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 #: 0061

Corresponding Measures:

De.2. Measure Title: Comprehensive Diabetes Care: Blood Pressure Control (<140/90 mm Hg)

Co.1.1. Measure Steward: National Committee for Quality Assurance

De.3. Brief Description of Measure: The percentage of patients 18-75 years of age with diabetes (type 1 and type 2) whose most recent blood pressure level taken during the measurement year is <140/90 mm Hg.

1b.1. Developer Rationale: This measure aims to improve the quality of care for patients with diabetes by assessing whether their blood pressure was adequately controlled (<140/90 mm Hg). High blood pressure is a major risk factor for microvascular and macrovascular complications in patients with diabetes. Cardiovascular disease is one of the leading causes for early death in people with diabetes. Uncontrolled high blood pressure contributes to the risk of complications and early death due to heart attack, stroke, angina and coronary heart disease. The benefits of quality envisioned by this measure include controlled blood pressure in patients with diabetes and a reduction in complications and early death.

S.4. Numerator Statement: Patients whose most recent blood pressure level was <140/90 mm Hg during the measurement year.

S.6. Denominator Statement: Patients 18-75 years of age by the end of the measurement year who had a diagnosis of diabetes (type 1 and type 2) during the measurement year or the year prior to the measurement year.

S.8. Denominator Exclusions: This measure excludes adults in hospice. It also excludes adults with advanced illness and frailty, as well as Medicare adults 65 years of age and older enrolled in an I-SNP or living long-term in institutional settings.

Additionally, exclude patients who had a diagnosis of gestational diabetes or steroid-induced diabetes, in any setting, during the measurement year or the year prior to the measurement year and who did NOT have a diagnosis of diabetes. These patients are sometimes pulled into the denominator via pharmacy data. They are then removed once no additional diagnosis of diabetes (Type 1 or Type II) is found.

De.1. Measure Type: Outcome: Intermediate Clinical Outcome

S.17. Data Source: Claims, Electronic Health Data, Electronic Health Records, Paper Medical Records

S.20. Level of Analysis: Health Plan

IF Endorsement Maintenance – Original Endorsement Date: Aug 10, 2009 **Most Recent Endorsement Date:** Jun 30, 2015

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

0731:Comprehensive Diabetes Care

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

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

Preliminary Analysis: Maintenance of Endorsement

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

Criteria 1: Importance to Measure and Report

1a. Evidence

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

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

The developer provides the following evidence for this measure:

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

Changes to evidence from last review

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

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

Updates:

- The developer provided a summary of the links between monitoring blood pressue in patients with diabetes type 1 and type 2 with an improved health outcome of reducing microvascular and macrovascular complications, hospitalizations, and death.

1) The developer provided the following clinical practice guideline: The American Diabetes Association's Standards of Medical Care in Diabetes—2019. Diabetes Care *2019 Jan; 37(1): 11-34*. https://doi.org/10.2337/cd18-0105
    - **Recommendation:** "For individuals with diabetes and hypertension at lower risk for cardiovascular disease (10-year ASCVD risk <15%), treat to a blood pressure target of <140/90 mmHg. (Grade A)
    - Grade A definition:

- 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., the "all or none" rule developed by the Centre for Evidence-Based Medicine at 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
- **Recommendation:** "For individuals with diabetes and hypertension at higher cardiovascular risk (existing ASCVD or 10-year ASCVD risk >15%), a blood pressure target of <130/80 mmHg may be appropriate, if it can be safely attained." (Grade C)
- Grade C definition:
  - Supportive evidence from poorly controlled or uncontrolled studies, including:
    - 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 to historical controls)
    - Evidence from case series or case reports
  - Conflicting evidence with the weight of evidence supporting the recommendation

- The developer provided a Quantity, Quality, and Consistency of guideline.
- Harms identified include adverse effects of antihypertensive therapy.

2) The developer provided the following clinical practice guideline: American College of Cardiology/ American Heart Association Guidelines for the Prevention, Detection, Evaluation and Management of High Blood Pressure in Adults November 2017 . https://www.ahajournals.org/doi/pdf/10.1161/HYP.0000000000000065
   - Recommendation: ""In adults with DM and hypertension, antihypertensive drug treatment should be initiated at a BP of 130/80 mm Hg or higher with a treatment goal of less than 130/DBP: 80 mm Hg (1-8)." *Class I; Level B-R (systolic), Level C-EO (diastolic)"*
     - These guidelines redefined the diagnosis of hypertension moving from > 140/ 90 to a new definition of stage 1 hypertension (130-139/ 80-89) .
   - Grade defininition:
     - Level B-R (Randomized)
       - Moderate-quality evidence from 1 or more RCTs
       - Meta-analyses of moderate-quality RCTs
     - Level C-EO (Expert Opinion)
       - Consensus of expert opinion based on clinical experience
     - Class 1 (Strong) Benefit >>> Risk
   - The developer provided a Quantity, Quality, and Consistency of guideline.
   - Harms identified include adverse effects of antihypertensive therapy.

Questions for the Committee:

- For intermediate outcome measures:
  - 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?

- The measure assumes that all patients with DM should have their last BP reading below 140/90. A single reading of elevated BP is not diagnostic nor always indicative of poor control. What does this imply for the appropriateness of this measure?

Guidance from the Evidence Algorithm

Intermediate outcome measure with systematic review (Box 3) → Summary of the QQC provided (Box 4) → Systematic review concludes moderate quality evidence (Box 5b).
The highest possible rating is "High" for Evidence

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

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

Maintenance measures – increased emphasis on gap and variation

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

- The developer provided rate of patients with diabetes with a blood pressure level <140/90 mm Hg, extracted from HEDIS data, stratified by year (2016, 2017, 2018) and at the health plan level (commercial, Medicare, Medicaid). The data indicate variation exists.
    - Commercial (2016, 2017, 2018): mean (56-59%); SD (22-22%); Interquartile Range (16-18%)
    - Medicaid (2016, 2017, 2018): mean (60-63%); SD (12-14%); Interquartile Range (14-16%)
    - Medicare (2016, 2017, 2018): mean (63-69%); SD (11-17%); Interquartile Range (12-16%)

Disparities

- The developer submitted disparities data on the measure from the CMS Racial, Ethnic, and Gender Disparities in Health Care in Medicare Advantage Annual Report.
    - The report described racial and ethnic disparities among beneficiaries 18-75 years old with diabetes who had their blood pressure under control.
    - Hispanic and Asian or Pacific Islander beneficiaries outperformed White beneficiaries by more than 3 percentage points.
    - Asian or Pacific Islanders performed best at 79.4%, with Hispanics following behind at 74.8% while only 64.7% of Whites were likely to have their blood pressure controlled.
    - Overall, only 58.1% of Black beneficiaries with diabetes were likely to have their blood pressure controlled, with more than a 3% difference in performance compared to Whites.
- The developer is unable to collect performance data at the health plan level stratified by race, ethnicity, or language.
- The developer also cited literature by the CDC reports and the 2017 National Diabetes Statistics report noting disparities in minority groups affected by diabetes at higher rates than Whites. Also similar disparities in cardiovascular disease among patients with diabetes.
    - The percentage of patients with diabetes and blood pressure levels <140/90 mm Hg is also lower in minority groups. Seventy-seven percent of Non-Hispanic Whites had a blood pressure <140/90 mm Hg between 2003 and 2006 (CDC, 2013).
    - In comparison 71.6% of Non-Hispanic Blacks and 66.8% of Mexican Americans had blood pressure levels <140/90 mm Hg between 2003 and 2006 (CDC, 2013).

Questions for the Committee:

- Is there a gap in care that warrants a national performance measure?
-

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

RATIONALE:

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

1a. Evidence
Comments:
**The percentage of patients 18-75 years of age with diabetes (type 1 and type 2) whose most recent blood pressure level taken during the measurement year is <140/90 mm Hg, an intermediate outcome measure. The application refers to the ACC/AHA Guideline of 2017 and the ADA Standards of Medical Care of 2019. Personally, I agree with selecting the higher of the two levels from the ADA recommendations over the lower BP recommenation of both the ADA and the ACC/AHA guideline, but an explanation of the measure developers explaining this choice would be helpful.
**There is a very long and detailed review of the 140 v. 130 or 120 criterion. This remains highly controversial, and it's difficult to extrapolate findings from non DM to DM patients. Nonetheless, we probably should discuss the evidence base, which has evolved since the last review
**This is an intermediate outcome measure. It is in alignment with Grade A Evidence based guidelines from the American Diabetes Association.
**Not aware of any new studies/information
**The relationship between blood pressure and CV outcomes is strong.

1b. Performance Gap
Comments:
**HEDIS data from 2016-18 of commercial, Medicare, and Medicaid health plans show that the process measure was met in the upper 50s to upper 60s percentages.
**Very impressive gap remains.
**A performance gap exists, supporting the need for a national performance measure.
**Yes, overall less than optimal performance
**Gap exists and warrants a national performance measure.

Disparities:
Comments:
**HEDIS data does not currently stratify by race, ethnicity, or language. The CMS Office of Minority Health/RAND study indicates that such disparities do exist.
** Egregious disparities remain
** Data from a CMS Report supports disparities in achievement of blood pressure control
** Yes, black beneficiaries rate lowest among all populations
** Evidence of disparity by ethnic/race from CDC but not from this performance measure.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

Reliability

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

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

Validity

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

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

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

Evaluators: NQF Scientifc Methods Panel

Methods Panel Review (Combined)

Methods Panel Evaluation Summary:

Scientific Methods Panel Votes: Measure passes

- Reliability: H-2, M-3, L-0, I-0 → Measure passes with Moderate rating
- Validity: H-3, M-1, L-0, I-1 → Measure passes with HIGH rating

This measure was reviewed by the Scientific Methods Panel and discussed on the in-person meeting on October 21, 2019. A summary of the measure and the Panel discussion is provided below.

Reliability

- Testing of performance measure score with beta binomial reliability.
- Tests conducted on 394 commercial plans, 250 Medicaid plans and 477 Medicare plans; Overall reliability ranged from 0.976 to 0.998 respectively.
- Specifications: There was some confusion as per whether a patient would be counted as compliant if both systolic and diastolic BP values are in compliance or if a patient would be counted as compliant if systolic and diastolic BP values are in compliance.
- The appropriateness of the telehealth exclusion was questioned, especially from the rural lens.
- Beta-binomial results: One panelist questioned whether it is likely that the within plan variation is low enough to yield a reliability coefficient >.90, and suggested it would have been useful to see the plan level ICCs and a distribution of plan level rates with standard error bars. Discuss the need for this.

Validity

- To establish construct validity, the developer correlated this measure to other measures that are similarly focused on diabetic patients.
    - CDC: Hemoglobin A1c (HbA1c) Testing: The percentage of adults 18-75 with diabetes that had an HbA1c test performed during the measurement year.

    - CDC: HbA1c Control (<8.0%): The percentage of adults 18-75 with diabetes whose most recent HbA1c level is <8.0% during the measurement year.
    - CDC: HbA1c Poor Control (> 9.0%): The percentage of adults 18-75 with diabetes whose most recent HbA1c level is >9% during the measurement year.
    - CDC: Eye Exam (Retinal) Performed: The percentage of adults 18-75 with diabetes that had an eye screening for diabetic retinal disease during the measurement year.
- Correlation scores ranged from 0.41 to 0.89, indicating moderate to strong correlations.
- Reviewers expressed concern regarding lack of analysis on use of multiple data sources and comparability of results when more than one source was available for a plan; lack of clarity around the handling of missing data.
- Some reviewers expressed concern regarding the lack of risk adjustment for clinical factors and the developers rationale for this decision.

Summary from October 21, 2019 SMP In-Person Meeting:

Methods Panel Subgroup 1 briefly discussed three measures of diabetes care (0575, 0059, and 0061). The measures were found to be reliable and valid in the subgroup's preliminary analyses, but nonetheless, they were pulled for discussion regarding a common issue. The Panel asked the developer to consider the inherent similarities in the measures and explore their potential as a composite. The measure developer (NCQA) noted that there is both an NQF-endorsed composite measure *Optimal Diabetes Care* (NQF 0729), stewarded by Minnesota Community Measurement, as well as NCQA's own composite measure *Comprehensive Diabetes Care* (NQF 0731), which is no longer NQF-endorsed. The Panel also expressed concern that the three measures draw on multiple data sources, but a comparative analysis of the performance by data source was not provided. The Panel then urged the developer to carefully consider the impact of social risk on scoring and performance on the measures. The Panel was not convinced by the developer's argument against the need for risk adjustment and emphasized that many social risk factors may predispose certain populations to have lower performance rates on diabetes-related intermediate outcome measures.

Questions for the Committee regarding reliability:

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

Questions for the Committee regarding validity:

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

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

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

2a1. Reliability – Specifications

Comments:

**The collection of the data fields for this measure is complicated as is, but it should be noted that there are significant concerns in basing the decision to place a patient in the adequately controlled or poorly controlled groups on the basis of one single BP measurement. I have concerns for the NCQA allowance to exclude adults with advanced illness and frailty, as well as Medicare adults 65 years of age and older enrolled in an I-SNP or living long-term in institutional settings. It all depends on how this exclusion is used, but a threat exists that preventive actions of value to older individuals with very serious but managed medical conditions would be dropped or worse, these persons would be dropped from patient care rosters and not on the basis of a truly limited lifespan. Frailty is very difficult to measure. The low 2% exclusion under these rules is surprising. I would like a discussion regarding the low rate of applying these exclusions and that might address my concerns.

**The main issue is whether or not frailty can be captured as outlined. I rather doubt the accuracy of this method. Fortunately the affected population may be small, according to the working document

**This was rated moderate reliability by the Methods Review Panel.

**No concerns, the measure has been in use for some time and the panelist rank the measure as moderate. The expansion of the code set to make it more granular has been helpful. Panel only noted some concern with the inconsistencies across the different data sources and risk adjustment

**As eHealth grows, will developer reconsider use of that data?

2a2. Reliability – Testing

Comments:

**It is unrealistic to expect that the BP measurements used would be judged to be done properly, but how often is this done in a clinical setting. it was wise to exclude the BP measurements from emergency and urgent care settings. It is impractical that more than one BP measurement be utilized, rather than the most recent. However, these limitations should be recognized and discussed.

**I remain concerned about the discrepancy in results depending on who measures BP with what and where. That said, the measure itself is reliable if one ignores the unreliability of BP measurement itself

**No concerns

**No concerns

**No concerns

2b1. Validity – Testing

Comments:

**There should be a discussion of risk adjustment.

**no

**No concerns

**No concerns, high ranking and the differences in the risk adjustment and social risk likely present across the entire population (balencing the impact on equally)

**no concerns

2b4-7. Threats to Validity

Comments:

**Matching information elements from billing/structural medical plans' data and from medical records (paper and electronic) is complicated. What is the frequency of the data auditor declaring a source is "materially biased" and, hence, not be used?

