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 #: 0059

Measure Title: Comprehensive Diabetes Care: Hemoglobin A1c (HbA1c) Poor Control (>9.0%)

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 >9.0% during the measurement year.

Developer Rationale: This measure assesses poor HbA1c control among diabetics. The improvement in quality envisioned by the use of this measure is to have fewer diabetic adults 18-75 years of age with HbA1c levels above 9.0%. This measure is critically important for clinical diabetes management, because the largest improvement in outcomes occurs by reduction of sugar levels in those patients with the highest glycohemoglobin levels.

Numerator Statement: Patients whose most recent HbA1c level is greater than 9.0% or is missing a result, or for whom an HbA1c test was not done 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 any setting during the measurement year or the year prior to the measurement year and who did NOT have a diagnosis of diabetes. These patients are sometimes pulled into the denominator via pharmacy data. They are then removed once no additional diagnosis of diabetes (Type 1 or Type II) is found.

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: Aug 10, 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 poor HbA1c control among diabetics. The improvement in quality envisioned by the use of this measure is to have fewer diabetic adults 18-75 years of age with HbA1c levels above 9.0%. This measure is critically important for clinical diabetes management, because the largest improvement in outcomes occurs by reduction of sugar levels in those patients with the highest glycohemoglobin levels." 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:

- 2019 Guidelines

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**

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 to data 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	38%	17%	12%	22%	56%	99%
2017	36	17	11	11	57	98
2018	34	16	15	15	49	99

Medicaid

	Mean	**Standard deviation:**	**Min**	**10th**	**90th**	**Max**
2016	43%	14%	15%	29%	59%	100%
2017	41	12	18	30	53	100
2018	41	14	20	28	58	100

Medicare

	Mean	**Standard deviation:**	**Min**	**10th**	**90th**	**Max**
2016	26%	15%	2%	12%	43%	100%
2017	25	15	6	12	42	100

2018	22	12	4	11	46	93

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:
**There is good evidence of the importance of improving poor diabetes control in the literature
**How does the evidence relate to the specific structure, process, or outcome being measured? Does it apply directly or is it tangential? Direct application of evidence to support this intermediate outcome measure, with applicable studies and level of evidence supplied. How does the structure, process, or outcome relate to desired outcomes? Improved glycemic control early in diabetes treatment reduces micro- and microvascular complications, and can reduce mortality, when intensity of treatment is appropriate to patient populations treated. For maintenance measures –are you aware of any new studies/information that changes the evidence base for this measure that has not been cited in the submission? No new evidence noted, but the 2019 ADA Standards of Medical Care in Diabetes referenced. (PASS)
**evidence remains good, though the data provided are not specific to an A1C greater than 9.0. How much greater is the risk of key outcomes with this specific threshold? I know it's higher, but by how much. Could not download an evidence table.
**The evidence for this measure directly applies to the outcome being measured. The measure developer referred to updated 2019 American Diabetes Association standards of care. The guidelines are supported by many randomized, controlled trials involving approximately 29,000 patients. The evidence presented confirms that control (though the evidence also points to better outcomes with HbA1c levels being <7%. I am not aware of new studies that were not presented. This measure received a Grade A recommendation (though it ties it back to nonpregnant adults <7% (53 mmol/mol). The evidence does not appear to have changed since last endorsement review.

**This maintenance intermediate clinical outcome measures 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. It is part of NQF approved composite measure 0731: Comprehensive Diabetes Care. This application includes reference to the 2019 American Diabetes Association Standards of Care regarding how the hemoglobin A1C result correlates with poor diabetes control but with the potential of prevention of complications if better diabetes control is achieved. However, the application does not explain or analyze the development of this Standard of Care. It mentions that different sources suggest different hemoglobin target numbers but does not discuss how 9.0 was chosen.
**I am not aware of any new study that changes the evidence base for this measure that has not been cited
**This is an intermediate outcome measure assessing gycemic control as evidenced by HbA1C. HbA1C goals are defined by Grade A level evidence and are included in the American Diabetes Association Standards of Care.
**No change in evidence

1b. Performance Gap

Comments:
**There is current performance data that continues to show suboptimal and variable performance amongst the different insurance plans. These performance gaps outline the importance as an outcome measure
**Was current performance data on the measure provided? Yes. Three populations presented (Commercial, Medicaid and Medicare) had variability in the means between these groups (22% to 41%), 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. 90th percentile performance, (indicator poor performance) was between 36% and 58% in the three payer groups reported. (PASS)
**Yes, but presented in a very confusing way. The performance data for >9 are not consistent with the disparities data for "good control." Please ask the developer to use numerical thresholds, not terms like poor or good. I must be missing something since 80+ performance of "good control" is not concordant with about 40+% on >9. I doubt if 40% are >9, actually, so I probably am missing a nuance in the rate calculation that can be resolved easily.
**Less than optimal performance was included in the documents for review. Commercial 2018 rate was 34%, Medicaid 41% and 2017 Medicare 25%. The low level of control being measured remains an area of concern. The opportunity for improvement continues to be high and does warrant a national performance measure.
**The percentage of diabetic patients with the most recent hemoglobin A1c > 9.0% varied from 25 to 43%, extracted from HEDIS data of commercial, Medicare, and Medicaid health care plans 2016-18.The numbers of health care plans in each category are not provided.
**Yes, overall less than optimal performance (2018 Medicare population still has a mean of 22)
**Performance data on the measure was provided by the developer. Although there is a trend toward improvement on this measure over time, opportunities to improve care remain.
**Gap exists and warrants a national performance measure.

Disparities:

**There is evidence of racial disparities with this measure specifically amongst African Americans, Asians, and Latinos.
**Yes - 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. (PASS)
**The disparities are moderate, but still important. It would be useful to see age-stratified data.

