____

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)*

This measure was tested for data element validity testing, content validity, face validity, and feasibility of the data elements.

Data element validity testing was conducted utilizing Parallel Forms Reliability Testing methodology. Parallel forms reliability testing considers the analysis of agreement, through assessing the extent to which multiple formats or versions of data abstraction yield the same results. Verification of the data elements was obtained through automated data search strategies against a reference strategy (considered the gold standard) for obtaining the data elements. Manual review of the data elements was used as the reference strategy against which automated data search and extraction strategies were evaluated.

For this electronic clinical quality measure (eCQM), testing was used to determine if data elements found through electronic data pulls could be confirmed by manual abstraction of the same data elements. Testing at the level of the data elements allows for the analysis of each individual required data element included in the performance measure.

Interrater reliability (Cohen's Kappa coefficient) is used to assess the reliability of the measure based on results from two independent reviewers trained in the same way reviewing the same patient record. To perform inter-rater reliability testing we created an electronic data collection tool and trained the reviewers (raters) on

its' use. The reviewers separately reviewed every sampled patient and collected all data elements necessary for computation of the performance measure (contained on the electronic data collection tool). Data received was analyzed using SAS to calculate frequencies, level of agreement, and agreement statistics (Cohen's Kappa). Cohen's Kappa coefficient measures inter-rater agreement for qualitative items and takes into account any agreement occurring by chance.

The following table displays the interpretation of the kappa statistic:

Kappa	Strength of Agreement[iii]
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

Evidence of content validity is provided by looking for agreement among subject matter experts. The performance measure was assessed for content validity by a panel of technical expert work group members during the development process. This subject matter expert panel had representation from measure methodologists, patient advocacy groups, and clinical specialties. Additional input on the content validity of draft measures is obtained through a 30-day public comment period. All comments received are reviewed by the expert work group and the measures adjusted as needed.

For face validity, an external group of clinical and methodological experts assessed the measure for face validity through an on-line survey. The survey introduction provided the following definition of face validity: Face validity is the extent to which an empirical measurement appears to reflect that which it is supposed to "at face value." Face validity of an individual measure poses the question of how well the definition and specifications of an individual measure appear to capture the single aspect of care or healthcare quality as intended. Face validity of the measure score as an indicator of quality was systematically assessed as follows:

After the measure was fully specified, the expert panel was asked to rate their agreement with the following statement: "The scores obtained from the measure as specified will accurately differentiate quality across providers".

Scale 1-5, where 1= Strongly Disagree; 3=Neither Agree nor Disagree; 5=Strongly Agree, N/A = Not Applicable

The face validity panel included 22 panel members from the following organizations:

1. American Geriatric Society
2. American College of Occupational and Environmental Medicine
3. Omada Health
4. University of Chicago
5. Drexel University
6. Ascension St. John Detroit
7. American Society of Addiction Medicine
8. Tufts Medical Center
9. Rush University
10. National Institutes of Health
11. NorthShore University Healthcare
12. Northwestern Medicine
13. Rush University

14. Omada Health
15. Northwestern Medicine
16. Centers for Disease Control and Prevention
17. Emory University
18. Cincinnati Children's
19. Northwestern Medicine
20. Stony Brook Medicine
21. Advocate Healthcare
22. University of California San Francisco

Regarding feasibility of the data elements, a 2018 feasibility assessment was performed to assess the extent to which the required data are readily available, can be captured without undue burden, and are feasible for implementation within electronic health record systems. Two entities participated in the feasibility assessment for this measure.

• Test Site #1: a multispecialty academic medical center using EPIC EHR
• Test Site #2: a medical center using Matrix Care EHR

For this process, a testing methodology using a Data Element Tool (DET) to assess the availability of the data and the technical feasibility and implementation feasibility of the measures was employed. The DET is an Excel workbook designed to capture information that will determine whether or not each site can feasibly collect the data for the measures. It is structured to collect metadata about each data element necessary to construct each measure stored in the EHR. It will also collect information related to integrity and validity of data collection. Specifically, the DET is designed to capture the following information:

1. Data element information: Whether or not the data element is captured in the EHR, the data source application, primary user interface data location, data type, coding system, unit of measure, frequency of collection, and calculability within the measure context.
2. Measure integrity information: An assessment by the testing site as to what degree the measure, as specified, retains the originally stated intention of the measure.
3. Measure validity information: An assessment by the testing site as to what degree the scores obtained from the measure, as specified, will accurately differentiate quality performance across providers.

The DETs collected responses used to assess technical and implementation feasibility for each measure. Measure technical feasibility was defined as "Can my EHR do this?" and measure implementation feasibility was defined as "Will workflow be used consistently?" The responses were captured in the form of a rating using the following responses:

· "Feasible. Can do today."
· "Feasible with workflow mod/changes to EHR."
· "Non-feasible. Unable to do today."