**no

**No concerns

**No concerns

**No concerns

2b2-3. Other Threats to Validity

2b2. Exclusions

2b3. Risk Adjustment

Comments:

**See #7 and #8 above.

**For P4P purposes risk adjustment is important, but only if combined with unadjusted rates to show disparities.

**No concerns

**moderate

**no concerns

Scientific Acceptability: Combined Preliminary Analysis Form by NQF Scientific Methods Panel

Measure Number: 0061

Measure Title: Comprehensive Diabetes Care: Blood Pressure Control (<140/90 mm Hg)

Type of measure:

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

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

Data Source:

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

Level of Analysis:

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

Measure is:

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

RELIABILITY: SPECIFICATIONS

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

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

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

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

   Panel Member #3: I got a bit confused when reading the measure specifications as to if a patient would be counted as compliant if BOTH his/sysystolic and diatolic BP values are in compliance or if a patient would be counted as compliant if EITHER his/sysystolic and diatolic BP values are in compliance. At one point I was reading it was both – then at another – it seemed either.

   Another question I have is concerning why telehealth encounters are excluded (Section S.7). I get that the measure does not want to count the patient's e devices due to reliability and validity issues – but the telehealth exclusion seems problematic and restrictive. Ususally, when telehealth is used, health care providers are on both sides of the interaction and the blood pressure should be taken by a provider at the remote site. Why would this be excluded? This may produce bias to those in extreme rurals areas.

   Another concern I have is minor and it is more of a question than a concern. In the specifications for the denominator, it is indicated that the patient must have "a diagnosis of DM during the measurement year or the year prior." I wonder, how often is a patient's diagnosis of DM carried forward in EHR/chart documentation? That is, if the provider diagnosed the patient 10 years ago and knows the patient well, is he or she likely to enter DM as a diagnosis annually or even every other year? If so, if this a way to "game the system"? (e.g., If the provider has a patient who has a high blood pressure can he/she eliminate this patient from the denominator simply by not carrying the DM diagnosis forward in the record?)

   Panel Member #4: When using prescription drug claims it does not indicate if at least 2 prescriptions on different dates are required, does this imply only 1 prescription? If so, inconsistent with requiring at least 2

outpatient dx for diabetes on 2 different dates. Also, language around nonacute inpatient encounts without telehealth versus only 1 of 2 visits may be outpatient telealth directly below is confusing.

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

   Panel Member #1: The developer used a commonly used beta-binomial approach to assess reliability, given that this measure is specified for health plan, therefore denominator will be relatively large, the reliability score will tend to be very high and can even reach 1.

   Panel Member #3: The Beta-binomial model seems aproriate.

   Panel Member #4: Used beta-binominal approach measureing signal to noise. This is acceptable method.

   Panel Member #5: Beta-binomial testing performed – appropriate for this type of measure.

   Panel member #6: It is unlikely that the within plan variation is low enough to yield a reliability coefficient >.90. It would have been useful to see the plan level ICCs and a distribution of plan level rates with standard error bars.

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

   Submission document: Testing attachment, section 2a2.3

   Panel Member #1: It is not clear what overall reliability means in Table 2 (Testing form page 5, 2a2.3). For example, for commercial plan, it is not median (0.992), but 0.998 cannnot be the mean either based on the information included in Table 2.

   In general, reliability scores for health plans across product lines are quite high, but not all are greater than 0.7 as stated in the testing form (2a2.4).

   Panel Member #3: The number and types of health plans used is acceptable (commercial (n=394), Medicare (n=477), Medicaid (n=250). The overall relability, min, max and percentiles are included for each plan type. There is a high level of confidence that the measure results are reliable. My only concern is the sampling method described in S.15 uses systematic sampling – which is prone to bias.

   Panel Member #4: Shows strong reliability.

   Panel Member #5: Reliability level is acceptable (overall and median > 0.96). There is some variability in the minimum reliability scores, but 10^{th} to 90^{th} percentils are all > 0.9.

   Panel member #6: Difficult to interpret.

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 Panel Member #1:** with caution as noted above (large n).

   ☐ **No**

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

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

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

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

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

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

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

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

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

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**

    Panel Member #1: The reliability scores are all quite high, partly due to the way they are derived. Even very small variation among health plans can still produce rather high reliability scores for health plan with large n.

    Panel Member #2: Signal to noise measures appear high for all health plans. Minimum/median sample size of 411 appears adequate to assure reliability.

    Panel Member #3: I would have rated the reliability as high – except that the minimum reliability for the Medicaid plans was significantly lower than for the commercial and Medicare, e.g., 0.607). This should be explored.

    Panel Member #4: Overall measure shows high reliability. However,sSome concern that data element level testing was not conducted since there are multiple sources of data that can be used and if CPT codes ranges are allowed but if medical record a value must be provided, would like to see if these two methods have consistent results when both are available?

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

    Panel Member #1: The exclusion criteria seem to be quite expansive, for example, patients with heart failure (aged 18-75) are excluded from this measure. They are not uncommon among older patients.

    Panel Member #2: None

    Panel Member #3: No concerns.

    Panel Member #4: None.

Panel Member #5: Most exclusions were not formally tested – no concerns with those tested

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

    Submission document: Testing attachment, section 2b4.

    Panel Member #1: The testing of the difference (2b4.1 2nd paragraph, testing form) as described doesn't make sense. Is it really "an independent sample t-test of the performance difference between two randomly selected palns at the 25th and 75t percentile?", it seems to be just comparing two proportions.

    Panel Member #2: Substantial variation across plans.

    The 10th percentile of the Commercial plans looks anomalously low. Is this number correct?

    Panel Member #3: No concerns.

    Panel Member #4: None.

    Panel Member #5: None

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
    Submission document: Testing attachment, section 2b5.
    Panel Member #2: Ideally, would like to see comparison of performance rates in manually abstracted data and EMR data, but don't a priori assume bias.
    Panel Member #3: No concerns.

    Panel Member #4: The measure does utilize multiple data sources but did not address possible comparable results when more than one source was available for a plan.
    Panel Member #5: NA

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

    Submission document: Testing attachment, section 2b6.

    Panel Member #2: The treatment of missing data is not clearly described. There is discussion of "material bias" at the plan level and suppression of reports of data for specific plans but not discussion of how missing data at the patient level within plans is dealt with. This cannot be evaluated.

    Panel Member #3: No concerns.

    Panel Member #4: None.

16. **Risk Adjustment**

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

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

    ☒ Yes ☒ No ☐ Not applicable

    16c. **Social risk adjustment:**

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

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

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

    16d.**Risk adjustment summary:**

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

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

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

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

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

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

Panel Member #1: I think it is appropriate not to risk adjust for this measure.

Panel Member #2: Technically, there is stratification at the plan type level, but there is no risk adjustment within plan types, and I characterize this as no risk adjustment.

The sponsor cites studies without any detail that show no variation in appropriate care by SES and implies this demonstrates no need to risk adjust the outcome measure. However, while one of the rationales for SES adjustment is that patients of different backgrounds access the system differently, a second rationale is that the community and neighborhood contexts in which they implement prescribed treatments impose constraints that can affect outcomes such as Hb1Ac control, constraints due to available foods, opportunities for exercise, stress, and work, among others. This is not addressed in the discussion.

Beyond SES issues, the lack of risk adjustment within strata implies that the medical conditions and circumstances of patients within plans are sufficiently homogeneous across plans that no adjustment for factors that influence tractability of Hb1Ac levels is required. I'm skeptical of this and would like to hear from clinicians on this issue.

Panel Member #4: No rationale was presented for not risk adjusting for clinical factors**.** The rationale for not analyzing social risk factors was that the measure is specified to be reported separately for commercial, Medicaid and Medicare plan types which serves as a proxy for income and other socioeconomic risk factors. There is absolutely no rationale, evidence or literature cited to back up this claim which I would dispute is accurate on several levels.

Panel Member #5: Stratification by plan type allows appropriate comparisons**.**

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

17. **Are the specifications in alignment with the stated measure intent?**
    ☐ Yes ☐ Somewhat ☐ No (If "Somewhat" or "No", please explain)
18. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**

VALIDITY: TESTING

19. **Validity testing level: ☒ Measure score ☐ Data element ☐ Both**
20. **Method of establishing validity of the measure score:**
    ☐ Face validity
    ☒ Empirical validity testing of the measure score
    ☐ N/A (score-level testing not conducted)
21. **Assess the method(s) for establishing validity**

    Submission document: Testing attachment, section 2b2.2

    Panel Member #1: To establish construct validity, the developer correlated this measure to five other measures that are similarly focused on diabetic patients. Three are process measures and two are similar intermediate outcome measures. These are reasonable choices.

Panel Member #2: Correlation with other outcome and process measures

Panel Member #3: Construct validity was assessed by using a Pearson's correlation to compare the BP measure to a few process measures (A1C Testing, Eye exam performed) as well as a few other intermediate clinical outcomes measures (A1C control and A1C poor control). I have only a few concerns: 1) the rationale for comparing to the process measures is not clear to me; 2) Pearson correlations should be used for interval/ratio level data and, in my opinion, the process measures are not interval/ratio level.

Panel Member #4: They used construct validity using 4 measures that should be correlated to the CDC measure.

Panel Member #5: Construct validity tested via assessing correlation with other CDC measures.

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

    Submission document: Testing attachment, section 2b2.3

    Panel Member #1: The correlations are in general fair with the process measures but moderate to high with the intermediate outcome measures. That blood pressure control measure is at least moderately correlated with HbA1c control measure is supportive evidence of construct validity.

    Panel Member #3: No concerns.

    Panel Member #4: Some concerns about low correlations between blood pressure control and HbA1c testing (0.409 and 0.433 for Commercial and Medicare plans respectively. There was better correlation to the HbA2c control measures however, indicating better care which should be associated with better blood pressure control

    Panel Member #5: All correlations statistically and practically significant.

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

    Submission document: Testing attachment, section 2b1.

    ☒ **Yes**

    ☒ **No**

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

24. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?**
    NOTE that data element validation from the literature is acceptable.

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

    ☒ **Yes**

    ☐ **No**

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

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

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

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

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

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

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

    Panel Member #1: The two selected intermediate outcome measures are really only one as they are measuring almost the same thing (just from different perspectives, they are functionally related.) The moderate to high correlation between blood pressure control measure and HbA1c control measure is reassuring, but that it varies substantially across product lines (commercial 0.888, Medicaid 0.756, Medicare 0.583) is somewhat surprising.

    Panel Member #2: Treatment of missing data not clearly described.

    Lack of risk adjustment is questionable.

    Panel Member #3: Construct validity was assessed by using a Pearson's correlation to compare the BP measure to a few process measures (A1C Testing, Eye exam performed) as well as a few other intermediate clinical outcomes measures (A1C control and A1C poor control). I have only a few concerns: 1) the rationale for comparing to the process measures is not clear to me; 2) Pearson correlations should be used for interval/ratio level data and, in my opinion, the process measures are not interval/ratio level.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

    ☐ **High**

    ☐ **Moderate**

    ☐ **Low**

    ☐ **Insufficient**

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

ADDITIONAL RECOMMENDATIONS

29. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**

    Panel Member #2: The lack of risk adjustment merits wider discussion. It is not just a statistical issue.

Criterion 3. Feasibility

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

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

Data Specifications and Elements

- The measure has Information gathered through multiple data sources (administrative data, electronic clinical data, and paper records).
- Some data elements are in defined fields in electronic sources.

- This is not an eMeasure

Data Collection Strategy

- NCQA conducts an independent audit of all HEDIS collection and reporting processes, as well as an audit of the data which are manipulated by those processes, in order to verify that HEDIS specifications are met.
- The developer also has a Policy Clarification Support System for any inquiries on the measure.
- The measure for broad public use is encouraged by developer. Written consent would be required for any "commercial use".

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?

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

RATIONALE:

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. Feasibility

Comments:

**What has been the feedback from the health care plans on the time/effort required to collect data on this measurement (combining information on the same patient from more than one source, billing/structural patient information and medical record information) compared to measures not requiring such "pairing"?

**OK

**The data required to assess this measure is generated in the routine delivery of care

**Moderate as there are various data collection methods

**Data elements routinely generated and available in electronic form.

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

- DIABETES RECOGNITION PROGRAM: This measure is used NCQA's Diabetes Recognition Program (DRP) that assesses clinician performance on key quality measures that are based on national evidence based guidelines in diabetes care.
- HEALTH PLAN ACCREDITATION: This measure is used in scoring for accreditation of Medicare Advantage health plans.
- HEALTH PLAN RANKINGS/REPORT CARDS: This measure is used to calculate health plan rakings which are reported on the NCQA website.
- INTEGRATED HEALTHCARE ASSOCIATION (IHA) CALIFORNIA PAY FOR PERFORMANCE: This measure is used in the California P4P program which is the largest non-governmental physician incentive program in the United States.
- QUALITY COMPASS: This measure is used in Quality Compass which is an indispensable tool used for selecting a health plan, conducting competitor analysis, examining quality improvement and benchmarking plan performance.
- ACCOUNTABLE CARE ORGANIZATION ACCREDITATION: This measure is used in NCQA's ACO Accreditation program, that helps health care organizations demonstrate their ability to improve quality, reduce costs and coordinate patient care.
- STATE OF HEALTH CARE ANNUAL REPORT: This measure is publicly reported nationally and by geographic regions in the NCQA State of Health Care annual report.

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

- NCQA publishes HEDIS results annually in its Quality Compass tool. The measure receives feedback through the Policy Clarification Support System. The feedback received has generally been centered around clarification on the use of blood pressure readings obtained during potentially stressful procedures or specific visit types, suggestions for exclusions, and the use of patient-reported readings.
- NCQA has provided minor clarifications about the measure during the annual update process in order to address questions received through the Policy Clarification Support system.

Additional Feedback:

The developer/steward did not provide any further feedback.

***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**

RATIONALE:

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

- From 2016-2018, performance on this measure has generally improved (2-5%) across the commercial, Medicare, and Medicaid product lines. Current average performance (MY 2018) is highest in Medicare plans (69%), followed by Medicaid plans (62%), and then commercial plans (59%).

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

- Per developer, there were no identified unexpected findings (positive or negative) during testing or since implementation of this measure.

Potential harms

- The developer did not report any unintended consequences.

