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 #: 0575

Measure Title: Comprehensive Diabetes Care: Hemoglobin A1c (HbA1c) Control (<8.0%)

Measure Steward: National Committee for Quality Assurance

Brief Description of Measure: The percentage of patients 18-75 years of age with diabetes (type 1 and type 2) whose most recent HbA1c level is <8.0% during the measurement year.

Developer Rationale: This measure assesses HbA1c control among diabetics. The improvement in quality envisioned by the use of this measure is to have more diabetic adults 18-75 years of age with HbA1c levels lower than 8.0%. This measure is critically important for clinical diabetes management, because keeping patients in this desirable range of HbA1c helps to prevent complications of diabetes.

Numerator Statement: Patients whose most recent HbA1c level is less than 8.0% during the measurement year.

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.

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 my 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 I or Type II) is found.

Measure Type: Outcome: Intermediate Clinical Outcome

Data Source: Claims, Electronic Health Data, Electronic Health Records, Paper Medical Records

Level of Analysis: Health Plan

IF Endorsement Maintenance – Original Endorsement Date: Dec 04, 2009 **Most Recent Endorsement Date:** Sep 02, 2014

Preliminary Analysis: Maintenance of Endorsement

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

Criteria 1: Importance to Measure and Report

1a. Evidence

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

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

The developer provides the following evidence for this measure:

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

Evidence Summary

- Based on the 2019 American Diabetes Association Standards of Care.
- Grade A recommendation that a reasonable A1C goal for many nonpregnant adults is <7% (53 mmol/mol).
- Alternative goals include more stringent A1C goals (such as <6.5% [48 mmol/mol]) for selected individual patients (Grade C) and less stringent A1C goals (such as <8% [64 mmol/mol]) may be appropriate for patients with certain other conditions (Grade B).
- Guidelines are supported by numerous randomized controllled trials with approximately 29,000 patients.
- The developer states that " This measure assesses HbA1c control among diabetics. The improvement in quality envisioned by the use of this measure is to have more diabetic adults 18-75 years of age with HbA1c levels lower than 8.0%. This measure is critically important for clinical diabetes management, because keeping patients in this desirable range of HbA1c helps to prevent complications of diabetes." Glycemic control, especially early in the course of the disease, is strongly associated with reductions in complications and cardiovascular disease, and therefore better patient outcomes. There may be more risk to stringent controls in patients with long-standing type 2 diabetes or at significant risk of cardiovascular disease.

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:

Questions for the Committee:

- *The developer attests the underlying evidence for the measure has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat discussion and vote on Evidence?*

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**

RATIONALE:

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 provides the following data to demonstrate the variation in the rate of patients with diabetes that had poor HbA1c control.

Commercial plans

	Mean	**Standard deviation:**	**Min**	**10th**	**90th**	**Max**
2016	51%	16%	0%	28%	65%	75%
2017	53	15	1	30	66	76
2018	54	15	1	37	66	77

Medicaid

	Mean	**Standard deviation:**	**Min**	**10th**	**90th**	**Max**
2016	47%	12	0	34	59	72
2017	49	10	0	37	60	72
2018	49	12	0	35	61	70

Medicare

	Mean	**Standard deviation:**	**Min**	**10th**	**90th**	**Max**
2016	64%	13%	0%	47%	76%	92%
2017	65	14	0	49	77	85
2018	67	12	4	53	78	90

Disparities

The developer cites a CMS/RAND report on blood sugar control disparities by race:

- Asian or Pacific Islander women: 90.2%
- White women: 83.2%,
- Hispanic women: 82.0%
- Black women: 78.3%
- Asian or Pacific Islander men: 88.8%
- White men: 83.5%
- Hispanic men: 80.9%
- Black men 76.5%

The developer also notes that "Although racial disparities in complications are somewhat less marked in populations receiving uniform access to care, disparities in HbA1c (A1C) level among African Americans, Asians, and Latinos have been shown compared with non-Hispanic whites."

Questions for the Committee:

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

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

RATIONALE:

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

1a. Evidence

Comments:

** No new studies. Cites studies with Grade A, B, and C evidence that supports A1C goal <7% to reduce rates of microvascular and macrovascular complications. Same evidence as for measure 0059

** Aligns with and is based on the 2019 American Diabetes Association Standards of Care. Randomized controlled studies that include approximately 29,000 individuals were conducted. There has not been any new evidence related to this measure. It continues to support the standard of care for diabetes.

** Appears evidence for measure has changed since last NQF endorsement review. No need to vote on evidence. Interesting that measure uses <8% rather than <7%.

** Not aware of any new studies/information

** This is an intermediate outcome measure. There has been no significant change in evidence since the last endorsement of this measure. The evidence Grade that this measure is based on is Grade A.

** There is no new studies to change the evidence base for this measure.

1b. Performance Gap

Comments:

** Was current performance data on the measure provided? Yes. Three populations presented (Commercial, Medicaid and Medicare) had variability in the means between these groups (49% to 67%), with trend toward improvement in the measure comparing data from 2016 thru 2018. How does it demonstrate a gap in care (variability or overall less than optimal performance) to warrant a national

performance measure? As above. 10th percentile performance, (indicator poor performance) was between 35% and 53% in the three payer groups reported. Interquartile range is ~11% in all three payer cohorts.

** The performance rate reported for the commercial population in 2018 was 54%, Medicaid 49% and Medicare 67% indicating that there is still a performance gap indicating a need for a national measure. Results indicate that there are still racial disparities for the African American, Latino and Asian populations when compared to non-hispanic whites. There is a lot of evidence that supports that there continues to be a performance gap.

** Data on plan performance provided (commercial and gov payer) indicating opportunity for improvement.

** Yes, overall less than optimal performance

** Although there has been improved performance on this measure over time, a performance gap remains

** Current performance data provided indicates that commercial and medicaid is about 50% with medicare around 65%. This indicates a significant gap warranting a national performence measure.There has been mild improvement over the three years reported.

Disparities:
Comments:
** Data from CMS Office of Minority Health reports that Black women and men had the lowest rates of diabetes control, followed by Hispanic women and men, whites and Asian/Pacific Islanders.
** Populations with consistent access to care show a lower level of disparities. However, overall, African American, Asian and hispanic populations continue to have disparities when compared to non-hispanic white populations.
** Information regarding differences by race re blood sugar control provided, but also states that multiple factors may drive differences. Health plan type used as proxy.
** Yes, black beneficiaries rate lowest among all populations
** Evidence exists to suggest disparities in care exist based on race/ethnicity
** HEDIS data is reported by estimates of racial/ethnicity from information from CMS administrative data, surname, and residential location

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:

Methods Panel Review (Combined)

Methods Panel Evaluation Summary:

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

- **Reliability:** H-1; M-4; L-0; I-0 → Measure passes with MODERATE rating
  - Score level testing was conducted using the beta-binomial methodology defined by Adams
  - 401 commercial plans, 250 Medicaid plans, and 477 Medicare plans were analyzed
  - 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	0.995	0.808	0.978	0.978	0.979	0.983	0.995	1.000
Medicaid	0.978	0.611	0.885	0.949	0.952	0.957	0.961	1.000
Medicare	0.975	0.768	0.964	0.968	0.969	0.976	0.979	1.000