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

The data element validity testing results including Kappa scores are presented below

		% Agree	Kappa	N
Site 1	Denominator	98	0.927	112
	Numerator	94	0.850	17
Site 2	Denominator	100	1	75
	Exclusions	93	0.836	70
	Numerator	100	**	23

**Kappa scores not calculable with multiple non-responses by raters (i.e., all No/No or all Yes/Yes)

For face validity, the panel rating of the validity statement for the measure were as follows:

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

N = 22; Mean rating = 4.14 and 86% of respondents either agree or strongly agree that this measure can accurately distinguish good and poor quality.

Frequency Distribution of Ratings

5 (Strongly Agree) – 8

4 (Agree) – 11

3 (Neither Agree nor Disagree) - 2

2 (Disagree) – 0

1 (Strongly Disagree) – 1

X (Not Applicable) – 0

For feasibility, overall, the measures are technically "Feasible. Can do today." in both EHR systems that tested the measures. The majority of the of the data elements are routinely collected as part of clinical care but additional time and programming resources would be needed to implement the missing elements below:

• Laboratory Test, Performed: 2-H Plasma Glucose During a 75g Oral Glucose Tolerance Test

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?*)

For this measure, we find almost perfect levels of agreement for all data elements. Kappa scores ranged from .83 to 1.0, which is considered almost perfect.

The results of the data element validity testing demonstrate that this measure is valid, supported by the results of the content validity, face validity, and feasibility testing that was conducted.

_____

2b2. EXCLUSIONS ANALYSIS

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

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*)

Each site initially pulled a random sample of patients. Site 1 pulled data that met the initial population criteria for the measure, and applied exclusion criteria to the denominator. The site provided a detailed spreadsheet that included the list of exclusions and reasons for exclusions that met the criteria. Site 2 pulled a random sample from the patient population and tested the exclusion criteria and applied inter-rater reliability testing using Cohen's Kappa Score.

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*)

Across the measured entities, there were 70 patients excluded for this measure, with a Kappa score of .836. Performance for this measure was 33%. We would expect a performance score within this range. Because the exclusions for this measure are also widely used in other diabetes-related measures, and are based on evidence-based clinical guidelines, the impact on performance is minimal.

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*)

With a Kappa score of .836, there is substantial agreement of reliability. The exclusions specified for this measure are consistent with other clinical exclusions that are used for measures with this clinical population. The testing results of the exclusions show moderate agreement for this measure.

Furthermore, the individual clinical exclusions specified in this measure are similar/and closely aligned with several already developed NQF endorsed measurement sets. The data elements for the measure exclusions are as follows: patients with diabetes, pregnancy, hospice care, ambulatory, and palliative care. The following NQF-endorsed measures have those data elements, so we are confident that the measure exclusions are appropriate and statistically demonstrate appropriateness:

- Diabetes Care for People with Serious Mental Illness: Hemoglobin A1c (HbA1c) Control (<8.0%): http://www.qualityforum.org/QPS/2608
    - Diabetes, Hospice Care Ambulatory
- Comprehensive Diabetes Care: Medical Attention for Nephropathy: http://www.qualityforum.org/QPS/0062
    - Hospice Care Ambulatory
- Diabetes: Foot Exam: http://www.qualityforum.org/QPS/0056
    - Diabetes
- Comprehensive Diabetes Care: Eye Exam (retinal) performed: http://www.qualityforum.org/QPS/0055
    - Diabetes
- Weight Assessment and Counseling for Nutrition and Physical Activity for Children/Adolescents (WCC) http://www.qualityforum.org/QPS/0024
    - Pregnancy, Hospice Care Ambulatory
- Depression remission at 12 months: http://www.qualityforum.org/QPS/0710e
    - Palliative Care

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

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

N/A

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

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with** Click here to enter number of factors **risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

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.

N/A

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.

N/A

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?

N/A

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)**

N/A

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

N/A

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.**

N/A

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)
N/A

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

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

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

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

2b3.9. Results of Risk Stratification Analysis:

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

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

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

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

Differences in performance were not tested, however during testing, performance was 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*)

Differences in performance were not tested

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?*)
Differences in performance were not tested

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*)

N/A

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*)

N/A

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*)

N/A

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*)

As part of the reliability testing, sites completed data element tables that assessed for missing elements. All elements that were missing on the sample cases were tracked. Since we found instances across the measures where more full and accurate information could be found in the manual abstraction process than through electronic reporting, this seems to be a consistent issue across all types of measures, not just this particular measure, given the nature of EHR capabilities and limitations.

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*)

We do not have the number of the overall frequency of 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*)

N/A

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: Copy_of_NQF_Feasibility_Scorecard_-_AMA_Retesting_For_Abnormal_Glucose.xlsx,Bonnie_Report_-_Retesting_of_Abnormal_BG.pdf

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.

N/A

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)*.

None

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
Public Reporting Payment Program	

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

This measure has not yet been implemented

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?*)
There are several discussions underway for this measure to be adopted and implemented in public programs, and we describe the plan and expected timeframes below in 4a 1.3

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.*)

Plans for this measure to be adopted and implemented in public programs are underway. The AMA's goal is for this measure to be included in the MIPS QPP to complement the prediabetes IAs, so we will be submitting to the CMS MUC list call for measures in 2020. Furthermore, CMS CMMI has reached out to the AMA to adopt this measure for the Maryland Primary Care Program´s public reporting program for 2021. Ongoing discussions are currently underway and there is a plan in place for this measure to be implemented into this program. Additionally, CMS has already met with the AMA to discuss this measure being included (as part of the set) in a Prediabetes MIPS Value Pathway (MVP) for the 2021 performance period.