Additional Feedback:

- This measure was last reviewed by the former NQF Endocrine Standing Committee in 2015. Endocrine Measures Final Report-Cycle 3. Some previous discussions at that time by the former Standing Committee revolved whether the evidence supported the blood pressure range of less than 140/90.

Questions for the Committee:

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

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

RATIONALE:

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

4a1. Use - Accountability and Transparency

Comments:

**Explained well on pages 58-61.

**It would be great to have a balancing measure regarding consequences of over-aggressive Rx. Also, to achieve good results, physicians may begin treating people based on an office measure that does not reflect that actual BP of the patient. It's too easy to treat and check the box.

**This measure is used in several accountability and public reporting programs

**Yes

**Publicly reported and used in accountability programs. Yes feedback obtained.

4b1. Usability – Improvement

Comments:

**See above re balancing measures and concerns.

**Use of this measure may help further the acheivement of blood pressure control.

**benefits outweigh harm

**Modest improvement reported. Benefits outweigh harms.

Criterion 5: Related and Competing Measures

Related or competing measures

The developer identified one related measure in the Primary Care and Chronic Illness project:

- 0729 Optimal Diabetes Care (Minnesota Community Measurement)

NQF staff noted there are other blood pressure related measures in other NQF CDP projects:

- 0018 Controlling High Blood Pressure (NCQA)
  - The percentage of adults 18-85 years of age who had a diagnosis of hypertension (HTN) and whose blood pressure was adequately controlled (<140/90 mm Hg) during the measurement year.
- 2602 Controlling High Blood Pressure for People with Serious Mental Illness (NCQA)
  - The percentage of patients 18-85 years of age with serious mental illness who had a diagnosis of hypertension (HTN) and whose blood pressure (BP) was adequately controlled during the measurement year.
- 2606 Diabetes Care for People with Serious Mental Illness: Blood Pressure Control (<140/90 mm Hg) (NCQA)
  - The percentage of patients 18-75 years of age with a serious mental illness and diabetes (type 1 and type 2) whose most recent blood pressure (BP) reading during the measurement year is <140/90 mm Hg.
- 0073 Ischemic Vascular Disease (IVD): Blood Pressure Control (NCQA)
  - The percentage of patients 18 to 75 years of age who were discharged alive with acute myocardial infarction (AMI), coronary artery bypass graft (CABG) or percutaneous coronary interventions (PCI) during the 12 months prior to the measurement year, or who had a diagnosis of ischemic vascular disease (IVD) during the measurement year and the year prior to the measurement year and who had the following during the measurement year: Blood pressure control (BP): reported as under control <140/90 mm Hg.
- 0076 Optimal Vascular Care (Minnesota Community Measurement)
  - The percentage of patients 18-75 years of age who had a diagnosis of ischemic vascular disease (IVD) and whose IVD was optimally managed during the measurement period as defined by achieving ALL of the following:
    • Blood pressure less than 140/90 mmHg
    • On a statin medication, unless allowed contraindications or exceptions are present
    • Non-tobacco user
    • On daily aspirin or anti-platelet medication, unless allowed contraindications or exceptions are present

Harmonization

- The developer noted key differences between 0729 and 0061 including:
  - 0729 is a composite measure where 0061 is a single measure.
  - 0729 is physician level of analysis and 0061 is health plan level

- 0729 uses medical record data source and 0061 uses administrative claims and medical records
- Measures are collected from different data sources by different entities

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

5. Related and Competing

Comments:

**This measure (0061) is included in the composite diabetes measure (0729). The differences in structure and analysis are explained on pages 62-63.

**no

**Harmonized

**A number of measures address blood pressure goals in different patient groups, either as a single measure or part of a composite.

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 01/31/2020

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

Brief Measure Information

NQF #: 0061

Corresponding Measures:

De.2. Measure Title: Comprehensive Diabetes Care: Blood Pressure Control (<140/90 mm Hg)

Co.1.1. Measure Steward: National Committee for Quality Assurance

De.3. Brief Description of Measure: The percentage of patients 18-75 years of age with diabetes (type 1 and type 2) whose most recent blood pressure level taken during the measurement year is <140/90 mm Hg.

1b.1. Developer Rationale: This measure aims to improve the quality of care for patients with diabetes by assessing whether their blood pressure was adequately controlled (<140/90 mm Hg). High blood pressure is a major risk factor for microvascular and macrovascular complications in patients with diabetes. Cardiovascular disease is one of the leading causes for early death in people with diabetes. Uncontrolled high blood pressure contributes to the risk of complications and early death due to heart attack, stroke, angina and coronary heart disease. The benefits of quality envisioned by this measure include controlled blood pressure in patients with diabetes and a reduction in complications and early death.

S.4. Numerator Statement: Patients whose most recent blood pressure level was <140/90 mm Hg during the measurement year.

S.6. Denominator Statement: Patients 18-75 years of age by the end of the measurement year who had a diagnosis of diabetes (type 1 and type 2) during the measurement year or the year prior to the measurement year.

S.8. Denominator Exclusions: This measure excludes adults in hospice. It also excludes adults with advanced illness and frailty, as well as Medicare adults 65 years of age and older enrolled in an I-SNP or living long-term in institutional settings.

Additionally, exclude patients who had a diagnosis of gestational diabetes or steroid-induced diabetes, in any setting, during the measurement year or the year prior to the measurement year and who did NOT have a diagnosis of diabetes. These patients are sometimes pulled into the denominator via pharmacy data. They are then removed once no additional diagnosis of diabetes (Type 1 or Type II) is found.

De.1. Measure Type: Outcome: Intermediate Clinical Outcome

S.17. Data Source: Claims, Electronic Health Data, Electronic Health Records, Paper Medical Records

S.20. Level of Analysis: Health Plan

IF Endorsement Maintenance – Original Endorsement Date: Aug 10, 2009 **Most Recent Endorsement Date:** Jun 30, 2015

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

0731:Comprehensive Diabetes Care

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

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

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

BP_Control_Evidence_Form_-61-.docx

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

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

Yes

1a. Evidence (subcriterion 1a)

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0061

Measure Title: Comprehensive Diabetes Care: Blood Pressure Control (<140/90 mm Hg)

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

Date of Submission: 8/1/2019

Instructions

- ***Complete 1a.1 and 1a.2 for all measures. If instrument-based measure, complete 1a.3.***
- ***Complete EITHER 1a.2, 1a.3 or 1a.4 as applicable for the type of measure and evidence.***
- ***For composite performance measures:***
  - ***A separate evidence form is required for each component measure unless several components were studied together.***
  - ***If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.***
- **All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.**
- **If you are unable to check a box, please highlight or shade the box for your response.**
- **Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.**

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- **Outcome: [3] Empirical data demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service. If not available, wide variation in performance can be used as evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.**
- **Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.**

- **Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.**
- **Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.**
- **Efficiency: [6] evidence not required for the resource use component.**
- **For measures derived from patient reports, evidence should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.**
- **Process measures incorporating Appropriate Use Criteria: See NQF's guidance for evidence for measures, in general; guidance for measures specifically based on clinical practice guidelines apply as well.**

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines and/or modified GRADE.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

☐ Outcome:

☐ Patient-reported outcome (PRO):

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

☒ Intermediate clinical outcome (*e.g., lab value*): Blood Pressure <140/90 mm Hg

☐ Process:

☐ Appropriate use measure:

☐ Structure:

☐ Composite:

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

Patient 18-75 years of age with diabetes (type 1 and type 2) >>> Health care provider monitors patient's blood pressure level >>> Patient's blood pressure level result is <140/90 mm Hg (adequately controlled) >>> Patient has a significant reduction in microvascular and macrovascular complications, hospitalization, and death.

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

N/A

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

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

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

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

☒Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Table 1. American Diabetes Association (ADA) Guidelines

Source of Systematic Review: **2) Title** **3) Author** **4) Date** **5) Citation, including page number** 6) **URL**	**7) Standards of Medical Care in Diabetes–2019.** **8) American Diabetes Association** **9) January 2019** **10) The American Diabetes Association's Standards of Medical Care in Diabetes—2019. Diabetes Care *2019 Jan; 37(1): 11-34*. https://doi.org/10.2337/cd18-0105** **11) URL: https://care.diabetesjournals.org/content/42/Supplement_1**
Quote the guideline or recommendation verbatim about the process, structure or	Recommendations (2019)

intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	12) For individuals with diabetes and hypertension at lower risk for cardiovascular disease (10-year ASCVD risk <15%), treat to a blood pressure target of <140/90 mmHg. (Grade A) 13) For individuals with diabetes and hypertension at higher cardiovascular risk (existing ASCVD or 10-year ASCVD risk >15%), a blood pressure target of <130/80 mmHg may be appropriate, if it can be safely attained. (Grade C)
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	The grades assigned by ADA to the guideline varied by the guideline recommendation. The grades varied from A – C. See question above for the grade given to each guideline. Level of Evidence & Description: Level A • Clear evidence from well-conducted, generalizable, randomized controlled trials that are adequately powered, including: o Evidence from a well-conducted multicenter trial o Evidence from a meta-analysis that incorporated quality ratings in the analysis • Compelling nonexperimental evidence, i.e., the "all or none" rule developed by the Centre for Evidence-Based Medicine at Oxford • Supportive evidence from well-conducted randomized controlled trials that are adequately powered, including: o Evidence from a well-conducted trial at one or more institutions o Evidence from a meta-analysis that incorporated quality ratings in the analysis Level C • Supportive evidence from poorly controlled or uncontrolled studies, including: o Evidence from randomized clinical trials with one or more major or three or more minor methodological flaws that could invalidate the results o Evidence from observational studies with high potential for bias (such as case series with comparison to historical controls) o Evidence from case series or case reports • Conflicting evidence with the weight of evidence supporting the recommendation
Provide all other grades and definitions from the evidence grading system	Level of Evidence & Description: Level B • Supportive evidence from well-conducted cohort studies, including: o Evidence from a well-conducted prospective cohort study or registry o Evidence from a well-conducted meta-analysis of cohort studies

	• Supportive evidence from a well-conducted case-control study E Expert consensus or clinical experience
Grade assigned to the **recommendation** with definition of the grade	No additional grading was provided, grades assigned to evidence is the same with grades assigned to recommendations.
Provide all other grades and definitions from the recommendation grading system	No additional grading was provided, grades assigned to evidence is the same with grades assigned to recommendations.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	Quantity and Quality The ADA references four main randomized controlled trials to support the recommendations around blood pressure targets: • ACCORD BP - 4,733 participants with type 2 diabetes aged 40–79 years with prior evidence of CVD or multiple cardiovascular risk factors • ADVANCE BP - 11,140 participants with type 2 diabetes aged 55 years and older with prior evidence of CVD or multiple cardiovascular risk factors • HOT - 18,790 participants, including 1,501 with diabetes • SPRINT - 9,361 participants without diabetes Below is an excerpt of their discussion of these studies: "Randomized clinical trials have demonstrated unequivocally that treatment of hypertension to blood pressure <140/90 mmHg reduces cardiovascular events as well as microvascular complications. Therefore, patients with type 1 or type 2 diabetes who have hypertension should, at a minimum, be treated to blood pressure targets of <140/90 mmHg. The benefits and risks of intensifying antihypertensive therapy to target blood pressures lower than <140/90 mmHg (e.g., <130/80 or <120/80 mmHg) have been evaluated in large randomized clinical trials and meta-analyses of clinical trials. Notably, there is an absence of high-quality data available to guide blood pressure targets in type 1 diabetes. The Action to Control Cardiovascular Risk in Diabetes blood pressure (ACCORD BP) trial provides the strongest direct assessment of the benefits and risks of intensive blood pressure control among people with type 2 diabetes. In ACCORD BP, compared with standard blood pressure control (target systolic blood pressure <140 mmHg), intensive blood pressure control (target systolic blood pressure <120 mmHg) did not reduce total

	major atherosclerotic cardiovascular events but did reduce the risk of stroke, at the expense of increased adverse events. The ACCORD BP results suggest that blood pressure targets more intensive than <140/90 mmHg are not likely to improve cardiovascular outcomes among most people with type 2 diabetes but may be reasonable for patients who may derive the most benefit and have been educated about added treatment burden, side effects, and costs, as discussed below. Additional studies, such as the Systolic Blood Pressure Intervention Trial (SPRINT) and the Hypertension Optimal Treatment (HOT) trial, also examined effects of intensive versus standard control, though the relevance of their results to people with diabetes is less clear. The Action in Diabetes and Vascular Disease: Preterax and Diamicron MR Controlled Evaluation–Blood Pressure (ADVANCE BP) trial did not explicitly test blood pressure targets; the achieved blood pressure in the intervention group was higher than that achieved in the ACCORD BP intensive arm and would be consistent with a target blood pressure of <140/90 mmHg. Notably, ACCORD BP and SPRINT measured blood pressure using automated office blood pressure measurement, which yields values that are generally lower than typical office blood pressure readings by approximately 5–10 mmHg (30), suggesting that implementing the ACCORD BP or SPRINT protocols in an outpatient clinic might require a systolic blood pressure target higher than <120 mmHg, such as <130 mmHg. A number of post-hoc analyses have attempted to explain the apparently divergent results of ACCORD BP and SPRINT. Some investigators have argued that the divergent results are not due to differences between people with and without diabetes but rather are due to differences in study design or to characteristics other than diabetes. Others have opined that the divergent results are most readily explained by the lack of benefit of intensive blood pressure control on cardiovascular mortality in ACCORD BP, which may be due to differential mechanisms underlying cardiovascular disease in type 2 diabetes, to chance, or both.”
Estimates of benefit and consistency across studies	Excerpt from ADA discussion of benefit: “Patients and clinicians should engage in a shared decision-making process to determine individual blood pressure targets. This approach acknowledges that the benefits and risks of intensive blood pressure targets are uncertain and may vary across patients and is consistent with a patient-focused approach to care that values patient priorities and provider judgment. Secondary analyses of ACCORD BP and SPRINT suggest that clinical factors can help determine individuals more likely to benefit and less likely to be harmed by intensive blood pressure control. Absolute benefit from blood pressure reduction correlated with absolute baseline cardiovascular risk in SPRINT and in earlier clinical trials conducted

	at higher baseline blood pressure levels. Extrapolation of these studies suggests that patients with diabetes may also be more likely to benefit from intensive blood pressure control when they have high absolute cardiovascular risk. Therefore, it may be reasonable to target blood pressure <130/80 mmHg among patients with diabetes and either clinically diagnosed cardiovascular disease (particularly stroke, which was significantly reduced in ACCORD BP) or 10-year ASCVD risk ≥15%, if it can be attained safely. This approach is consistent with guidelines from the American College of Cardiology/American Heart Association, which advocate a blood pressure target <130/80 mmHg for all patients, with or without diabetes."
What harms were identified?	Excerpt from ADA discussion of harms: "Potential adverse effects of antihypertensive therapy (e.g., hypotension, syncope, falls, acute kidney injury, and electrolyte abnormalities) should also be taken into account. Patients with older age, chronic kidney disease, and frailty have been shown to be at higher risk of adverse effects of intensive blood pressure control. In addition, patients with orthostatic hypotension, substantial comorbidity, functional limitations, or polypharmacy may be at high risk of adverse effects, and some patients may prefer higher blood pressure targets to enhance quality of life. Patients with low absolute cardiovascular risk (10-year ASCVD risk <15%) or with a history of adverse effects of intensive blood pressure control or at high risk of such adverse effects should have a higher blood pressure target. In such patients, a blood pressure target of <140/90 mmHg is recommended, if it can be safely attained."
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	There have been no new studies that contradict the current body of evidence.