  - Panelist: The beta-binomial approach has been commonly used, but reliability score obtained with this approach may not support the assertion that "the higher the reliability score, the greater is the confidence with which one can distinguish the performance of one plan from another." (Testing form page 5, 2a2.2). If n is sufficiently large, which is the case for this measure, it is very easy to obtain a very high reliability score.
  - Panelist: The number and types of health plans used is acceptable (commercial (n=401), Medicare (n=477), Medicaid (n=250). The overall reliability, 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.
- **Validity:** H-2; M-2; L-0; I-1 → Measure passes with MODERATE rating
  - Score level testing was conducted using correlation analyses for construct validity
  - Developer tested for construct validity of the Comprehensive Diabetes Care (CDC): HbA1c Control (<8.0%) 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 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.
        - CDC: Medical Attention for Nephropathy: The percentage of adults 18-75 with diabetes that had a nephropathy screening test or evidence of nephropathy during the measurement year.
        - CDC: Blood Pressure Control (<140/90 mm Hg): The percentage of adults 18-75 with diabetes whose most recent blood pressure level taken during the measurement year is <140/90 mm Hg.
    - Results ranged from 0.35 to 0.99 indicating moderate to very strong correlation.
    - Panelist: IQR is reasonable, but the t-test described in 2b4.1 doesn't make sense. It is just comparing two proportions.
    - Panelist: 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.
    - Panelist: The exclusion criteria for advanced illness are very expansive, for example, patients with heart failure (among 18 – 75) are excluded. I am surprised that only 2% patients were excluded by applying the advanced illness and frailty criteria for patients aged 66 and older (Testing form page 10, 2b2.2) as noted in the testing form, the prevalence rate of heart failure alone would seem to be much higher than that.
    - Panelist: The measure does utilize multiple data sources but did not address possible comparable results when more than one source was available for a plan.

Scientific Methods Panel Votes: Measure passes

Reliability: H-1; M-4; L-0; I-0
Validity: H-2; M-2; L-0; I-1

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

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.

The subgroup members achieved consensus on reliability and validity in their preliminary analyses with a vote on reliability and validity of moderate for both. This measure was discussed in conjunction with measures NQF 0059 and NQF 0061. The Panel elected to retain the vote captured before the meeting after the discussion. The Primary Care and Chronic Illness Standing Committee will evaluate this measure in the fall 2019 cycle. The

Panel agreed that a reconsideration of the measure was not warranted, and the votes submitted for the preliminary analysis will stand as the final vote.

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:

** Which data elements, if any, are not clearly defined? None Which codes with descriptors, if any, are not provided? None Which steps, if any, in the logic or calculation algorithm or other specifications (e.g., risk/case-mix adjustment, survey/sampling instructions) are not clear? None What concerns do you have about the likelihood that this measure can be consistently implemented? None

** Measure implemented with 401 commercial plans, 250 Medicaid plans, and 477 Medicare plans where data was anlayzed. The results indicate that the measure is reliable. Data elements are clearly defined. The measure includes a value-set for consistency in reporting. The measure does not have risk or case-mix adjusting. The sampling instructions are clear and should result in a random, non-biased sample. The measure is able to be consistently implemented.

** no concerns

** no concerns

** no concerns

** Data elements are clear and specifications indicate it could be consistently implemented.

2a2. Reliability – Testing

Comments:

** Concerns about reliability - None. Testing of beta-binomial in 2018 shows the statistic to be greater than 0.7, with overall reliability greater than 0.9. (PASS)

** No concerns with the reliability of this measure.

** Reliability testing seems adequate. Methods Panel rates Moderate. Little need to discuss or vote on reliability.

** No concerns

** No concerns

** No concerns

2b1. Validity – Testing

Comments:

**no

** The Developer tested for construct validity of the Comprehensive Diabetes Care (CDC): HbA1c Control (<8.0%) measure by exploring whether it was correlated with other similar measures of quality. All measures included in this process were part of the NCQA comprehensive diabetes care measure. One comment related to how missing data were treated. Based on how bias of rates is determined for NCQA HEDIS measures, and whether the missing data or records would have had a greater than 5 perent impact to the rates, I do not have any concerns with the validity of this measure.

** While deemed acceptable, few questions around validity might suggest that committee should discuss and/or vote.

** No concerns

** No concerns

** No concerns

2b4-7. Threats to Validity

Comments:

** No threats to validity. Meaningful differences (statistically significant) in the interquartile range in all three plan types, showing meaningful differences. Exclusions removed on average 2% of plans population in the measure. Missing data is rare due to it be claims based information.

** This measure does use multiple data sources, claims, encounters, EHR, chart. However, all data elements from the data sources (lab value) are consistent so I do not view this as a threat to validity. We are evolving measures to use more electronically driven data sources and this measure supports that with the inclusion of EHR data. No concerns with missing data being a threat to validity. NCQA guidelines ensure consistent treatment of missing records and when it would bias results. I still questin whether it is appropriate to exclude patients that are frail or have advanced illnesses.

** General HEDIS response to missing data.

** No concerns

** No concerns

** No concerns

2b2-3. Other Threats to Validity

2b2. Exclusions

2b3. Risk Adjustment

Comments:

** No threats to validity - construct validity evaluated using A1C testing, Poor Control, Eye exam, nephropathy screening, and diabetes blood pressure control. Pearson Correlation coefficients were moderately or strongly correlated. Exclusions did not impact this measures performance. Also, risk adjustment (proxy of SES) did not impact measure performance.

** As discussed in the previous response field, patients that are frail or are diagnosed with advanced illness are excluded from the denominator. I recognize that prior to excluding clinical care recommendations were

considered and measure testing occured, however, I think it leaves room for an incremental threat to validity. Validation of the performance measure does minimize the threat.

** No risk adjustment; Differences of opinion regarding need.

** Measure captured across various plan types but no risk adjustment. Developer did a a study using a qualitative assessment and found SES did not have a meaningful impact on the results

** The question was raised with respect to heart failure being used as an exclusion

** Scientific panel raised concerns about social risk factorsand encouraged developer to consider adjustment.But not high enough concern to threaten measure.

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.

- Generated during the provision of care
- Some data is available in defined fields in electronic sources. Measure is collected using a range of sources to allow the greatest participation; developer anticipates more electronic data and less requirements for paper record review in the future.
- NCQA conducts audits for all HEDIS collection and reporting processes
- Commercial use ("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") requires written consent; non commercial use does not.

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:

** No concerns about data collection. Readily available within electronic health records, billing systems, and at the payer level, through claims data

** The data for this measure is routinely generated during care delivery. Data are available electronically and in charts. No concerns with the data collection strategy for this measure.

** Information seems to be obtainable with minimal effort. Patient information is often available in electronic format; plan level data collected by online submission system.