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.

This measure has not yet been implemented

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.

This measure has not yet been implemented

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.

This measure has not yet been implemented

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

This measure has not yet been implemented

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

This measure has not yet been implemented

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

This measure has not yet been implemented

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.

This measure has not yet been implemented

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.

This measure has not yet been implemented

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

This measure has not yet been implemented

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

No

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

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

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

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

Yes

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

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.)
There are no competing measures for prediabetes, this the first set of measures in U.S. to address this condition.

Appendix

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

No appendix **Attachment:**

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): American Medical Association

Co.2 Point of Contact: Beth, Tapper, beth.tapper@ama-assn.org, 312-933-6636-

Co.3 Measure Developer if different from Measure Steward: American Medical Association

Co.4 Point of Contact: Beth, Tapper, beth.tapper@ama-assn.org, 312-933-6636-

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.

Ronald T Ackermann, MD, MPH (Co-Chair) Northwestern University

Senior Associate Dean for Public Health

Director, Institute for Public Health and Medicine (IPHAM) - Center for Community Health

Director, Center for Diabetes and Metabolism

Professor of Medicine (General Internal Medicine and Geriatrics), Medical Social Sciences and Medicine (Endocrinology)

William Golden, MD, MACP (Co-Chair) Professor of Medicine and Public Health

University of Arkansas for Medical Sciences

Medical Director

Arkansas DHS/Medicaid

Mary Carol Greenlee, MD, FACP, FACE Endocrinologist

Faculty for TCPi (national faculty and Colorado Practice Transformation Network faculty)

Mary E Krebs, MD

Family Medicine Physician and Faculty

HealthSource of Ohio and Soin Family Medicine Residency

Ameldia R. Brown MDiv, BSN, RN

Director Faith and Community Health

Henry Ford Health System; Henry Ford Macomb Hospital

Leslie Kolb, RN, BSN, MBA

Vice President of Science and Practice

American Association of Diabetes Educators

Jennifer Torres Mosst, PhD, MscPH, MSSW

Program Manager, Diabetes Prevention and Health System Strategies

Los Angeles County Department of Public Health

Tannaz Moin, MD, MBA, MSHS

Assistant Professor

UCLA and VA Greater Los Angeles

Anita Stewart, MD, MPH, JD

Medical Director for Medicare/Medicaid Programs

BlueCross BlueShield Illinois

Maria Prince, MD, MPH

Medical Director

Aetna

Laura Clapper, MD, MPPA, CPE, FAAPL

Regional Vice President

Anthem

Elizabeth Joy, MD, MPH

Physician, Medical Director

Community Health and Food & Nutrition

Intermountain Healthcare

Stephen Benoit, MD, MPH Medical Epidemiologist

Centers for Disease Control

James L. Rosenzweig, MD Endocrinologist

CDC Subject Matter Expert

Ann Albright, PhD, RD

AMA Staff

Kate Kirley, MD, MS

Karen Kmetik, PhD

Koryn Rubin

Beth Tapper, MA

Greg Wozniak, PhD

PCPI Foundation-consultants to this measure development project

Beth Bostrom, MPH

Kerri Fei, MSN, RN

Diedra Gray, MPH

Courtney Hurt, MSW, LCSW

Sam Tierney, MPH

Patrick Yep, MS, MPH

Technical expert panel members played a key role in the evidence review, development of the draft measures through an in-person consensus development process, and refinement and revision of the measures post-public comment. TEP members also helped with final measure revisions and approval of the measures in their current form.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2019

Ad.3 Month and Year of most recent revision: 12, 2019

Ad.4 What is your frequency for review/update of this measure? yearly

Ad.5 When is the next scheduled review/update for this measure? 12, 2020

Ad.6 Copyright statement:

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

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

Commercial uses of the Measures require a license agreement between the user and American Medical Association (AMA). The AMA shall not be responsible for any use of the Measures. The AMA encourages use of the Measures by other health care professionals, where appropriate.

Ad.8 Additional Information/Comments: We believe this measure, as part of the full measure set is necessary to reduce chronic disease burden. An estimated 30 million Americans have diabetes. This epidemic will continue to grow unless clinicians screen patients for prediabetes and manage at risk patients with preventive

interventions. This measure addresses important areas that are critical to quality of care, improved outcomes, and lowered costs in the prevention and treatment of chronic disease, specifically:

- Improving patient outcomes by preventing or delaying progression of type 2 diabetes
- Reducing medical expenditures associated with type 2 diabetes and its complications by identifying and addressing prediabetes before progression to type 2 diabetes
- Improving clinical practice burden associated with treating diabetes by referring patients for treatment of their prediabetes

The United States has 84 million adults with prediabetes, putting them at a higher risk for developing type 2 diabetes.