Table 2. ACC/AHA 2017 Guideline

Source of Systematic Review: 14) **Title** 15) **Author** 16) **Date** 17) **Citation, including page number**	**19) Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults** **20) ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCN** **21) November 2017** **22) Whelton PK, Carey RM, Aronow WS, Casey DE Jr, Collins KJ, Dennison Himmelfarb C, DePalma SM, Gidding S, Jamerson KA, Jones DW, MacLaughlin EJ, Muntner P, Ovbiagele B, Smith SC Jr, Spencer CC, Stafford RS, Taler SJ, Thomas RJ, Williams KA Sr, Williamson JD, Wright**

18) **URL**	**JT Jr. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Hypertension. 2018;71:e13–e115. DOI: 10.1161/ HYP.0000000000000065.** **23) URL:** **https://www.ahajournals.org/doi/pdf/10.1161/HYP.0000000000000065**
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.	Recommendation 1: "In adults with DM and hypertension, antihypertensive drug treatment should be initiated at a BP of 130/80 mm Hg or higher with a treatment goal of less than 130/DBP: 80 mm Hg (1-8)." *Class I; Level B-R (systolic), Level C-EO (diastolic)*
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	The grades assigned by the ACC and AHA to the evidence associated with the recommendation varied by the guideline recommendation. See the question above for the grade given to each guideline recommendation. Grade of Evidence: Level B-R (Randomized) • Moderate-quality evidence from 1 or more RCTs • Meta-analyses of moderate-quality RCTs Level C-EO (Expert Opinion) • Consensus of expert opinion based on clinical experience
Provide all other grades and definitions from the evidence grading system	Grade of Evidence: Level A • High-quality evidence from more than 1 RCT • Meta-analyses of high-quality RCTs • One or more RCTs corroborated by high-quality registry studies Level B-NR (Nonrandomized) • Moderate-quality evidence from 1 or more well-designed, well-executed nonrandomized studies, observational studies, or registry studies • Meta-analyses of such studies Level C-LD (Limited Data) • Randomized or nonrandomized observational or registry studies with limitations of design or execution • Meta-analyses of such studies

	• Psychological or mechanistic studies in human subjects
Grade assigned to the **recommendation** with definition of the grade	The grades assigned by the ACC and AHA to the guideline varied by the guideline recommendation. See the question above for the grade given to each guideline recommendation. Grade of Recommendation: Class 1 (Strong) Benefit >>> Risk Suggested phrases for writing recommendations: • Is recommended • Is indicated/useful/effective/beneficial • Should be performed/administered/other • Comparative-Effectiveness Phrases †: o Treatment/strategy A is recommended/indicated in preference to treatment B o Treatment A should be chosen over treatment B † For comparative-effectiveness recommendations (COR 1 and IIa; LOE A and B only), studies that support the use of comparator verbs should involve direct comparisons of the treatments or strategies being evaluated
Provide all other grades and definitions from the recommendation grading system	Grade of Recommendation: Class IIa (Moderate) Benefit >> Risk Suggested phrases for writing recommendations: • Is reasonable • Can be useful/effective/beneficial • Comparative-Effectiveness Phrases †: o Treatment/strategy A is probably recommended/indicated in preference to treatment B o It is reasonable to choose treatment A over treatment B Class IIb (Weak) Benefit ≥ Risk Suggested phrases for writing recommendations: • May/might be reasonable • May/might be considered • Usefulness/effectiveness is unknown/unclear/uncertain or not well established Class III: No Benefit (Moderate) Benefit = Risk Suggested phrases for writing recommendations: • Is not recommended • Is not indicated/useful/effective/beneficial • Should not be performed/administered/other

	Class III: Harm (Strong) Risk > Benefit Suggested phrases for writing recommendations: • Potentially harmful • Causes harm • Associated with excess morbidity/mortality • Should not be performed/administered/other † For comparative-effectiveness recommendations (COR 1 and IIa; LOE A and B only), studies that support the use of comparator verbs should involve direct comparisons of the treatments or strategies being evaluated
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	Recommendation 1 Quantity: • A total of 33 publications from 15 studies were considered for inclusion. Publication dates ranged from 1998 through 2015. Target BPs varied for the lower and standard BP groups with 9 studies having a systolic BP target <130 mm Hg for the lower therapy group (48, 49, 56-64). Many of the studies included patients with comorbid conditions such as diabetes mellitus, CKD, or were at high risk for cardiovascular disease or progression to end-stage renal disease. Most studies excluded those with prior or recent MI or stroke, secondary hypertension, CHF, or other serious illnesses. In only 3 studies did all participants have diabetes mellitus or CKD. Quality: • There is limited quality evidence to determine a precise BP target in adults with DM. No RCTs have explicitly 1) documented whether treatment to an SBP goal <140 mm Hg versus a higher goal improves clinical outcomes in adults with hypertension and DM or 2) directly evaluated clinical outcomes associated with SBP <130 mm Hg (2). However, 2 high-quality systematic reviews of RCTs support an SBP target of <140 mm Hg (4, 7). • For each outcome, the relative risk (RR) and 95% CI from the number of events and participants were calculated. The risk of each outcome for any lower BP target versus any standard BP target was examined first, then the effect of a lower systolic BP target <130 mm Hg versus any higher BP target for all outcomes. In prior guidelines, there was insufficient evidence to demonstrate benefit of a BP goal <140/90 mm Hg (80). Given the completion of additional studies with lower target BP goals, the ACC/AHA examined available evidence to determine whether a lower BP goal conferred additional benefit either in the general population or in a specific subpopulation. The lower goal of 130 mm Hg was selected because it was the lower limit of high-normal BP and was the goal BP set by other guidelines for certain subpopulations. There were an adequate

	number of studies with a BP target of ≤130 mm Hg to study the question. Publication dates ranged from 1998 through 2015. • Meta-analysis and systematic review of trials that compare more intensive BP reduction to standard BP reduction among adults with elevated ASCVD risk, report that more intense BP lowering significantly reduces the risk of stroke, coronary events, major cardiovascular events, and cardiovascular mortality (1). In a stratified analysis of these data, achieving an additional 10–mm Hg reduction in SBP reduced CVD risk when compared with an average SBP of 158/82 to 143/76 mm Hg, 144/85 to 137/81 mm Hg, and 134/79 to 125/76 mm Hg. Patients with DM and CKD were included in the analysis (1, 2, 11-13, 18). (Specific management details are in Section 9.3 for CKD and Section 9.6 for DM.)
Estimates of benefit and consistency across studies	Studies reviewed for Recommendation 1: The ACCORD trial compared CVD outcomes in adults with DM and hypertension who were randomized to an SBP target of <140 mm Hg (standard therapy) or <120 mm Hg (intensive therapy), did not document a significant reduction in the primary outcome (CVD composite) with the lower BP goal, but the trial was underpowered to detect a statistically significant difference between the 2 treatment arms. It was found that more adverse events (2% increase in absolute risk) were identified in the lower BP group. The SPRINT trial demonstrated cardiovascular benefit from intensive treatment of BP to a goal of <120 mm Hg as compared with <140 mm Hg but did not include patients with DM. However, the results of ACCORD and SPRINT were generally consistent (26). Previous trials have shown similar quantitative benefits from lowering BP in persons with and without DM (9). The strongest effect was seen among patients with diabetes mellitus where greater BP lowering reduced the risk of stroke by 44% (RR: 0.56; 95% CI: 0.42–0.74). Of the studies that were included, subgroup results could not be pooled in a meaningful way. Among the subgroup analyses reported by individual studies, there were no significant findings. The ACC/AHA examined the reduction in risk of morbidity and mortality for 7 outcomes, comparing results from trials that randomly assigned individuals to lower targets versus standard targets for BP reduction. They found that greater BP lowering significantly reduced the risk of major cardiovascular events, MI, stroke, and heart failure. To determine whether an optimal target for BP reduction could be identified, the ACC/AHA also examined reduction in risk of these outcomes for RCTs with a systolic BP target <130 mm Hg in the lower BP target group and again found a reduced risk of stroke and major cardiovascular events with marginally significant reductions in risk of MI and all-cause mortality. Limiting the analyses to studies that included only participants with diabetes

	mellitus or CKD or with a mean participant age ≥60 years had little impact on the findings.
What harms were identified?	The ACC and AHA didn't provide an explicit discussion of the harms that were discussed in each study supporting the recommendations. However, there was some overall discussion of harms associated with treatment with antihypertensive medications, which largely consist of increased risk of adverse events such as, hypotension, syncope, electrolyte abnormalities, and acute kidney injury. Discussion also pointed to a potential risk of falls and serious fall injuries due to hypotension. Overall, the studies supporting these recommendations found that the benefits of blood pressure-lowering treatments outweigh the harms.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	There have been no new studies that contradict the current body of evidence.

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.

N/A

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

N/A

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

N/A

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.

This measure aims to improve the quality of care for patients with diabetes by assessing whether their blood pressure was adequately controlled (<140/90 mm Hg). High blood pressure is a major risk factor for microvascular and macrovascular complications in patients with diabetes. Cardiovascular disease is one of the leading causes for early death in people with diabetes. Uncontrolled high blood pressure contributes to the risk of complications and early death due to heart attack, stroke, angina and coronary heart disease. The benefits of quality envisioned by this measure include controlled blood pressure in patients with diabetes and a reduction in complications and early death.

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.*

The following data are extracted from HEDIS data collection and reflect the most recent years of measurement for this measure. Performance data is summarized at the health plan level and summarized by the mean, standard deviation, minimum health plan performance, maximum health plan performance, performance percentiles (10th, 25th, 50th, 75th, and 90th percentile) and the interquartile range. Data is stratified by year and product line (i.e. commercial, Medicare, Medicaid) at the health plan level.

The following data demonstrate the variation in the rate of patients with diabetes that had a blood pressure level <140/90 mm Hg.

Comprehensive Diabetes Care: Blood Pressure Control (<140/90 mm Hg)

N = Number of Health Plans

YEAR = Measurement Year

Commercial

YEAR|N|MEAN|ST DEV|MIN|10th|25th|50th|75th|90th|MAX|Interquartile Range

2016|395|56%|21%|0%|27%|51%|60%|69%|76%|86%|17%

2017|394|56%|22%|0%|4%|52%|62%|70%|76%|88%|18%

2018|394|59%|21%|0%|12%|56%|64%|72%|77%|89%|16%

Medicaid

YEAR|N|MEAN|ST DEV|MIN|10th|25th|50th|75th|90th|MAX|Interquartile Range

2016|266|60%|14%|0%|46%|53%|61%|69%|76%|85%|16%

2017|265|63%|12%|0%|50%|56%|64%|71%|78%|88%|14%

2018|250|62%|14%|0%|47%|57%|64%|71%|77%|85%|15%

Medicare

YEAR|N|MEAN|ST DEV|MIN|10th|25th|50th|75th|90th|MAX|Interquartile Range

2016|473|63%|17%|0%|44%|58%|66%|74%|80%|98%|16%

2017|473|67%|14%|0%|53%|61%|69%|75%|80%|94%|14%

2018|477|69%|11%|0%|57%|64%|70%|76%|81%|91%|12%

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

N/A

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

The CMS Office of Minority Health in collaboration with the RAND Corporation produces an annual report: CMS Racial, Ethnic, and Gender Disparities in Health Care in Medicare Advantage. We provide below summary data for this measure from that report. The authors note that "for reporting HEDIS data stratified by race and ethnicity, racial and ethnic group membership is estimated using a methodology that combines information from CMS administrative data, surname, and residential location."

The report described racial and ethnic disparities among beneficiaries 18-75 years old with diabetes who had their blood pressure under control. Hispanic and Asian or Pacific Islander beneficiaries outperformed White beneficiaries by more than 3 percentage points. Asian or Pacific Islanders performed best at 79.4%, with Hispanics following behind at 74.8% while only 64.7% of Whites were likely to have their blood pressure controlled. Overall, only 58.1% of Black beneficiaries with diabetes were likely to have their blood pressure controlled, with more than a 3% difference in performance compared to Whites.

2019 CMS Racial, Ethnic, and Gender Disparities in Health Care in Medicare Advantage report. https://www.cms.gov/About-CMS/Agency-Information/OMH/Downloads/2019-National-Level-Results-by-Race-Ethnicity-and-Gender.pdf

HEDIS data are stratified by type of insurance (e.g. commercial, Medicaid, Medicare). NCQA does not currently collect performance data stratified by race, ethnicity, or language. Escarce et al. have described in detail the difficulty of collecting valid data on race, ethnicity, and language at the health plan level (Escarce, 2011). While not specified in the measure, this measure can also be stratified by demographic variables, such as race/ethnicity or socioeconomic status, in order to assess the presence of health care disparities. The HEDIS Health Plan Measure Set contains two measures that can assist with stratification to assess health care disparities. The Race/Ethnicity Diversity of Membership and the Language Diversity of Membership measures were designed to promote standardized methods for collecting these data and follow Office of Management and Budget and Institute of Medicine guidelines for collecting and categorizing race/ethnicity and language data. In addition, NCQA's Multicultural Health Care Distinction Program outlines standards for collecting, storing and using race/ethnicity and language data to assess health care disparities.