** No concerns

** Required data is captured in the routine delivery of care

** All data elements are routinely generated and available in EHR. No concerns.

Criterion 4: Usability and Use

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

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

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

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

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

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

OR

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

Accountability program details

This is used in several accountability and public reporting programs

- California Pay for Performance program (largest non-governmental physician incentive program in the United States)
- CMS Quality Payment Program
- CMS Medicare Star Rating Program (included in composite Medicare Advantage Star Rating)
- CMS Medicaid Adult Core Set
- Used in scoring for accreditation of Medicare Advantage Heath Plans
- NCQA Report Cards, accreditation programs, certification programs, HEDIS Quality Compass

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 the Quality Compass tool and also presents data via conferences and webinars.
- NCQA measures are reviewed using a consensus-based process to consider input from multiple stakeholders; developer states this process includes multi-stakeholder advisory panels, public comment posting, and review of questions submitted to the Policy Clarification Support System.
- NCQA states this is long-standing measure and few questions are received. Minor clarifications are made during the annual update process, to address questions recieved.

Additional Feedback:

N/A

Questions for the Committee:

- How have 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

- The developer states that "Performance across all plan types has generally improved over the past three years, with Medicare, Medicaid, and commercial plan performance increasing each year by about 1-2%."

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

Unexpected findings (positive or negative) during implementation

- None reported

Potential harms

- None reported

Additional Feedback:

N/A

Questions for the Committee:

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

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

RATIONALE:

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

4a1. Use - Accountability and Transparency

Comments:

** Reported by regional and federal entities and available outside of organizations/practices. Reported as part of health plan scorecards and for value based plans. Data is shared with entities being measured. Feedback on this measure from those being measured is reported to have been used in improving the use of the measure.

** This measure is being used for Medicare STARR rating system, accreditation of health plans, California pay-for-performance program, certification programs, and scoring Medicare Advantage plans. Very little feedback is received by the measure developer most likely because the measure has been in place for a long time.

** Measure included in multiple programs.

** HEDIS measure along with other PFP programs -yes and utilize open comment period for measures

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

** Measure is used in several accountability and public reporting areas.Feedback is given in reporting bt NCQA.

4b1. Usability – Improvement

Comments:

**Identifies practices/systems with populations with better diabetes control. This information can become part of learning collaboration wtihin and between plans, practices, systems. No unintended consequences are not apparent

**No harms in endorsing this measure. I prefer this measure to #0059, as this measure is not an inverse measure, and shows how well a system is performing in relation to recommended diabetes care. The rate does not need to be explained (such as in the #0059 measure where you have to state that a lower rate is good). Improved results should mean high-quality care. The rates are improving at least one percent per year.

**Improving the overall populations results outweighs the potential unintended negative consequences

**One concern that needs to be noted is the potential harm from attempts to intensify glycemic control (hypoglycemia) vs. the benefits of improved glycemic control

**Mild improvement noted. No harms anticipated or identified.

Criterion 5: Related and Competing Measures

Related or competing measures

- 2608 : Diabetes Care for People with Serious Mental Illness: Hemoglobin A1c (HbA1c) Control
- 0729 Optimal Diabetes Care
- NQF staff also identified NQF 0059 Comprehensive Diabetes Care: Hemoglobin A1c (HbA1c) Poor Control (>9.0%) as a competing measure. This measure calculates the percentage of patients 18-75 years of age with diabetes (type 1 and type 2) whose most recent HbA1c level is >9.0% during the measurement year. Does the Committee feel there is added benefit associated with having these two measures both endorsed?

Harmonization

- The developer states that 2608 looks at a different population and they are harmonized to the extent possible.
- The developer states that 0729 is an all-or-none composite that relies on medical record abstraction and is reported at the physician level of accountability; this measure is health plan level and uses admin data or record review.

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

5. Related and Competing

Comments:

**none

**2608, 0729 and 0059 are similar measures. However, 2608 is specific to a subset of the population that may be diagnosed with diabetes from medications used to treat specific mental health conditions. 0059 is the inverse measure that measures patients in poor control of their diabetes. 0729 is a physician level reporting rather than a health plan level reporting so they are not measuring the same thing. No additional steps for harmonization needed.

**Multiple other measures in HEDIS comprehensive diabetes care. Unless there are specific unaddressed differences in the inclusions / exclusion criteria, I can't image that there is the need for both this measure and measure 59.

**Competing measures appear to be harmonized

**Measure 0059 is potentially competing, but may be deemed complementarty

**A related measure focuses on a different population ie mental health. The question of the need for both this measure and the Hemoglobin A1c (HbA1c) Poor Control (>9.0%) as a competing measure may warrant discussion.

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 02/04/2020

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

Combined Methods Panel Scientific Acceptability Evaluation

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 0575

Measure Title: Comprehensive Diabetes Care: Hemoglobin A1c (HbA1c) Control (<8.0%)

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 #1: One 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 extremely 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, is 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 #2: This measure is almost the opposite of measure #0059, as evidenced by an extremely high negative correlation coefficient (-0.99). It is not clear whether both are needed.

   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 <u>VALIDITY</u> testing** of <u>patient-level data</u> conducted?

   ☒ **Yes** ☒ **No**

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

   Submission document: Testing attachment, section 2a2.2

   Panel Member #1: The Beta-binomial model seems aproriate.

Panel Member #2: The beta-binomial approach has been commonly used, but reliability score obtained with this approach may not support the assertion that "the higher the reliability socre, the greater is the confidence with which one can distinguish the performance of one plan from another." (Testin form page 5, 2a2.2). If n is sufficiently large, which is the case for this measure, it is very easy to obtain a very high reliability score.

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.

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

   Submission document: Testing attachment, section 2a2.3

   Panel Member #1: The number and types of health plans used is acceptable (commercial (n=401), 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 #2: I wish the developer had described how they obtained the overall reliability.

   I assume the overall reliability included in Table 2 is the average reliability of health plans, obviously it is not median (which is 0.979 for commercial). However, based on the information provided in the table, mathmetaiclly mean cannot be 0.995.

   Panel Member #4: The distribution of scores is not very large across percentiles. Actual performance rates and could be affected by a handful of patients; eg. For Medicare given the denominator of 411 patients, the difference between the 75th percentile and 90th percentile is 10 patients, which is not a large number in practice, and depending on where the plan falls relative to the measure cut-points, a difference of 1-2 patients could mean difference between 4 Star or 5 Star. The test results do indicate a high level of reliability however.

   Panel Member #5: Reliability level is acceptable (overall and median > 0.96). There is some variability in the minimum reliability scores, but 10th to 90th percentils are all > 0.91.

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

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

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

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

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

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

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

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

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

☐ **Low** (NOTE: Should rate 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: 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., 0.611 for a minimum). Why is this? This should be explored.

Panel Member #2: I think the beta-binomial approach is ok, it does produce an unrealistically high number. The overall reliability cannot be mean mathematically, so it is not clear what it is.

Panel Member #3: Signal to noise measures appear high for all health plans. Minimum/median sample size of 411 appears adequate to assure reliability.

Panel Member #4: See comments above about concerns about differentiating performance in practice.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

Submission document: Testing attachment, section 2b2.

Panel Member #1: No concerns.

Panel Member #2: The exclusion criteria for advanced illness are very expansive, for example, patients with heart failure (among 18 – 75) are excluded. I am surprised that only 2% patients were excluded by applying the advanced illness and frailty criteria for patients aged 66 and older (Testing form page 10, 2b2.2) as noted in the testing form, the prevalence rate of heart failure alone would seem to be much higher than that.

Panel Member #3: None

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: No concerns.

Panel Member #2: IQR is reasonable, but the t-test described in 2b4.1 doesn't make sense. It is just comparing two proportions.

Panel Member #3: Substantial variation across plans

Panel Member #4: See above.

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 #1: No concerns.

Panel Member #2: To identify numerator compliance, two data sources can be used, either based on administrative codes with three categories or medical record review that requires "a distinct numeric

result is required fro numerator compliance." (MIF page 5 S.5). These seem to be inconsistent across data sources and may potentially lead to systematic bias.

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: None

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

Submission document: Testing attachment, section 2b6.

Panel Member #1: No concerns.

Panel Member #3: 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: 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 #4: None.

Panel Member #5: NA

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'm not completely convinced that risk assessment is not necessary for this measure (especially for socioeconomic status).

Panel Member #2: I think it is appropriate that this measure is not risk adjusted**.**

Panel Member #3: 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**.**

For cost/resource use measures ONLY:

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

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

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

VALIDITY: TESTING

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

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: Construct validity was assessed by using a Pearson's correlation to compare the A1C measure to a few process measures (A1C Testing, Eye exam performed, etc.) as well as a few other intermediate clinical outcomes measures (BP 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 #2: The developer correlated the measure score with several other measures that are similarly focused on diabetic patients to estiablish construct validity.

    Panel Member #3: Correlation with other outcome and process measures

    Panel Member #4: They used construct validity using 5 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: No concerns.

    Panel Member #2: The correlations with three related process measures (HbA1c testing ,eye exam, nephropathy screening) are negative and moderate. The correlations with two similar intermediate outcomes measures are quite high, although it is not surprising that HbA1c good control is highly and

negatively correlated with HbA1c poor control, particularly if these two measures were derived from the same sample of patients.

The high correlation with blood pressure control measure is more informative and helpful although it seemd to vary a lot across types of health plans (0.888, 0.756, 0.583 for commercial, Medicaid, and Medicare, respectively).

Panel Member #4: Some concern about low correlations with several measures for commercial plans (i.e., HbA1c testing only .3571, eye exam 0.4217, attention for diabetic nephropathy 0.3738). If there was good control you would expect there to be a HIGH correlation with with this measure for good control. This may indicate there were many missing values or no testing. Correlation was higher for Medicaid and Medicare but would expect it to be higher.

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 are threats to validity and/or relevant threats to validity were not assessed OR if testing methods/results are not adequate)

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

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

    Panel Member #1: Construct validity was assessed by using a Pearson's correlation to compare the A1C measure to a few process measures (A1C Testing, Eye exam performed, etc.) as well as a few other intermediate clinical outcomes measures (BP 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 #3: Treatment of missing data not clearly described.

    Lack of risk adjustment is questionable.

    Panel Member #4: See above re low validity with some of measures expect high correlation.

Panel Member #5: None

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**

    Panel Member #2: Of the five measures that were used to establish construct validity, three are process measures, one is a meansure that is functionally related to this measure, only blood pressure control measure is more useful and informative.

    Panel Member #3: The lack of risk adjustment merits wider discussion. It is not just a statistical issue.

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.**

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

Good_Cntrl_Evidence_Form_-575--637088170316593685.docx

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

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

No

1a. Evidence (subcriterion 1a)

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0575

Measure Title: Comprehensive Diabetes Care: Hemoglobin A1c (HbA1c) Control (<8.0%)

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: Click here to name the health outcome

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

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

☒ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name what is being measured

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

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

1a.2 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram

should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

Adults with diabetes (type 1 or 2) >>> HbA1c test (screening) is performed >>> Test (screening) results are evaluated for adequate control >>> HbA1c results (<8.0%) >>> Health provider determines treatment to maintain HbA1c to desirable level >>> improvement in HbA1c level and/or quality of life (desired outcome).

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.

N/A

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

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

☒ 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: • **Title** • **Author**	• **Standards of Medical Care in Diabetes–2019.** • **American Diabetes Association** • **January 2019**

• **Date** • **Citation, including page number** • **URL**	• **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** • **URL: https://care.diabetesjournals.org/content/42/Supplement_1**
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.	Recommendations (2019): • A reasonable A1C goal for many nonpregnant adults is <7% (53 mmol/mol). (Grade A) • Providers might reasonably suggest more stringent A1C goals (such as <6.5% [48 mmol/mol]) for selected individual patients if this can be achieved without significant hypoglycemia or other adverse effects of treatment (i.e., polypharmacy). Appropriate patients might include those with short duration of diabetes, type 2 diabetes treated with lifestyle or metformin only, long life expectancy, or no significant cardiovascular disease. (Grade C) • Less stringent A1C goals (such as <8% [64 mmol/mol]) may be appropriate for patients with a history of level 3 hypoglycemia (altered mental and/or physical state requiring assistance), limited life expectancy, advanced microvascular or macrovascular complications, extensive comorbid conditions, or long-standing diabetes in whom the goal is difficult to achieve despite diabetes self-management education, appropriate glucose monitoring, and effective doses of multiple glucose-lowering agents including insulin. (Grade B)
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 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
	
	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 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 several randomized controlled trials of varying size to support the recommendations around glycemic targets: • Diabetes Control and Complications (DCCT) – 1441 Type 1 DM patients • Epidemiology of Diabetes Interventions and Complications (EDIC) – 1394 Type 1 DM patients • Kumamoto – 110 Type 2 DM patients • UK Prospective Diabetes Study - 3867 Type 2 DM patients • Action to Control Cardiovascular Risk in Diabetes (ACCORD) – 10251 Type 2 DM patients • Action in Diabetes and Vascular Disease (ADVANCE) – 11140 Type 2 DM patients • Veterans Affairs Diabetes Trial (VADT) – 673 Type 2 DM patients