Escarce, J.J., Carreon, R., Veselovskiy, G., Lawson, E.G. Collection of Race and Ethnicity Data by Health Plans has Grown Substantially, but Opportunities Remain to Expand Efforts. Health Affairs (Millwood) 2011; 30(10):1984-91. http://www.ncbi.nlm.nih.gov/pubmed/21976343

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

The 2017 National Diabetes Statistics report shows that minority groups are affected by diabetes at higher rates than Whites. Between 2013 and 2015, American Indians/Alaska Natives had the highest rates of diagnosed diabetes (15.1%). Non-Hispanic Blacks and Hispanics had the second and third highest rates at 12.7% and 12.1%, respectively. Eight percent of Asian Americans and 7.4% of Non-Hispanic Whites had diagnosed diabetes (CDC, 2017). The Office of Minority Health also reports on disparities in diabetic complications. Both African Americans and American Indian/Alaska Natives with diabetes are twice as likely to die from diabetes than non-Hispanic whites. African Americans also have higher rates of diabetic complications in comparison to non-Hispanic whites. African American men are 2.7 times as likely to undergo renal replacement therapy related to diabetes when compared to non-Hispanic white men. In addition, African Americans are 1.7 times as likely to be hospitalized for diabetes than Whites (OMH, 2014).

The Centers for Disease Control and Prevention reports similar disparities in cardiovascular disease among patients with diabetes. Between 1998-2006, African Americans with diabetes were discharged from the hospital for major cardiovascular disease at rates 1.5 times higher than Whites with diabetes (CDC, 2011). In 2006, the discharge rates for stroke in patients with diabetes were twice as high in African Americans than Whites (CDC, 2011). The percentage of patients with diabetes and blood pressure levels <140/90 mm Hg is also lower in minority groups. Seventy-seven percent of Non-Hispanic Whites had a blood pressure <140/90 mm Hg between 2003 and 2006 (CDC, 2013). In comparison 71.6% of Non-Hispanic Blacks and 66.8% of Mexican Americans had blood pressure levels <140/90 mm Hg between 2003 and 2006 (CDC, 2013).

Centers for Disease Control and Prevention. National Diabetes Statistics Report: Estimates of Diabetes and Its Burden in the United States, 2017. Atlanta, GA: U.S. Department of Health and Human Services; 2017.

Centers for Disease Control and Prevention, National Center for Health Statistics, data from the National Health and Nutrition Examination Survey. Statistical analysis by the Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Division of Diabetes Translation. 2013. Accessed from http://www.cdc.gov/diabetes/statistics/bp/bp_pct2byraceth.htm

Centers for Disease Control and Prevention. Diabetes Public Health Resource. Age-Adjusted Hospital Discharge Rates for Major Cardiovascular Disease as First-Listed Diagnosis per 1,000 Diabetic Population, by Race, United States, 1988-2006. 2011. Accessed from http://www.cdc.gov/diabetes/statistics/cvdhosp/cvd/fig6.htm

Centers for Disease Control and Prevention. Diabetes Public Health Resource. Age-Adjusted Hospital Discharge Rates for Stroke as First-Listed Diagnosis per 1,000 Diabetic Population, by Race, United States, 1988-2006. 2011. Accessed from http://www.cdc.gov/diabetes/statistics/cvdhosp/cvd/fig6.htm

Office of Minority Health, U. S. Department of Health and Human Services. Diabetes and American Indians/Alaska Natives. 2014. Accessed from http://minorityhealth.hhs.gov/omh/browse.aspx?lvl=4&lvlid=33

Office of Minority Health, U. S. Department of Health and Human Services. Diabetes and African Americans. 2014. Accessed from http://minorityhealth.hhs.gov/omh/browse.aspx?lvl=4&lvlID=18

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):*

Cardiovascular : Hypertension, Endocrine : Diabetes

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

Primary Prevention

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

Populations at Risk

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

This is not an eMeasure **Attachment:**

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

Attachment **Attachment:** 0061_CDC_BP_Control_Value_Sets_Fall_2019-637088223907626862.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. <u>For maintenance of endorsement:</u> Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

Yes

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

There have been minor changes to the value sets and medication lists to reflect current practice.

NCQA added a hospice exclusion to most HEDIS measures in 2016. The focus of hospice care is not to cure illnesses of patients, but rather to improve comfort and quality of life for those with less than six months to live. Most HEDIS quality measures are focused on health screenings or treatments that are not clinically appropriate or beneficial for those who are at end of life. Many of these screenings and treatments would also be uncomfortable for hospice patients, add undue burden and have no impact on improving length or quality of life. Therefore, including individuals who are receiving hospice in our HEDIS quality measures is inappropriate.

In addition, NCQA added exclusion criteria for adults with advanced illness and frailty, as well as Medicare adults 65 years of age and older enrolled in an I-SNP or living long-term in institutional settings. We recognize that for individuals with limited life expectancy, advanced illness or more complex clinical situations, the focus of this measure may not be relevant or in line with the patient's goals of care. By implementing this set of exclusions, those providing care to the frail and advanced illness population can focus on care that's more appropriate for their conditions and health status. Attention can be more focused on quality measures that capture services and care processes that are most relevant for this population (e.g., improving care transitions, getting follow-up after acute care episodes, or avoiding preventable hospitalizations).

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.*

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

Patients whose most recent blood pressure level was <140/90 mm Hg during the measurement year.

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)*

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

There are two data sources and approaches used for collecting data and reporting the numerator for this measure: Administrative Claims and Medical Record Review.

ADMINISTRATIVE CLAIMS

Use codes (See code value sets located in question S.2b.) to identify the most recent blood pressure reading taken during an outpatient visit or a nonacute inpatient encounter or remote monitoring event during the measurement year.

The patient is numerator compliant if the blood pressure is <140/90 mm Hg. The patient is not compliant if the blood pressure is =140/90 mm Hg, if there is no blood pressure reading during the measurement year or if the reading is incomplete (e.g., the systolic or diastolic level is missing). If there are multiple blood pressure readings on the same date of service, use the lowest systolic and lowest diastolic blood pressure on that date as the representative blood pressure.

Organizations that use CPT Category II codes to identify numerator compliance for this indicator must search for all codes in the following value sets and use the most recent codes during the measurement year to determine numerator compliance for both systolic and diastolic levels.

VALUE SET / NUMERATOR COMPLIANCE

Systolic Less Than 140 Value Set / Systolic compliant

Systolic Greater Than or Equal to 140 Value Set / Systolic noncompliant

Diastolic Less Than 80 Value Set / Diastolic compliant

Diastolic 80-89 Value Set / Diastolic compliant

Diastolic Greater Than or Equal to 90 Value Set / Diastolic not compliant

See attached code value sets.

MEDICAL RECORD REVIEW

The most recent BP level (taken during the measurement year) is <140/90 mm Hg, as documented through administrative data or medical record review.

The organization should use the medical record from which it abstracts data for the other measures in the Comprehensive Diabetes Care set. If the organization does not abstract for other measures, it should use the medical record of the provider that manages the patient's diabetes. If that medical record does not contain a BP, the organization may use the medical record of another PCP or specialist from whom the patient receives care.

Identify the most recent blood pressure reading noted during the measurement year. Do not include blood pressure readings that meet the following criteria:

-Taken during an acute inpatient stay or an ED visit.

-Taken on the same day as a diagnostic test or diagnostic or therapeutic procedure that requires a change in diet or change in medication on or one day before the day of the test or procedure, with the exception of fasting blood tests.

-Reported by or taken by the patient.

Blood pressure readings from remote monitoring devices that are digitally stored and transmitted to the provider may be included. There must be documentation in the medical record that clearly states the reading was taken by an electronic device, and results were digitally stored and transmitted to the provider and interpreted by the provider.

Identify the lowest systolic and lowest diastolic blood pressure reading from the most recent blood pressure notation in the medical record. If there are multiple blood pressure readings recorded for a single date, use

the lowest systolic and lowest diastolic blood pressure on that date as the representative blood pressure. The systolic and diastolic results do not need to be from the same reading when multiple readings are recorded for a single date.

The patient is not numerator compliant if the blood pressure does not meet the specified threshold or is missing, or if there is no blood pressure reading during the measurement year or if the reading is incomplete (i.e., the systolic or diastolic level is missing).

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

Patients 18-75 years of age by the end of the measurement year who had a diagnosis of diabetes (type 1 and type 2) during the measurement year or the year prior to the measurement year.

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).

There are two ways to identify patients with diabetes: by claim/encounter data and by pharmacy data. The organization must use both methods to identify the eligible population, but a patient only needs to be identified by one method to be included in the measure. Patients may be identified as having diabetes during the measurement year or the year prior to the measurement year.

CLAIM/ENCOUNTER DATA

Patients who met any of the following criteria during the measurement year of the year prior to the measurement year (count services that occur over both years):

- At least one acute inpatient encounter with a diagnosis of diabetes without telehealth.

- At least one acute inpatient discharge with a diagnosis of diabetes on the discharge claim. To identify an acute inpatient discharge:

1. Identify all acute and nonacute inpatient stays.

2. Exclude nonacute inpatient stays.

3. Identify the discharge date for the stay.

- At least two outpatient visits, observation visits, telephone visits, online assessments, ED visits, nonacute inpatient encounters or nonacute inpatient discharges, on different dates of service, with a diagnosis of diabetes. Visit type need not be the same for the two visits. To identify a nonacute inpatient discharge:

1. Identify all acute and nonacute inpatient stays.

2. Confirm the stay was for nonacute care based on the presence of a nonacute code on the claim.

3. Identify the discharge date for the stay.

-- Only include nonacute inpatient encounters without telehealth.

-- Only one of the two visits may be an outpatient telehealth visit, a telephone visit or an online assessment. Identify telehealth visits by the presence of a telehealth modifier or the presence of a telehealth POS code associated with the outpatient set.

See attached code value sets.

PHARMACY DATA

Patients who were dispensed insulin or hypoglycemics/antihyperglycemics on an ambulatory basis during the measurement year or the year prior to the measurement year.

PRESCRIPTIONS TO IDENTIFY MEMBERS WITH DIABETES

DESCRIPTION / PRESCRIPTION

Alpha-glucosidase inhibitors / Acarbose, Miglitol

Amylin analogs / Pramlintide

Antidiabetic combinations / Alogliptin-metformin, Alogliptin-pioglitazone, Canagliflozin-metformin, Dapagliflozin-metformin, Empagliflozin-linagliptin, Empagliflozin-metformin, Glimepiride-pioglitazone, Glipizide-metformin, Glyburide-metformin, Linagliptin-metformin, Metformin-pioglitazone, Metformin-repaglinide, Metformin-rosiglitazone, Metformin-saxagliptin, Metformin-sitagliptin

Insulin / Insulin aspart, Insulin aspart-insulin aspart protamine, Insulin degludec, Insulin detemir, Insulin glargine, Insulin glulisine, Insulin isophane human, Insulin isophane-insulin regular, Insulin lispro, Insulin lispro-insulin lispro protamine, Insulin regular human, Insulin human inhaled

Meglitinides / Nateglinide, Repaglinide

Glucagon-like peptide-1 (GLP1) agonists / Dulaglutide, Exenatide, Albiglutide, Liraglutide

Sodium glucose cotransporter 2 (SGLT2) inhibitor / Canagliflozin, Dapagliflozin, Empagliflozin

Sulfonylureas / Chlorpropamide, Glimepiride, Glipizide, Glyburide, Tolazamide, Tolbutamide

Thiazolidinediones / Pioglitazone, Rosiglitazone

Dipeptidyl peptidase-4 (DDP-4) inhibitors / Alogliptin, Linagliptin, Saxagliptin, Sitagliptin

Note: Glucophage/metformin as a solo agent is not included because it is used to treat conditions other than diabetes; members with diabetes on these medications are identified through diagnosis codes only.

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

This measure excludes adults in hospice. It also excludes adults with advanced illness and frailty, as well as Medicare adults 65 years of age and older enrolled in an I-SNP or living long-term in institutional settings.

Additionally, exclude patients who had a diagnosis of gestational diabetes or steroid-induced diabetes, in any setting, during the measurement year or the year prior to the measurement year and who did NOT have a diagnosis of diabetes. These patients are sometimes pulled into the denominator via pharmacy data. They are then removed once no additional diagnosis of diabetes (Type 1 or Type II) is found.

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.)*

ADMINISTRATIVE CLAIMS

Exclude patients who use hospice services or elect to use a hospice benefit any time during the measurement year, regardless of when the service began. These patients may be identified using various methods, which may include but are not limited to enrollment data, medical record or claims/encounter data.

Exclude adults who meet any of the following criteria:

- Medicare adults 66 years of age and older as of December 31 of the measurement year who meet either of the following:

-- Enrolled in an Institutional SNP (I-SNP) any time during the measurement year.

-- Living long-term in an institution any time during the measurement year as identified by the LTI flag in the Monthly Membership Detail Data File. Use the run data of the file to determine if a member had an LTI flag during the measurement year.

- Adults 66 years of age and older as of December 31 of the measurement year (all product lines) with frailty and with advanced illness. Patients must meet BOTH of the following frailty and advanced illness criteria to be excluded:

1. At least one claim/encounter for frailty during the measurement year.

2. Any of the following during the measurement year or the year prior to the measurement year (count services that occur over both years):

-- At least two outpatient visits, observation visits, ED visits, nonacute inpatient encounters nonacute inpatient discharges on different dates of services, with an advanced illness diagnosis. Visit type need not be the same for the two visits. To identify a nonacute inpatient discharge:

1. Identify all acute and nonacute inpatient stays.

2. Confirm the stay was for nonacute care based on the presence of a nonacute code on the claim.

3. Identify the discharge date for the stay.

-- At least one acute inpatient encounter with an advanced illness diagnosis.

-- At least one acute inpatient discharge with an advanced illness diagnosis. To identify an acute inpatient discharge:

1. Identify all acute and nonacute inpatient stays.

2. Exclude nonacute inpatient stays.

3. Identify the discharge date for the stay.

-- A dispensed dementia medication

DEMENTIA MEDICATIONS

DESCRIPTION / PRESCRIPTION

Cholinesterase inhibitors / Donepezil; Galantamine; Rivastigmine

Miscellaneous central nervous system agents / Memantine

Exclude patients with gestational diabetes or steroid diabetes. Codes associated with identifying these identifying exclusions are attached in a separate file with code value sets.

See attached code value sets.

MEDICAL RECORD

Exclusionary evidence in the medical record must include a note indicating the patient did NOT have a diagnosis of diabetes, in any setting, during the measurement year or the year prior to the measurement year AND had a diagnosis of gestational diabetes or steroid-induced diabetes, in any setting, during the measurement year or the year prior to the measurement year.

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

No stratification

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

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Higher score

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

STEP 1: Determine the eligible population. To do so, identify patients who meet all the specified criteria.

- AGES: 18-75 years as of December 31 of the measurement year.

- EVENT/DIAGNOSIS: Identify patients with diabetes in two ways: by claim/encounter data and by pharmacy data. SEE S.6 and S.7 for eligible population and denominator criteria and details.