Estimates of benefit and consistency across studies	Below is an excerpt of the ADA discussion of benefit estimated in these studies: "A1C and Microvascular Complications Hyperglycemia defines diabetes, and glycemic control is fundamental to diabetes management. The Diabetes Control and Complications Trial (DCCT), a prospective randomized controlled trial of intensive (mean A1C about 7% [53 mmol/mol]) versus standard (mean A1C about 9% [75 mmol/mol]) glycemic control in patients with type 1 diabetes, showed definitively that better glycemic control is associated with 50–76% reductions in rates of development and progression of microvascular (retinopathy, neuropathy, and diabetic kidney disease) complications. Follow-up of the DCCT cohorts in the Epidemiology of Diabetes Interventions and Complications (EDIC) study demonstrated persistence of these microvascular benefits over two decades despite the fact that the glycemic separation between the treatment groups diminished and disappeared during follow-up. The Kumamoto Study and UK Prospective Diabetes Study (UKPDS) confirmed that intensive glycemic control significantly decreased rates of microvascular complications in patients with short-duration type 2 diabetes. Long-term follow-up of the UKPDS cohorts showed enduring effects of early glycemic control on most microvascular complications. Therefore, achieving A1C targets of <7% (53 mmol/mol) has been shown to reduce microvascular complications of type 1 and type 2 diabetes when instituted early in the course of disease. Epidemiologic analyses of the DCCT and UKPDS demonstrate a curvilinear relationship between A1C and microvascular complications. Such analyses suggest that, on a population level, the greatest number of complications will be averted by taking patients from very poor control to fair/good control. These analyses also suggest that further lowering of A1C from 7% to 6% [53 mmol/mol to 42 mmol/mol] is associated with further reduction in the risk of microvascular complications, although the absolute risk reductions become much smaller. Given the substantially increased risk of hypoglycemia in type 1 diabetes trials and with polypharmacy in type 2 diabetes, the risks of lower glycemic targets may outweigh the potential benefits on microvascular complications. Three landmark trials (Action to Control Cardiovascular Risk in Diabetes [ACCORD], Action in Diabetes and Vascular Disease: Preterax and Diamicron MR Controlled Evaluation [ADVANCE], and Veterans Affairs Diabetes Trial [VADT]) were conducted to test the effects of near normalization of blood glucose on cardiovascular outcomes in

	individuals with long-standing type 2 diabetes and either known cardiovascular disease (CVD) or high cardiovascular risk. These trials showed that lower A1C levels were associated with reduced onset or progression of some microvascular complications." "A1C and Cardiovascular Disease Outcomes CVD is a more common cause of death than microvascular complications in populations with diabetes. There is evidence for a cardiovascular benefit of intensive glycemic control after long-term follow-up of cohorts treated early in the course of type 1 diabetes. In the DCCT, there was a trend toward lower risk of CVD events with intensive control. In the 9-year post-DCCT follow-up of the EDIC cohort, participants previously randomized to the intensive arm had a significant 57% reduction in the risk of nonfatal myocardial infarction (MI), stroke, or cardiovascular death compared with those previously randomized to the standard arm. The benefit of intensive glycemic control in this cohort with type 1 diabetes has been shown to persist for several decades and to be associated with a modest reduction in all-cause mortality." "Cardiovascular Disease and Type 2 Diabetes In type 2 diabetes, there is evidence that more intensive treatment of glycemia in newly diagnosed patients may reduce long-term CVD rates. During the UKPDS, there was a 16% reduction in CVD events (combined fatal or nonfatal MI and sudden death) in the intensive glycemic control arm that did not reach statistical significance (P = 0.052), and there was no suggestion of benefit on other CVD outcomes (e.g., stroke). However, after 10 years of observational follow-up, those originally randomized to intensive glycemic control had significant long-term reductions in MI (15% with sulfonylurea or insulin as initial pharmacotherapy, 33% with metformin as initial pharmacotherapy) and in all-cause mortality (13% and 27%, respectively). ACCORD, ADVANCE, and VADT suggested no significant reduction in CVD outcomes with intensive glycemic control in participants followed for shorter durations (3.5–5.6 years) and who had more advanced type 2 diabetes than UKPDS participants. All three trials were conducted in relatively older participants with longer known duration of diabetes (mean duration 8–11 years) and either CVD or multiple cardiovascular risk factors. The target A1C among intensive-control subjects was <6% (42 mmol/mol) in ACCORD, <6.5% (48 mmol/mol) in ADVANCE, and a 1.5% reduction in A1C compared with control subjects in VADT, with achieved A1C of 6.4% vs. 7.5% (46 mmol/mol vs. 58 mmol/mol) in ACCORD, 6.5% vs. 7.3% (48 mmol/mol vs. 56 mmol/mol) in ADVANCE, and 6.9% vs. 8.4% (52 mmol/mol vs. 68

	mmol/mol) in VADT. Details of these studies are reviewed extensively in "Intensive Glycemic Control and the Prevention of Cardiovascular Events: Implications of the ACCORD, ADVANCE, and VA Diabetes Trials". The glycemic control comparison in ACCORD was halted early due to an increased mortality rate in the intensive compared with the standard treatment arm (1.41% vs. 1.14% per year; hazard ratio 1.22 [95% CI 1.01–1.46]), with a similar increase in cardiovascular deaths. Analysis of the ACCORD data did not identify a clear explanation for the excess mortality in the intensive treatment arm. Longer-term follow-up has shown no evidence of cardiovascular benefit or harm in the ADVANCE trial. The end-stage renal disease rate was lower in the intensive treatment group over follow-up. However, 10-year follow-up of the VADT cohort showed a reduction in the risk of cardiovascular events (52.7 [control group] vs. 44.1 [intervention group] events per 1,000 person-years) with no benefit in cardiovascular or overall mortality. Heterogeneity of mortality effects across studies was noted, which may reflect differences in glycemic targets, therapeutic approaches, and population characteristics. Mortality findings in ACCORD and subgroup analyses of VADT suggest that the potential risks of intensive glycemic control may outweigh its benefits in higher-risk patients. In all three trials, severe hypoglycemia was significantly more likely in participants who were randomly assigned to the intensive glycemic control arm. Those patients with long duration of diabetes, a known history of hypoglycemia, advanced atherosclerosis, or advanced age/frailty may benefit from less aggressive targets."
What harms were identified?	Below is an excerpt from the ADA discussion of the harms identified in these studies: "The concerning mortality findings in the ACCORD trial and the relatively intense efforts required to achieve near euglycemia should also be considered when setting glycemic targets for individuals with longstanding diabetes such as those studied in ACCORD, ADVANCE, and VADT. Findings from these studies suggest caution is needed in treating diabetes aggressively to near-normal A1C goals in people with long-standing type 2 diabetes with or at significant risk of CVD. However, on the basis of physician judgment and patient preferences, select patients, especially those with little comorbidity and long life expectancy, may benefit from adopting more intensive glycemic targets (e.g., A1C target <6.5% [48 mmol/mol]) if they can achieve it safely without hypoglycemia or significant therapeutic burden."

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 assesses HbA1c control among diabetics. The improvement in quality envisioned by the use of this measure is to have more diabetic adults 18-75 years of age with HbA1c levels lower than 8.0%. This measure is critically important for clinical diabetes management, because keeping patients in this desirable range of HbA1c helps to prevent complications of diabetes.

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 good HbA1c control.

Comprehensive Diabetes Care: Hemoglobin A1c (HbA1c) Control (<8.0%)

N = Number of Health Plans

YEAR = Measurement Year

Commercial

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

2016|411|51%|16%|0%|28%|48%|55%|61%|65%|75%|13%

2017|403|53%|15%|1%|30%|49%|57%|62%|66%|76%|13%

2018|401|54%|15%|1%|37%|52%|58%|63%|66%|77%|11%

Medicaid

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

2016|271|47%|12%|0%|34%|42%|49%|54%|59%|72%|12%

2017|267|49%|10%|0%|37%|44%|51%|55%|60%|72%|11%

2018|250|49%|12%|0%|35%|44%|51%|56%|61%|70%|12%

Medicare

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

2016|472|64%|13%|0%|47%|57%|66%|73%|76%|92%|16%

2017|475|65%|14%|0%|49%|61%|69%|74%|77%|85%|13%

2018|477|67%|12%|4%|53%|63%|70%|74%|78%|90%|11%

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 sugar levels under control. Asian or Pacific Islander women were the highest performing group to control their blood sugar levels with performance at 90.2%. Compared to White women who performed at 83.2%, Asian or Pacific Islander women overall had a difference of greater than 3 percentage points. White women were more likely to have their blood sugar levels controlled than Black or Hispanic women by more

than 3 percentage points. Black women had the lowest rates of controlled HbA1c at 78.3%, followed by Hispanic women at 82.0% and again, White women performing at 83.2%. Similar trends were also found among Asian or Pacific Islander men, whose rates of controlled HbA1c levels were 88.8%. There was a difference of more than 3 percentage points between Asian or Pacific Islander Men and White Men, who performed at 83.5%. As seen with the women, Black men performed the worst at 76.5%, followed by Hispanic men who performed slightly better at 80.9%. There is an overall difference greater than 3 percentage points between White men and Black men whose blood sugar levels are controlled. Hispanic men and White men had a difference of less than a 3 percentage points in blood sugar control levels.

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

Although HEDIS measures are not stratified by race and ethnicity, researchers have explored disparities in HbA1c levels among adults with diabetes. Although racial disparities in complications are somewhat less marked in populations receiving uniform access to care, disparities in HbA1c (A1C) level among African Americans, Asians, and Latinos have been shown compared with non-Hispanic whites. Improvements in glycemic control have been shown to prevent microvascular complications, and large trials have demonstrated the need for glucose control among patients with diabetes. Literature has suggested that A1C control may be poorer among minority populations than among nonminority populations. A number of factors may drive differences in A1C control: biological, socioeconomic, and quality-of-care factors have been suggested. Lack of access to health care may also affect diabetes care among minority individuals.

Kirk, JK, et. al. 2006. Disparities in HbA1c Levels between African-American and Non-Hispanic White Adults with Diabetes. Diabetes Care. 2006; 29(9): 2130-2136.

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):*

Endocrine, Endocrine : Diabetes

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

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

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

This is not an eMeasure **Attachment:**

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

Attachment **Attachment:** 0575_CDC_HbA1c_Good_Control_Value_Sets_Fall_2019-637088131732250530.xlsx

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

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

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

Not an instrument-based measure

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

Yes

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

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 limited life expectancy. 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 in 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 more 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.*

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

Patients whose most recent HbA1c level is less than 8.0% 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)*

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

There are two data sources and approaches used for collecting data 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 HbA1c test during the measurement year. The member is not numerator compliant if the result for the most recent HbA1c test is =8.0% or is missing a result, or if an HbA1c test was not done during the measurement year.

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 code during the measurement year to evaluate whether the patient is numerator compliant.

VALUE SET / NUMERATOR COMPLIANCE

HbA1c Level Less Than 7.0 Value Set / Compliant

HbA1c Level 7.0-9.0 Value Set / Not compliant*

HbA1c Level Greater Than 9.0 Value Set / Not compliant

* The CPT Category II code (3045F) in this value set indicates most recent HbA1c (HbA1c) level 7.0%-9.0% and is not specific enough to denote numerator compliance for this indicator. For patients with this code, the organization must use other sources (laboratory data, hybrid reporting method) to identify the actual value and determine if the HbA1c result was <8%.

MEDICAL RECORD REVIEW

The most recent HbA1c level (performed during the measurement year) is <8.0% as identified by laboratory data or medical record review.

At a minimum, documentation in the medical record must include a note indicating the date when the HbA1c test was performed and the result. The member is numerator compliant if the most recent HbA1c level during the measurement year is <8.0%. The member is not numerator compliant if the result for the most recent HbA1c level during the measurement year is >/=8.0% or is missing, or if a HbA1c test was not performed during the measurement year.

Ranges and thresholds do not meet criteria for this indicator. A distinct numeric result is required for numerator compliance.

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 my 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 I 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 services 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 members 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 on or between July 1 of the year prior to the measurement year and the end of the measurement year.

-- Living long-term in an institution any time on or between July 1 of the year prior to the measurement year and the end of the measurement year as identified by the LTI flag in the Monthly Membership Detail Data File. Use the run date of the file to determine if an adult had an LTI flag any time on or between July 1 of the year prior to the measurement year and the end of the measurement year.

- Adults 66 years of age and older as of December 31 of the measurement year (all product lines) with frailty and advanced illness. Adults 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 or nonacute inpatient discharges (instructions below) on different dates of service, 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 (Inpatient Stay Value Set).

2. Confirm the stay was for nonacute care based on the presence of a nonacute code (Nonacute Inpatient Stay Value Set) 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 (Inpatient Stay Value Set).

2. Exclude nonacute inpatient stays (Nonacute Inpatient Stay Value Set).

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 recent HbA1c test during the measurement year through the search of administrative data systems.

STEP 4: Identify patients with a most recent HbA1c test performed.

STEP 5: Identify the most recent result. If that result has an HbA1c level <8.0%, then that patient is numerator compliant. If the most recent result is instead with an HbA1c level >/=8.0% or a missing result or if no HbA1c test was done during the measurement year, then the member is not in the numerator.

STEP 6: Calculate the rate 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

Good_Cntrl_Testing_Form_-575-.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*)**:** 0575
Measure Title: Comprehensive Diabetes Care: HbA1c Control (<8.0%)
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 specificaitons (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:	Measure Tested with Data From:

(*must be consistent with data sources entered in S.17*)	
☒ abstracted from paper record	☒ abstracted from paper record
☒ claims	☒ claims
☐ registry	☐ registry
☒ abstracted from electronic health record	☒ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

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

N/A

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

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 all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

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

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

This measure assesses whether adults enrolled in commercial, Medicare, and Medicaid plans who have diabetes (type 1 and type 2) had their most recent HbA1c level less than 8.0%. 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 401 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: HbA1c Control (<8.0%), 2018.

Product Type	Number of Plans	Median Denominator Size/Plan
Commercial	401	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) and through a systematic assessment of face validity with expert panels.

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	0.995	0.808	0.978	0.978	0.979	0.983	0.995	1.000
Medicaid	0.978	0.611	0.885	0.949	0.952	0.957	0.961	1.000
Medicare	0.975	0.768	0.964	0.968	0.969	0.976	0.979	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 exceeded the minimally accepted

threshold of 0.7, and the majority of plans 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): HbA1c Control (<8.0%) 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 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.
- CDC: Medical Attention for Nephropathy: The percentage of adults 18-75 with diabetes that had a nephropathy screening test or evidence of nephropathy during the measurement year.
- CDC: Blood Pressure Control (<140/90 mm Hg): The percentage of adults 18-75 with diabetes whose most recent blood pressure level taken during the measurement year is <140/90 mm Hg.