STEP 2: Exclude patients who meet the exclusion criteria. SEE S.8 and S.9 for denominator exclusion criteria and details.

STEP 3: Determine the number of patients in the eligible population who had a blood pressure reading during the measurement year through the search of administrative data systems or medical record data.

STEP 4: Identify the lowest systolic and lowest diastolic blood pressure reading from the most recent blood pressure notation in the medical record.

STEP 5. Determine whether the result was <140/90 mm Hg.

STEP 6: Calculate the rate by dividing the numerator (STEP 5) by the denominator (after exclusions) (STEP 2).

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

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

Plans may report this measure using a systematic sample of 411 members. Plans are instructed to list and sort all eligible members for a measure. NCQA then provides plans with a Random Number Table that is released towards the end of the measurement year. The Random Number Table lists a value that is used to determine which members from the eligible population (i.e., every nth member) for whom numerator compliance will be determined.

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

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

N/A

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

If other, please describe in S.18.

Claims, Electronic Health Data, Electronic Health Records, Paper Medical 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.

This measure is based on administrative claims and medical record documentation collected in the course of providing care to health plan patients. NCQA collects the Healthcare Effectiveness Data and Information Set (HEDIS) data for this measure directly from health plans via NCQA's online data submission system.

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)*

Health Plan

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

BP_Control_Testing_Form_-61-_8.1.2019-637088230996654394.docx

2.1 For maintenance of endorsement

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

Yes

2.2 For maintenance of endorsement

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

Yes

2.3 For maintenance of endorsement

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

No - This measure is not risk-adjusted

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

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:** 0061
Measure Title: Comprehensive Diabetes Care: Blood Pressure Control <140/90 mm Hg
Date of Submission: 8/1/2019

Type of Measure:

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

☐ Process *(including Appropriate Use)*	☐ Efficiency
☐ Structure	

Instructions

- **Measures must be tested for all the data sources and levels of analyses that are specified. *If there is more than one set of data specifications or more than one level of analysis, contact NQF staff* about how to present all the testing information in one form.**
- **For all measures, sections 1, 2a2, 2b1, 2b2, and 2b4 must be completed.**
- **For outcome and resource use measures, section 2b3 also must be completed.**
- **If specified for multiple data sources/sets of specifications (e.g., claims and EHRs), section 2b5 also must be completed.**
- **Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b1-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.**
- **If you are unable to check a box, please highlight or shade the box for your response.**
- **Maximum of 25 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). *Contact NQF staff if more pages are needed.***
- **Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.**
- **For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for version 7.1 of the Measure Testing Attachment.**

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates 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. For instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.

2b1. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument-based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b3. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration**

OR

- **rationale/data support no risk adjustment/ stratification.**

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful [16] differences in performance;

OR

there is evidence of overall less-than-optimal performance.

2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b6. Analyses 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.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions.

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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

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

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data*

specified and intended for measure implementation. ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

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

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

N/A

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

Testing of performance measure score with beta binomial reliability and testing of construct validity with the Pearson Correlation were performed using HEDIS 2019 plan level data, measurement year 2018.

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

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

1.5. How many and which 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*)

This measure assesses whether adults enrolled in commercial, Medicare, and Medicaid plans who have diabetes (type 1 and type 2) had their recent blood pressure level less than 140/90 mm Hg. Therefore, testing was done at the health-plan level, which is appropriate for the level of reporting for this measure.

We calculated the measure score reliability and construct validity from HEDIS data that included 394 commercial health plans, 477 Medicare health plans, and 250 Medicaid health plans. The sample included all commercial, Medicare, and Medicaid health plans submitting data to NCQA for HEDIS. The plans were geographically diverse and varied in size.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)
Table 1 below provides a description of the data submitted for 2018, including the median denominator size per plan. Data are summarized at the health plan level and stratified by plan type (i.e. commercial, Medicaid, Medicare). Since data can be collected and reported from two data sources (administrative claims and medical record review), the vast majority of plans use a combination of data from administrative claims data and a sample of 411 of medical records they review to report their performance rates.

Table 1. Commercial, Medicaid, and Medicare plans reporting the Comprehensive Diabetes Care: Blood Pressure Control <140/90 mm Hg, 2018

Product Line	Number of Plans	Median Denominator Size/Plan
Commercial	394	411
Medicaid	250	411
Medicare	477	411

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:

Reliability of the health plan measure score was tested using a beta-binomial calculation. This analysis included the entire HEDIS data sample (described above).

Validity:

Validity of the health plan measure was demonstrated through construct validity using the entire HEDIS data sample (described above).

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.

We did not analyze social risk factors. This measure of health plan performance is specified to be reported separately by commercial, Medicaid and Medicare plan types, which serves as a proxy for income and other socioeconomic factors.

2a2. RELIABILITY TESTING

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

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

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Reliability was estimated by using the Beta-binomial model (Adams, 2009) for this health plan measure. Beta-binomial is appropriate for estimating the reliability of pass/fail rate measures. Reliability used here is the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance. A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to real differences in performance. The higher the reliability score, the greater is the confidence with which one can distinguish the performance of one plan from another. A reliability score greater than or equal to 0.7 is considered very good.

Adams, J.L. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California: RAND Corporation. TR-653-NCQA, 2009

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*)

Table 2 provides the reliability for the overall measure as shown by the Beta-binomial model as well as the distribution of individual plan reliability.

Table 2. Overall Beta-binomial statistic and distribution of plan reliability for commercial, Medicaid, and Medicare product lines, 2018

Product Line	Overall Reliability	Min	Percentiles					Max
			10th	25th	50th	75th	90th	
Commercial	0.998	0.918	0.990	0.991	0.992	0.994	0.999	1.000
Medicaid	0.987	0.607	0.901	0.948	0.952	0.958	0.965	1.000
Medicare	0.976	0.787	0.977	0.979	0.981	0.985	0.989	1.000

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?*)

The values for the beta-binomial statistic across all product lines for the health plan level measure are all greater than 0.7, indicating the measure has very good reliability. The 10-90th percentile distribution of health plan level-reliability on this measure show the vast majority of health plans not only exceeded the minimally accepted threshold of 0.7 but also exceeded 0.9. Strong reliability is demonstrated since the majority of variance is due to signal and not to noise.

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)*

We tested for construct validity of the Comprehensive Diabetes Care (CDC): Blood Pressure Control (<140/90 mm Hg) measure by exploring whether it was correlated with other similar measures of quality hypothesized which are listed below.

- CDC: Hemoglobin A1c (HbA1c) Testing: The percentage of adults 18-75 with diabetes that had an HbA1c test performed during the measurement year.
- CDC: HbA1c Control (<8.0%): The percentage of adults 18-75 with diabetes whose most recent HbA1c level is <8.0% during the measurement year.
- CDC: HbA1c Poor Control (> 9.0%): The percentage of adults 18-75 with diabetes whose most recent HbA1c level is >9% during the measurement year.
- CDC: Eye Exam (Retinal) Performed: The percentage of adults 18-75 with diabetes that had an eye screening for diabetic retinal disease during the measurement year.

These measures were chosen for construct validity because they are similarly focused on a population with diabetes (type 1 and type 2) and focus on evidenced-based monitoring and treatment for patients with diabetes. We hypothesized that a plan that does well on these measures for diabetes would also do well on this blood pressure control measure for patients who have diabetes.

To test these correlations, we used a Pearson correlation test. This test estimates the strength of the linear association between two continuous variables; the magnitude of correlation ranges from -1 to +1. A value of 1 indicates a perfect linear dependence in which increasing values on one variable is associated with increasing values of the second variable. A value of 0 indicates no linear association. A value of -1 indicates a perfect linear relationship in which increasing values of the first variable is associated with decreasing values of the second variable. Coefficients with absolute value of less than 0.3 are generally considered indicative of weak associations whereas absolute values of 0.3 or higher denote moderate to strong associations. The significance of a correlation coefficient is evaluated by testing the hypothesis that an observed coefficient calculated for the sample is different from zero. The resulting p-value indicates the probability of obtaining a difference at least as large as the one observed due to chance alone. We used a threshold of 0.05 to evaluate the test results. P-values less than this threshold imply that it is unlikely that a non-zero coefficient was observed due to chance alone.

** Note: All HEDIS value sets are updated annually with the most current codes available. The information below details the process we used to convert value sets that used ICD-9 codes to ICD-10 codes in 2015. **

ICD-10 CONVERSION:

In preparation for the national implementation of ICD-10 in 2015, NCQA conducted a systematic mapping of all value sets maintained by the organization to ensure the new values used for reporting maintained the reliability, validity, and intent of the original specification.

Steps in ICD-9 to ICD-10 Conversion Process

1. NCQA first identified value sets within the measure that included ICD-9 codes. We used General Equivalence Mapping (GEM) to identify ICD-10 codes that map to ICD-9 codes and reviewed GEM mapping in both directions (ICD-9 to ICD-10 and ICD-10 to ICD-9) to identify potential trending issues.
2. NCQA then searched for additional codes (not identified by GEM mapping step) that should be considered due to the expansion of concepts in ICD-10. Using ICD-10 tabular list and ICD-10 Index, searches by diagnosis or procedure name were conducted to identify appropriate codes.
3. NCQA HEDIS Expert Coding Panel review: Updated value set recommendations were presented for expert review and feedback.
4. NCQA RMAP clinical review: Due to increase specificity in ICD-10, new codes and definitions require review to confirm the diagnosis or procedure is consistent and appropriate given the scope of the measure.
5. New value sets containing ICD-10 code recommendations were posted for public review and comment in 2014 and updated in 2015. Comments received were reconciled with additional feedback from HEDIS Expert Coding Panel and MAPs as needed.
6. NCQA staff finalized value sets containing ICD-10 codes for publication in 2015.

Tools Used to Identify/Map to ICD-10
All tools used for mapping/code identification from CMS ICD-10 website (http://www.cms.gov/Medicare/Coding/ICD10/2012-ICD-10-CM-and-GEMs.html).
GEM, ICD-10 Guidelines, ICD-10-CM Tabular List of Diseases and Injuries, ICD-10-PCS Tabular List.

Expert Participation
The NCQA HEDIS Expert Coding Panel reviewed and provided feedback on staff recommendations. Names and credentials of the experts who served on these panels are listed under Additional Information, Ad. 1. Workgroup/Expert Panel Involved in Measure Development.

2b1.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)
The results from construct validity testing of the health plan level measure are presented by product line in Tables 3a, 3b, and 3c below.

Table 3a. Correlations among Diabetes Measures in Commercial Health Plans, 2018.

	Pearson Correlation Coefficients			
	HbA1c Testing	HbA1c Control (<8.0%)	HbA1c Poor Control (>9.0%)	Eye Exams
Blood Pressure Control	0.4094	0.8882	-0.8986	0.4755

Note: All correlations are significant at p<0.0001

Table 3b. Correlations among Diabetes Measures in Medicaid Health Plans, 2018.

	Pearson Correlation Coefficients			
	HbA1c Testing	HbA1c Control (<8.0%)	HbA1c Poor Control (>9.0%)	Eye Exams
Blood Pressure Control	0.5867	0.7562	-0.7820	0.5413

Note: All correlations are significant at p<0.0001

Table 3c. Correlations among Diabetes Measures in Medicare Health Plans, 2018.

	Pearson Correlation Coefficients			
	HbA1c Testing	HbA1c Control (<8.0%)	HbA1c Poor Control (>9.0%)	Eye Exams
Blood Pressure Control	0.4334	0.5832	-0.5914	0.5545

Note: All correlations are significant at p<0.0001

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?*)
Across all product lines, the correlations are moderate to very strong and statistically significant. These results confirm the hypothesis that plan performance on these diabetes measures are correlated with each other. Coefficients with absolute value of less than .3 are generally considered indicative of weak associations. Absolute values of .3 to .59 are considered moderate associations, absolute values of .6 to .69 indicate a strong positive relationship, and absolute values of .7 or higher indicate a very strong positive relationship. These correlation results suggest that this plan level measure has sufficient validity.

Note: Correlation values with the HbA1c Poor Control measure are all negative because it is a "lower is better quality" measure, while the other measures are all "higher is better". All other measures show that plans that higher rates on one measure will have high rates on the other.

2b2. EXCLUSIONS ANALYSIS

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

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

We did not perform testing of the following exclusions for this submission:

- Gestational diabetes
- Steroid-induced diabetes

NCQA engaged expert panels to inform the face validity of these exclusions for this measure, which aligns with evidence focused on the general population of people with Type I or Type II diabetes. This measure has been reviewed by NCQA's Diabetes Measurement Advisory Panel, Cardiovascular Measurement Advisory Panel, Technical Measurement Advisory Panel, and the Committee on Performance Measurement. The measure also received public comment feedback upon initial development.

Hospice, I-SNPs and Long-Term Care Institutions

These exclusions were also not formally tested for this submission. This measure is designed to be scientifically valid and feasible for comparing the quality of care provided to general populations, such as healthy older adults or those with a single condition. Patients receiving hospice, enrolled in an I-SNP, or residing in a long-term care institution would likely have different care needs and quality concerns, therefore they are excluded from this measure.

Advanced Illness and Frailty

For HEDIS 2019 (measurement year 2018), NCQA added exclusions for advanced illness and frailty to the Comprehensive Diabetes Care: Blood Pressure Control (<140/90 mm Hg) measure. NCQA decided to explore

implementing these exclusions, recognizing that for individuals with limited life expectancy, advanced illness and frailty, the focus of this measure may not be clinically appropriate, relevant or in line with the patient's goals of care. We performed a review of literature on different approaches to defining advanced illness and used this, along with feedback received from expert work groups, measurement advisory panels and public comment to create a list of illnesses, conditions and service codes to be included in testing. The conditions included: dementia and other neurodegenerative conditions, emphysema, end stage renal disease (ESRD), heart failure, liver failure, metastatic cancer, pulmonary fibrosis and respiratory failure.

NCQA then conducted a search of ICD-10 codes that were relevant to each of the conditions to create value sets for testing. To identify those with dementia, NCQA also included drug codes for medications such as donepezil hydrochloride and galantamine hydrobromide, to capture those who may not carry a diagnosis of dementia but are prescribed a drug for treatment.

The proxy for frailty was developed based on previously studied approaches[1, 2, 3] and feedback received from expert work groups and measurement advisory panels. The proxy is comprised of HCPCS and ICD-10 codes for diagnoses or services that can indicate when an individual is frail or dependent in activities of daily living. Examples include: gait abnormality, abnormal loss of weight and underweight, adult failure to thrive, debility, fall, pressure ulcer, durable medical equipment (hospital bed, walker, portable or home oxygen, wheelchair), bed confinement, palliative care and age-related physical debility. Members met the frailty proxy criteria if they had a claim for any of the codes included in the frailty code set in the measurement year.