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 values 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 Poor Control (>9.0%)	Eye Exams	Medical Attention for Diabetic Nephropathy	Blood Pressure Control <140/90
HbA1c Control (<8.0%)	0.3571	-0.9896	0.4217	0.3738	0.8882

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 Poor Control (>9.0%)	Eye Exams	Medical Attention for Diabetic Nephropathy	Blood Pressure Control <140/90
HbA1c Control (<8.0%)	0.6656	-0.9868	0.6210	0.3297	0.7562

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 Poor Control (>9.0%)	Eye Exams	Medical Attention for Diabetic Nephropathy	Blood Pressure Control <140/90
HbA1c Control (<8.0%)	0.5646	-0.9657	0.5881	0.4348	0.5832

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 at the plan level the 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: HbA1c Control (<8.0%) 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-9 and 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, ICD-9 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

---

[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.

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: HbA1c Control <8.0%) measure.

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: HbA1c Control (<8.0%) 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** Click here to enter number of factors **risk factors**

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

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

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

N/A

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

NCQA conducted a study on a measure similar to the Comprehensive Diabetes Care: HbA1c Control (<8.0%) measure, the Comprehensive Diabetes Care: HbA1c Poor Control (>9.0%) measure, among Medicare Advantage plans to assess whether to account for a member's socioeconomic status (SES) when comparing plan performance. A qualitative assessment included key informant interviews exploring ways in which SES may affect performance on this and other select HEDIS measures, and whether there was a conceptual basis for case-mix adjustment or other strategies. In the quantitative analysis, we assessed whether SES affected plan performance, using member low-income status, dual eligibility, and disability as proxies for SES. For this measure, adjusting for SES did not have a meaningful impact on results. When adjusting for disparity in performance between low- and high-SES populations, plan ranks were not substantially impacted. When accounting for clinical and demographic factors, we found that low-SES beneficiaries were as likely, or more likely, to receive recommended care as high-SES beneficiaries. Our results suggest there is neither a conceptual nor empirical basis for risk adjustment for the Comprehensive Diabetes Care: HbA1c Poor Control (>9.0%) measure. Given the similarities between the Poor Control measure and the Comprehensive Diabetes Care: HbA1c Control (<8.0%) measure, we concluded that the findings of the study are applicable to the latter measure as well.

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 measure. 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 4. Variation in Performance for commercial, Medicaid, and Medicare health plans, 2018.

Plan Type	N	Average (%)	St Dev (%)	P10th (%)	P25th (%)	P50th (%)	P75th (%)	P90th (%)	IQR (%)	p-value

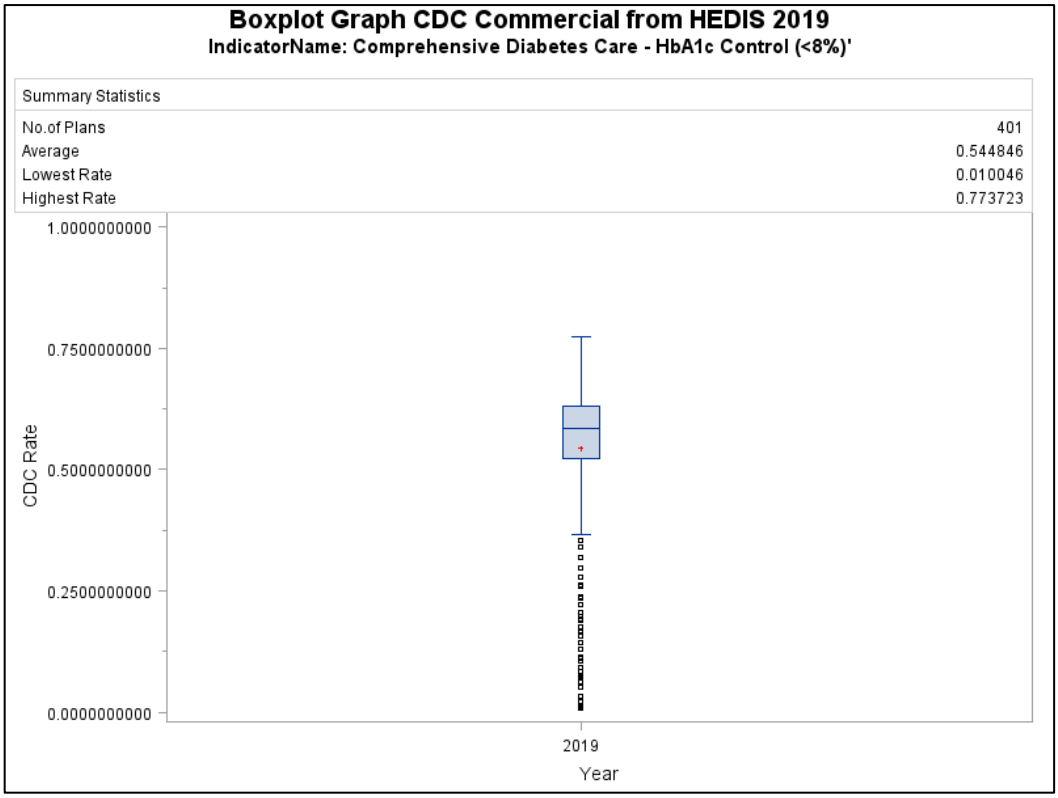
Commercial	401	54.48	14.59	36.87	52.31	58.39	63.11	66.18	10.80	<0.0001
Medicaid	250	48.74	11.54	34.54	44.04	51.22	55.96	60.68	11.92	<0.0001
Medicare	477	66.69	11.52	52.55	62.53	69.59	73.97	77.88	11.44	<0.0001

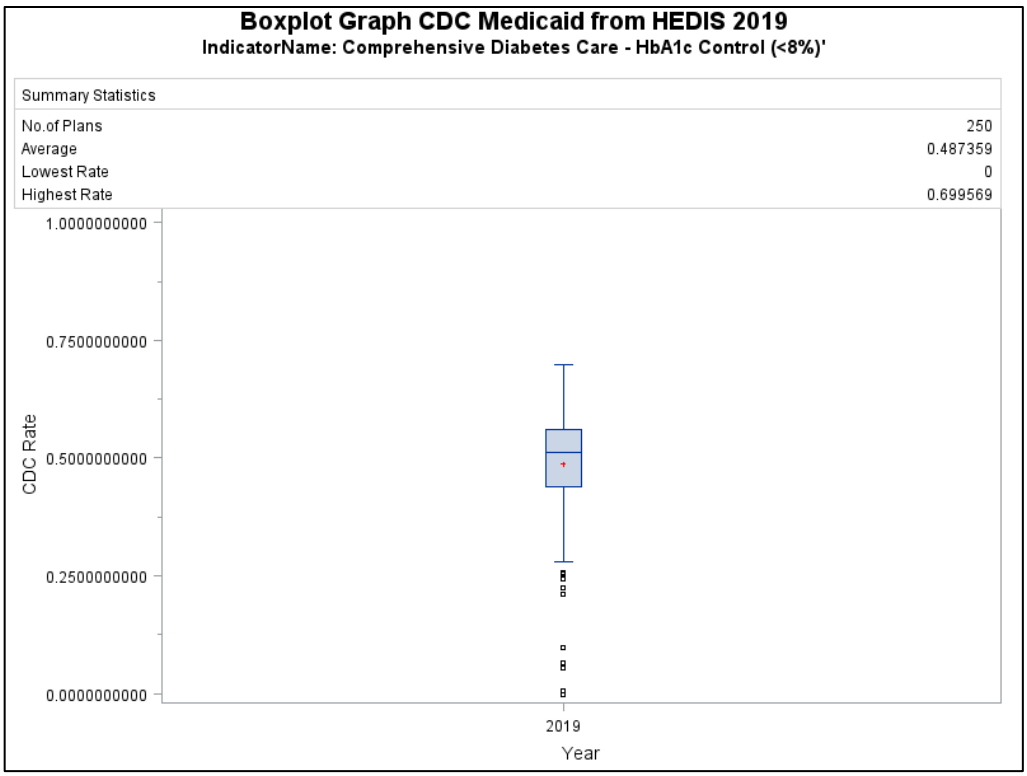
N = total number of plans reporting data

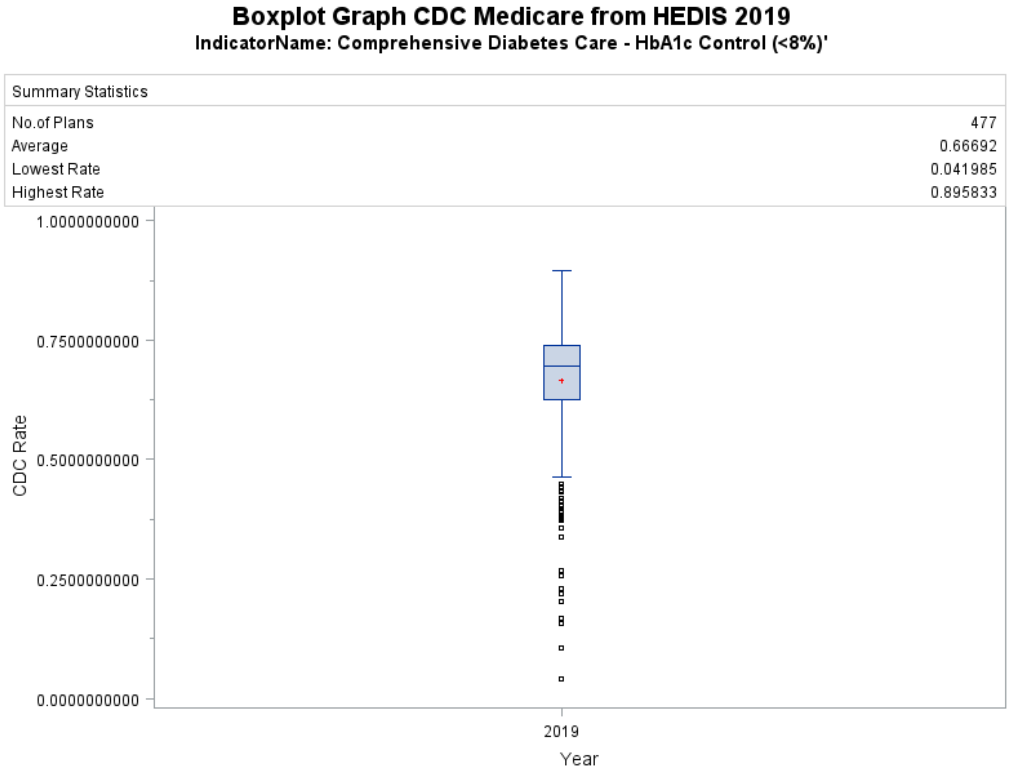
IQR: Interquartile range

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.







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: HbA1c Control (<8.0%) 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 https://www.ncqa.org/hedis/reports-and-research/ratings-2019/ Report Cards https://reportcards.ncqa.org/#/health-plans/list Health Plan Ratings https://www.ncqa.org/hedis/reports-and-research/ratings-2019/ Report Cards https://reportcards.ncqa.org/#/health-plans/list Payment Program IHA California Pay for Performance http://www.iha.org/manuals_operations_2014.html 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 Annual State of Health Care Quality https://www.ncqa.org/report-cards/health-plans/state-of-health-care-quality-report/

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

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.

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.

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.

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.
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.

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.

This is a long-standing, well-understood measure so NCQA receives very few questions or requests for clarification about it. Questions received through the Policy Clarification System have generally centered around clarification on optional exclusions in relation to the other Comprehensive Care Diabetes measures (HbA1c Control <7, poor control >9, eye exam or attention to nephropathy), guidelines supporting the age ranges for the measure, and methods used to convert units for the HbA1c result.

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

This measure has been deemed a priority measure by NCQA and other entities, as illustrated by its use in programs such as the Annual State of Healthcare Quality and the Health Plan Rating.

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.

Performance across all plan types has generally improved over the past three years, with Medicare, Medicaid, and commercial plan performance increasing each year by about 1-2%. We are encouraged by this continued improvement across health plans. Current average performance (MY 2018) is highest in Medicare plans (67%), followed by commercial plans (54%), and then Medicaid plans (49%).

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)

2608 : Diabetes Care for People with Serious Mental Illness: Hemoglobin A1c (HbA1c) Control (<8.0%)

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

0729 Optimal Diabetes Care. The measure steward is MN Community Measurement. This measure is NQF endorsed, but was not showing up in the previous question.

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.

There are two related measures that assess HbA1c control of <8% but they are either focused on different population, use different data sources or are specified at different levels of accountability than NQF 0575. Measure 2608 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 a serious mental illness and diabetes (type 1 and type 2) whose most recent HbA1c level during the measurement year is <8.0%. Measure 0729 is a composite measure (all or nothing) that uses physician reported data to assess the percentage of adult diabetes patients, 18-75 years of age, who have optimally managed modifiable risk factors HbA1c control (<8%) and four other indicators. HARMONIZED MEASURE ELEMENTS: All measures focus on an HbA1c target of <8% for adults age 18-75. DIFFERENCES: - Population Focus: While NQF 0575 and 0729 are focused on the general population of people with diabetes, NQF 2608 is focused on people with a serious mental illness and diabetes. - Data Source and Level of Accountability: Measure 00575 is collected through administrative claims and/or medical record review using health plan reported data. Measure 0729 is collected through medical record abstraction and reported at the physician level of accountability. IMPACT ON INTERPRETABILITY?AND DATA COLLECTION BURDEN:? The differences between measures 0575 and 2608 do not have an impact on interpretability of?publicly?reported rates or an impact on data collection burden as the measures are focused on different populations.

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

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.