To determine the feasibility and impact of applying this exclusion to the measure, NCQA used a research database that consisted of two years of inpatient, outpatient, and pharmacy claims for members age 18 and older enrolled in a sample of Medicare Advantage plans (N=25). NCQA compared several approaches for identifying the advanced illness and frailty populations, examining different age ranges and diagnosis positions and their impact on the denominator. The results of those queries along with input from the expert work groups, measurement advisory panels and public comment led us to determine that the best approach for identifying the advanced illness and frailty population that should be excluded from the measure was to apply the following criteria:

- Adults 66 and older as of December 31 of the measurement year (all product lines) with frailty and advanced illness

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*)

Table 4 shows the results of applying the exclusion of adults 66 and older with advanced illness and frailty to the Comprehensive Diabetes Care: Blood Pressure Control (<140/90 mm Hg) measure.

---

[1] Faurot, K.R., Funk, M.J., Pate, V., Brookhart, M.A., Patrick, A., Hanson, L.C., Castillo, W.C., Stürmer, T. 2015. Using Claims Data to Predict Dependency in Activities of Daily Living as a Proxy for Frailty. Pharmacoepidemiology and Drug Safety. 24(1): 59-66.

[2] Segal, J.B., Chang, H.Y., Du, Y., Walston, J.D., Carlson, M.C., Varadhan, R. 2017. Development of a

Claims-Based Frailty Indicator Anchored to a Well-Established Frailty Phenotype. Medical Care. 55(7): 716-722.

[3] Davidoff A.J., A. Hurrida, I.H. Zuckerman, S.M. Lichtman, N. Pandya, A. Hussain, F. Hendrick, J.P. Weiner, X. Ke, M.J. Edelman. 2013. A Novel Approach to Improve Health Status Measurement in Observational Claims-Based Studies of Cancer Treatment and Outcomes. J Geriatr Oncol. 4(2):157–165.

Table 4. Impact of applying the advanced illness and frailty for patients aged 66 and older

Number of Plans (N)	Average Number Excluded	Average % Removed by Exclusion
25	350	2.0

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*)

Advanced Illness and Frailty

The advanced illness and frailty exclusion had a small impact on the eligible population: 2.0% on average were removed for advance illness and frailty. Feedback from NCQA's expert work groups and measurement advisory panels, as well as public comment feedback, supported the application of this exclusion to the Comprehensive Diabetes Care: Blood Pressure Control (<140/90 mm Hg) measure for clinical reasons. By implementing this exclusion, those providing care to patients with advanced illness and frailty can focus on care that is more appropriate for their conditions and health status. Attention can be more focused on quality measures that capture services and care processes that are most relevant for this population (e.g., improving care transitions, getting follow-up after acute care episodes, or avoiding preventable hospitalizations).

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

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

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

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with _risk factors**

☐ **Stratification by _risk categories**

☐ **Other,**

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

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.

NCQA recognizes that there is a growing body of literature that might support risk adjustment or stratification of intermediate outcome measures. However, at this time, NCQA does not currently risk adjust this measure given the potential to mask poor performance and disparities in care. This measure of health plan performance is specified to be reported separately by commercial, Medicaid and Medicare plan types, which serves as a proxy for income and other socioeconomic factors.

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)**

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)*

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

To determine if this difference is statistically significant, NCQA calculates an independent sample t-test of the performance difference between two randomly selected plans at the 25th and 75th percentile. The t-test method calculates a testing statistic based on the sample size, performance rate, and standardized error of each plan. The test statistic is then compared against a normal distribution. If the p value of the test statistic is less than 0.05, then the two plans' performance is significantly different from each other.

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*)

Table 5. Variation in Performance for commercial, Medicaid, and Medicare health plans, 2018.

Plan Type	N	Average (%)	St Dev (%)	10th (%)	25th (%)	50th (%)	75th (%)	90th (%)	IQR (%)	p-value
Commercial	394	58.57	21.36	11.96	56.20	64.23	71.72	76.61	15.52	<0.0001
Medicaid	250	62.15	14.11	47.20	56.73	63.90	71.30	77.02	15.57	<0.0001
Medicare	477	68.98	11.06	57.30	64.48	69.59	76.07	81.14	11.59	<0.0001

N = Number of plans reporting

IQR = Interquartile range

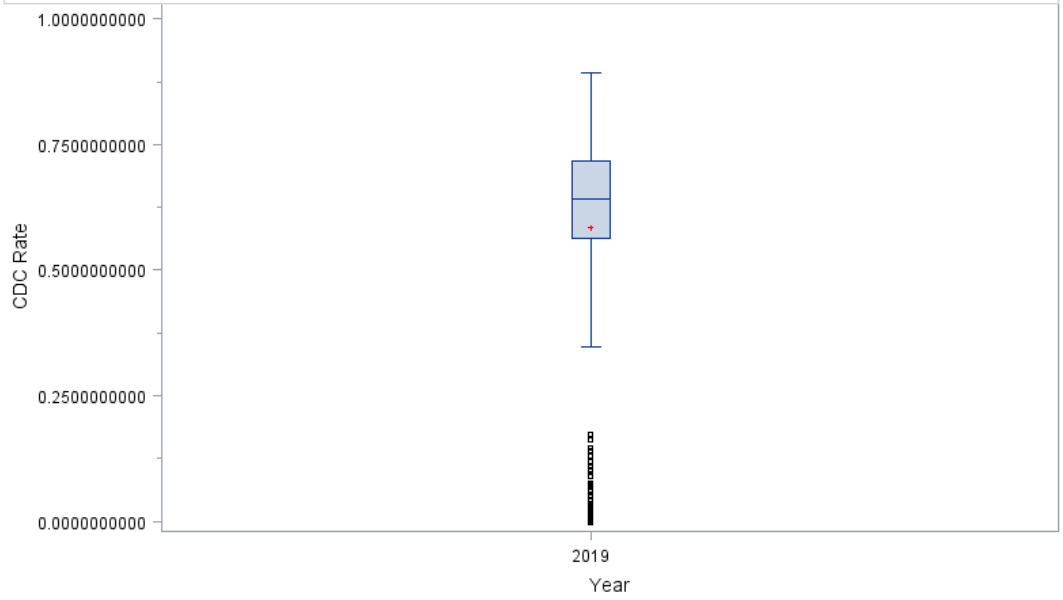
p-value = p-value of independent samples t-test comparing plans at the 25th percentile to plans at the 75th percentile.

Box plots for HEDIS 2019 (Measurement year 2018) Variation in Performance Across Health Plans are included below for your reference.

Boxplot Graph CDC Commercial from HEDIS 2019

IndicatorName: Comprehensive Diabetes Care - Blood Pressure Control (<140/90)'

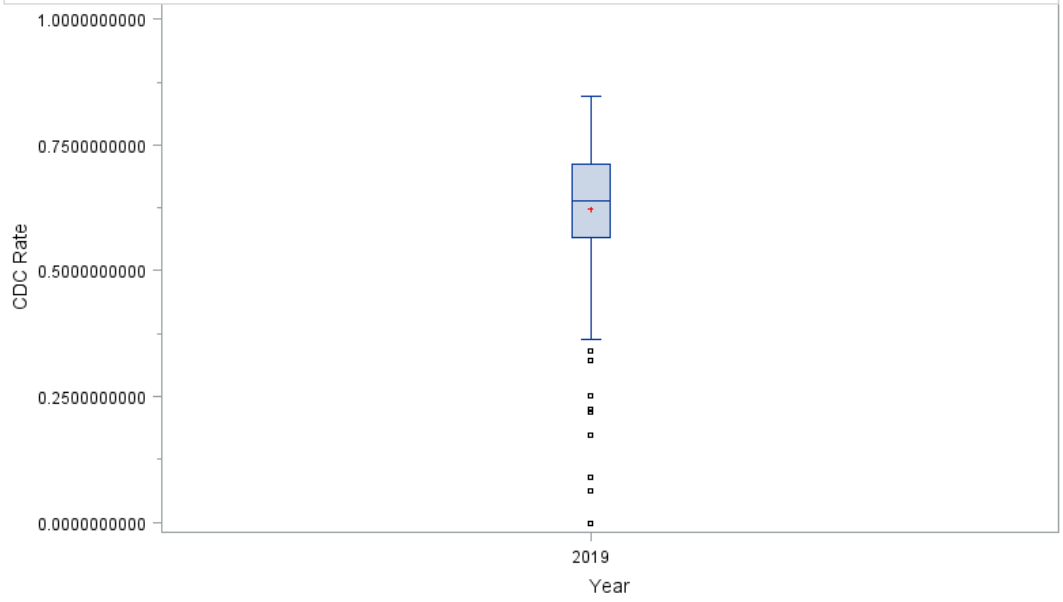
Summary Statistics	
No.of Plans	394
Average	0.585726
Lowest Rate	0
Highest Rate	0.891847

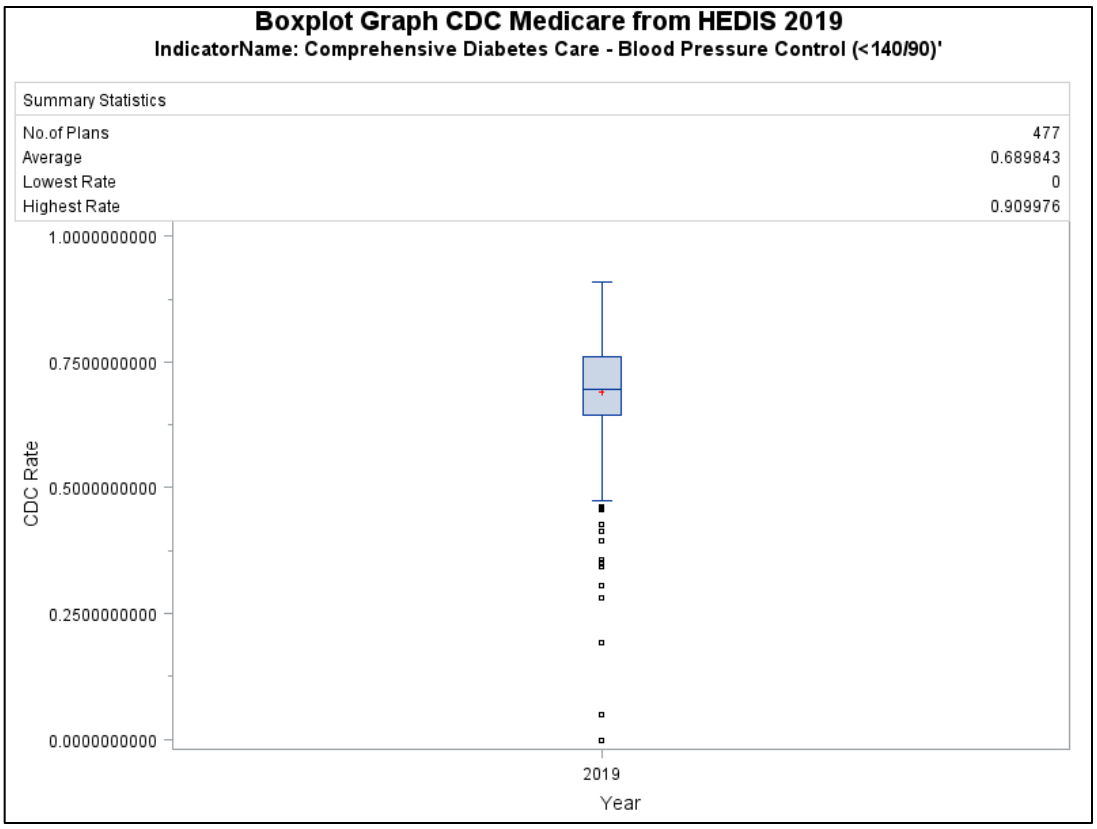


Boxplot Graph CDC Medicaid from HEDIS 2019

IndicatorName: Comprehensive Diabetes Care - Blood Pressure Control (<140/90)'

Summary Statistics	
No.of Plans	250
Average	0.621506
Lowest Rate	0
Highest Rate	0.847666





2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)
The results above indicate there is meaningful difference in performance. Across all product lines, the difference between the 25th and 75th percentile (better performance) is statistically significant.

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.***

The Comprehensive Diabetes Care: Blood Pressure Control (<140/90 mm Hg) measure has only one set of specifications.

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*)
HEDIS measures apply to enrolled members in a health plan, and NCQA has a rigorous audit process to ensure the eligible population and numerator events for each measure are correctly identified and reported. The audit process is designed to verify primary data sources used to populate measures and ensure specifications are correctly implemented.

The HEDIS Compliance Audit addresses the following functions:
- Information practices and control procedures
- Sampling methods and procedures
- Data integrity
- Compliance with HEDIS specifications
- Analytic file production
- Reporting and documentation

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*)
HEDIS addresses missing data in a structured way through its audit process. HEDIS measures apply to enrolled members in a health plan, and NCQA-certified auditors use standard audit methodologies to assess whether data sources are missing data. If a data source is found to be missing data, and the issues cannot be rectified, the auditor will assign a "materially biased" designation to the measure for that reporting plan, and the rate will not be used. Once measures are added to HEDIS, NCQA conducts a first-year analysis to assess the measure's feasibility once widely implemented in the field. This analysis includes an assessment of how many plans report valid rates vs. rates that are materially biased (or have other issues, such as small denominators). These considerations are weighed in the deliberation process before measures are approved for public reporting.

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

The denominator of this measure is identified using claims data and not subject to difference between response or nonresponse. This measure goes through the NCQA audit process each year to identify potential errors or bias in results. Only performances rates that have been reviewed and determined not to be "materially biased" are reported and used.

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

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

If other:

3b. Electronic Sources

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

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

Some data elements are in defined fields in electronic sources

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

To allow for widespread reporting across health plans and health care practices, this measure is collected through multiple data sources (administrative data, electronic clinical data, and paper records). We anticipate as electronic health records become more widespread, the reliance on paper record review will decrease.

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

Attachment:

3c. Data Collection Strategy

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

3c.1. Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and

frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

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

NCQA conducts an independent audit of all HEDIS collection and reporting processes, as well as an audit of the data which are manipulated by those processes, in order to verify that HEDIS specifications are met. NCQA has developed a precise, standardized methodology for verifying the integrity of HEDIS collection and calculation processes through a two-part program consisting of an overall information systems capabilities assessment followed by an evaluation of the MCO's ability to comply with HEDIS specifications. NCQA-certified auditors using standard audit methodologies will help enable purchasers to make more reliable "apples-to-apples" comparisons between health plans.

The HEDIS Compliance Audit addresses the following functions:

1) Information practices and control procedures

2) Sampling methods and procedures

3) Data integrity

4) Compliance with HEDIS specifications

5) Analytic file production

6) Reporting and documentation

In addition to the HEDIS audit, NCQA provides a system to allow "real-time" feedback from measure users. Our Policy Clarification Support System receives thousands of inquiries each year on over 100 measures. Through this system, NCQA responds immediately to questions and identifies possible errors or inconsistencies in the implementation of the measure. This system informs both annual updates to the measures as well as routine re-evaluation of measures. These processes include updating value sets and clarifying the specifications. Measures are re-evaluated on a periodic basis and when there is a significant change in evidence.

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

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

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
	Public Reporting Health Plan Ratings/Report Cards https://www.ncqa.org/hedis/reports-and-research/ratings-2019/ https://www.ncqa.org/report-cards/health-plans/state-of-health-care-quality-report/ Annual State of Health Care Quality Health Plan Ratings/Report Cards https://reportcards.ncqa.org/#/health-plans/list Health Plan Ratings/Report Cards https://www.ncqa.org/hedis/reports-and-research/ratings-2019/ https://www.ncqa.org/report-cards/health-plans/state-of-health-care-quality-report/ Annual State of Health Care Quality Health Plan Ratings/Report Cards https://reportcards.ncqa.org/#/health-plans/list Payment Program IHA California Pay for Performance https://www.iha.org/our-work/accountability/value-based-p4p/measure-set Regulatory and Accreditation Programs NCQA Accreditation https://www.ncqa.org/programs/health-plans/health-plan-accreditation-hpa/ NCQA Accreditation https://www.ncqa.org/programs/health-plans/health-plan-accreditation-hpa/ Professional Certification or Recognition Program NCQA Diabetes Recognition Program http://www.ncqa.org/Programs/Recognition/Clinicians/DiabetesRecognitionProgramDRP.aspx Quality Improvement (external benchmarking to organizations) Quality Compass http://www.ncqa.org/hedis-quality-measurement/quality-measurement-products/quality-compass https://www.ncqa.org/report-cards/health-plans/state-of-health-care-quality-report/ Annual State of Health Care Quality

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

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

DIABETES RECOGNITION PROGRAM: This measure is used in NCQA's Diabetes Recognition Program (DRP), that assesses clinician performance on key quality measures that are based on national evidence-based guidelines in diabetes care. The DRP Program has 6 measures which cover areas such as: HbA1c control, blood pressure control, eye examinations, nephropathy assessment, foot examination, and smoking and tobacco use cessation

advice or treatment. Eligible clinicians will abstract data from the charts of diabetes patients (25 patients for a single applicant) and submit this information to NCQA for review.
HEALTH PLAN ACCREDITATION: This measure is used in scoring for accreditation of Medicare Advantage Heath Plans. As of Fall 2018, a total of 184 Medicare Advantage health plans were accredited using this measure among others covering 9.2 million Medicare beneficiaries; 451 commercial health plans covering 113 million lives; and 125 Medicaid health plans covering 35 million lives. Health plans are scored based on performance compared to benchmarks.
HEALTH PLAN RATING/REPORT CARDS: This measure is used to calculate health plan rankings which are reported on the NCQA website. These rankings are based on performance on HEDIS measures among other factors. In 2019, a total of 255 Medicare health plans, 515 commercial health plans and 188 Medicaid health plans across 50 states were included in the rankings.
INTEGRATED HEALTHCARE ASSOCIATION (IHA) CALIFORNIA PAY FOR PERFORMANCE: This measure is used in the California P4P program which is the largest non-governmental physician incentive program in the United States. Founded in 2001, it is managed by the Integrated Healthcare Association (IHA) on behalf of eight health plans representing 10 million insured persons. IHA is responsible for collecting data, deploying a common measure set, and reporting results for approximately 35,000 physicians in nearly 200 physician groups. This program represents the longest running U.S. example of data aggregation and standardized results reporting across diverse regions and multiple health plans. California consumers benefit from the availability of standardized performance results from a common measure set, which are available to the public through the State of California, Office of the Patient Advocate.
QUALITY COMPASS: This measure is used in Quality Compass which is an indispensable tool used for selecting a health plan, conducting competitor analysis, examining quality improvement and benchmarking plan performance. Provided in this tool is the ability to generate custom reports by selecting plans, measures, and benchmarks (averages and percentiles) for up to three trended years. Results in table and graph formats offer simple comparison of plans' performance against competitors or benchmarks.
STATE OF HEALTH CARE ANNUAL REPORT: This measure is publicly reported nationally and by geographic regions in the NCQA State of Health Care annual report. This annual report published by NCQA summarizes findings on quality of care. In 2019, the report included results from calendar year 2018 for health plans covering a record 136 million people, or 43 percent of the U.S. population.

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

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

N/A

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

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

Health plans that report HEDIS calculate their rates and know their performance when submitting to NCQA. NCQA publicly reports rates across all plans and also creates benchmarks in order to help plans understand how they perform relative to other plans. Public reporting and benchmarking are effective quality improvement methods.

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.

NCQA publishes HEDIS results annually in our Quality Compass tool. NCQA also presents data at various conferences and webinars. For example, at the annual HEDIS Update and Best Practices Conference (now the Health Care Quality Congress), NCQA presents results from all new measures' first year of implementation or analyses from measures that have changed significantly and insight into new measure development projects. NCQA also regularly provides technical assistance on measures through its Policy Clarification Support System, as described in Section 3c.1.

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

Describe how feedback was obtained.

NCQA measures are evaluated regularly using a consensus-based process to consider input from multiple stakeholders, including but not limited to entities being measured. We use several methods to obtain input, including vetting of the measure with several multi-stakeholder advisory panels, public comment posting, and review of questions submitted to the Policy Clarification Support System. This information enables NCQA to comprehensively assess a measure's adherence to the HEDIS Desirable Attributes of Relevance, Scientific Soundness and Feasibility.

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

Questions received through the Policy Clarification Support System have generally centered around clarification on the use of blood pressure readings obtained during potentially stressful procedures or specific visit types, suggestions for exclusions, and the use of patient-reported readings.

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

This measure has been deemed a priority measure by NCQA, as illustrated by its use in programs such as Health Plan Rating, NCQA Accreditation and Quality Compass. States, employers and regional health quality organizations value this measure (and other HEDIS measures) for shining a light on quality.

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.

We have provided minor clarifications about the measure during the annual update process in order to address questions received through the Policy Clarification Support System.

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.

From 2016-2018, performance on this measure has generally improved (2-5%) across the commercial, Medicare, and Medicaid product lines. Current average performance (MY 2018) is highest in Medicare plans (69%), followed by Medicaid plans (62%), and then commercial plans (59%).

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.

There were no identified unexpected findings during testing or since implementation of this measure.

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

There were no identified unexpected findings during testing or since implementation of this measure.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

Yes

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

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

NQF 0729 Optimal Diabetes Care (Minnesota Community Measurement) was not listed above but is NQF endorsed.

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

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

Yes

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

Measure 0061 is NQF endorsed as a single measure that uses health plan reported data to assess the percentage of patients 18-75 years of age with diabetes (type 1 and type 2) whose most recent blood pressure level is <140/90 mm Hg. Measure 0729 is a composite measure (all or nothing) that uses physician reported data to assess the percentage of adult diabetes patients who have optimally managed modifiable risk factors including blood pressure and four other indicators. NCQA's measure 0061 is included with five other NCQA diabetes measures. The five other diabetes measures are individually NQF endorsed (Endocrine Maintenance Phase 1). Together, the six NCQA individual diabetes measures (including measure 0061) make a set of diabetes HEDIS measures but are not considered all or nothing. NCQA uses individual measures to provide health plans and others the opportunity to measure, report and incentivize each aspect of quality care for the diabetes population. HARMONIZED MEASURE ELEMENTS: Measures 0061 and 0729 both focus on an adult patient population 18-75 years of age with diabetes (type 1 and type 2). Both measures assess whether the patient's most recent blood pressure level in the measurement period was <140/90 mm Hg. Both measures also specify denominator visit criteria to include patients with at least two outpatient visits in the last two years with a diagnosis of diabetes. UNHARMONIZED MEASURE ELEMENTS: - Data Source: Measure 0061 is collected through administrative claims and/or medical record. Measure 0729 is collected through medical record abstraction. - Level of Accountability: Measure 0061 is a health plan level measure and is used in NCQA's

clinical quality and recognition programs (See 4.1 Usability and Use). Measure 0729 is a physician level measure. - Data Elements: Measure 0061 uses two methods to identify patients in the denominator 1) claims/encounter data with a diagnosis of diabetes and 2) pharmacy data for insulin or hypoglycemic/antihyperglycemics (see S.7 Denominator Details). Measure 0729 uses encounter data with a diagnosis for diabetes to identify patients in the denominator. NCQA uses two identification methods to ensure that only patients with diagnosed diabetes are included in the denominator. - Exclusions: Exclusions for measures 0061 and 0729 are substantially aligned with some variation due to differences in health plan and clinician level reporting. IMPACT ON INTERPRETABILITY AND DATA COLLECTION BURDEN: The differences between these measures do not have an impact on interpretability of publicly reported rates. There is no added burden of data collection because the data for each measure is collected from different data sources by different entities.

5b. Competing Measures

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

OR

Multiple measures are justified.

5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):

Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

N/A

Appendix

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

No appendix **Attachment:**

Contact Information

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

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

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

Co.4 Point of Contact: Kristen, Swift, Swift@ncqa.org, 202-955-5174-

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.

NCQA follows a standard process of vetting members of the measurement advisory panel for conflicts of interest.

DIABETES MEASUREMENT ADVISORY PANEL

David Aron, MD, MS, Department of Veteran's Affairs

Jerry Cavallerano, OD, PhD, Joslin Diabetes Center

Mark Cziraky, PharmD, FAHA, CLS, HealthCore, Inc.

Stephen Fadem, MD, Kidney Associates PPLC

Ted Ganiats, MD, University of California, San Diego

Richard Hellman, MD, FACP, FACE, Private Practice, Endocrinology

William Herman (Chair), MD, MPH, University of Michigan

Lynne Levitsky, MD, Partners Healthcare

Seth Rubenstein, DPM, American Podiatric Medical Association

John Thompson, MD, Private Practice, Ophthalmology

CARDIOVASCULAR MEASUREMENT ADVISORY PANEL

Kathy Berra, RN, MSN, ANP-BC, FAHA, FAAN, FPCNA, The LifeCare Company

Donald Casey, MD, MPH, MBA, FACP, FAHA, FAAPL, DFACMQ, American College of Medical Quality

Tom Kottke, MD, MSPH, HealthPartners

Eduardo Ortiz, MD, MPH, Tennessee Valley Healthcare System

Stephen Persell (Chair), MD, MPH, Northwestern University

Randall Stafford, MD, PhD, Stanford University

Kim Williams, MD, MACC, MASNC, FAHA, FESC, Rush University Medical Center

Tracy Wolff, MD, Agency for Healthcare Research and Quality

COMMITTEE ON PERFORMANCE MEASUREMENT

Andrew Baskin, MD, Aetna

Elizabeth Drye, MD, SM, Yale School of Medicine

Andrea Gelzer, MD, MS, FACP, AmeriHealth Caritas

Kate Goodrich, MD, MHS, Centers for Medicare & Medicaid Services

David Grossman, MD, MPH, Washington Permanente Medical Group

Christine Hunter, (Co-Chair), MD, WPS Health Solutions

David Kelley, MD, MPA, Pennsylvania Department of Human Services

Jeffrey Kelman, MMSc, MD, Department of Health and Human Services

Nancy Lane, PhD, Independent Consultant

Bernadette Loftus, MD, Freelance

Adrienne Mims, MD, MPH, AGSF, FAAFP, Alliant Health Solutions

Amanda Parsons, MD, MBA, Metroplus

Wayne Rawlins, MD, MBA, ConnectiCare

Misty Roberts, MSN, RN, CPHQ, PMP, Humana

Rudy Saenz, MD, MMM, FACOG, Riverside Medical Clinic

Marcus Thygeson, (Co-Chair), MD, MPH, Blind On-Demand

JoAnn Volk, MA, Georgetown University

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2009

Ad.3 Month and Year of most recent revision: 07, 2018

Ad.4 What is your frequency for review/update of this measure? Approximately every 3 years, sooner if the clinical guidelines have changed significantly.

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

Ad.6 Copyright statement: The HEDIS® measures and specifications were developed by and are owned by the National Committee for Quality Assurance (NCQA). The HEDIS measures and specifications are not clinical guidelines and do not establish a standard of medical care. NCQA makes no representations, warranties, or endorsement about the quality of any organization or physician that uses or reports performance measures and NCQA has no liability to anyone who relies on such measures or specifications. NCQA holds a copyright in these materials and can rescind or alter these materials at any time. These materials may not be modified by anyone other than NCQA. Anyone desiring to use or reproduce the materials without modification for a non-commercial purpose may do so without obtaining any approval from NCQA. All commercial uses must be approved by NCQA and are subject to a license at the discretion of NCQA.

Calculated measure results, based on unadjusted HEDIS specifications, may not be termed “Health Plan HEDIS rates” until they are audited and designated reportable by an NCQA-Certified Auditor. Such unaudited results should be referred to as “Unaudited Health Plan HEDIS Rates.” Accordingly, “Heath Plan HEDIS rate” refers to and assumes a result from an unadjusted HEDIS specification that has been audited by an NCQA-Certified HEDIS Auditor.

Limited proprietary coding is contained in the measure specifications for convenience. Users of the proprietary code sets should obtain all necessary licenses from the owners of these code sets. NCQA disclaims all liability for use or accuracy of any coding contained in the specifications.

Content reproduced with permission from HEDIS, Volume 2: Technical Specifications for Health Plans. To purchase copies of this publication, including the full measures and specifications, contact NCQA Customer Support at 888-275-7585 or visit

www.ncqa.org/publications.

Ad.7 Disclaimers: This HEDIS® performance measure is not a clinical guideline and does not establish a standard of medical care and has not been tested for all potential applications.

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

Ad.8 Additional Information/Comments: NCQA Notice of Use. Broad public use and dissemination of these measures, without modification, are encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Modifications to, and/or commercial use of, a measure requires the prior written consent of NCQA and is subject to a license at the discretion of NCQA. As used herein, “commercial use” refers to any sale, license, or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed, or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.