**Race/ethnicity disparities continued in HbA1c level among African Americans, Asians, and Latinos have been shown compared with non-Hispanic whites. Gaps remain an area of concern. The data continues to demonstrate potential disparities in diabetes HbA1c testing.
**It is mentioned in other measure applications reviewed in this NQF Measure Review Cycle that HEDIS data does not stratify by race, ethnicity, or language. This application refers to a CMS/RAND report on blood sugar control disparities by race and gender, which shows significant disparities. However, the description of this study does not cite a reference or define the studies' definition of blood sugar control.
**Yes, black men rate lowest among all populations
**Data was provided to support the existance of disparities of care based on race and ethnicity.
**Not specific to the measure, but in general disparities presented in DM control.

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:

Scientific Methods Panel Votes: Measure passes

Reliability: H-2; M-3; L-0; I-0

Validity: H-1; M-3; L-0; I-1

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-2; M-3; L-0; I-0 → Measure passes with MODERATE rating
    - Score level reliability was tested using the beta-binomial approach.
    - Developer tested data from 378 commercial plans, 241 Medicaid plans, and 477 Medicare plans:

Product Line	Overall Reliability	Min	Percentiles					Max
			10th	25th	50th	75th	90th	
Commercial	0.996	0.831	0.980	0.982	0.983	0.987	0.992	1.000
Medicaid	0.983	0.627	0.915	0.955	0.966	0.973	0.977	1.000
Medicare	0.980	0.792	0.970	0.975	0.977	0.982	0.985	1.000

    - Panelist: I assume the overall reliability included in Table 2 is the average reliability of health plans, obviously it is not median (which is 0.983 for commercial). However, based on the information provided in the table, mathematically mean cannot be 0.996.
    - Panelist: concern that data element level testing was not conducted since there are multiple sources of data that can be used and if CPT codes ranges are allowed but if medical record a value must be provided, would like to see if these two methods have consistent results when both are available?
    - Panelist: 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 from numerator compliance." (MIF page 5 S.5). These seem to be inconsistent across data sources and may potentially lead to systematic bias.
    - Panelist: Ideally, would like to see comparison of performance rates in manually abstracted data and EMR data
    - Panelist: Since between vs. within plan variation is of concern for using this measure to discriminate between plan, intraclass correlation coefficients with plotted plan measures and standard errors would have provided more useful information to assess plan level variation
- **Validity:** H-1; M-3; L-0; I-1 → Measure passes with MODERATE rating
    - Developer tested for construct validity by Pearson's correlations with other similar measures of quality hypothesized which are listed below:
        - CDC: Hemoglobin A1c (HbA1c) Testing: The percentage of adults 18-75 with diabetes that had an HbA1c test performed during the measurement year.
        - CDC: HbA1c Control (<8.0%): The percentage of adults 18-75 with diabetes whose most recent HbA1c level is <8% 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 exhibited moderate to very strong inverse correlation, ranging from -0.32 to -0.99.
    - Note: The correlation values are all negative because the HbA1c Poor Control measure is a "lower is better quality" measure, while the other measures are "higher is better". This indicates that plans that have low rates on this measure will have high rates on the others.

- 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.

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 of these measures. The Panel noted the similarity between this measure and measure 0575 and retained the vote for Scientific Acceptability of the measure from the preliminary analysis. The Primary Care and Chronic Illness Standing Committee will evaluate this measure in the fall 2019 cycle.

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:

**I have no concerns

**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 (PASS)

**OK. Please note that I'm not sure why missing data and not performed are included in the rate. Is this standard practice?

**The code set (value set) is complete and clearly defined. The measure can be consistently implemented.

**The numerator is "Patients whose most recent HbA1c level is greater than 9.0% or is missing a result, or for whom an HbA1c test was not done during the measurement year". I would be comfortable with this numerator definition if the measure title was "... Poor Control (>9.0%) or Unknown Control". The demominator understandably excludes those diagnosed with gestational diabetes and steroid induced diabetes and those under hospice care. The application states that NCQA now 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. "Long-term in institutional settings" and "advanced illness and frailty" are hard to define as "yes" or "no". There are lots of shades of gray. This change may add to the risks of poor care for the older population, which has an increasing range of physical and mental capabilities, not a shrinking range. The data extraction is more complicated than some of the other meausres in this Cycle in which all needed data can be extracted from claims alone. Finding the most recent HbA1c level during the measurement year may require a good deal of "digging" through electronic and paper medical records and then matching this information with claims data - a much more complicated and time comsuming task.

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

**This measure was reviewed by the Methods Panel and received a moderate rating by the panel. There are no concerns wtih respect to consistent implementation of the measure.

**No concerns

2a2. Reliability – Testing

Comments:

**I do not have concerns about reliability testing of this measure

**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)

**Not more than in general for NCQA measures

**Developer tested data from 378 commercial plans, 241 Medicaid plans, and 477 Medicare plans. Reliability was calculated to be .098 to .996. No concerns with the reliability of the measure.

**Yes. As this is a maintenance measure, I would like to see an explanation of how the concerns for reliability and for data integrity were determined and addressed in the first years of implementation of this measure.

**No

**No concerns

**No concerns

2b1. Validity – Testing

Comments:

**I do not have concerns about validity testing of this measure

**Construct validity correlating poor control with A1C testing, A1C control, eye examinations, attention to nephropathy screening and BP control showed moderate to high negative correlations (because in this measure, lower is better, and the other measures, higher is better) (PASS)

**no

**The only concerns I continue to have is with the definition of advanced illness and whether an individual with advanced illness would benefit from HbA1c testing and poor control or in control determinations and the associated exclusion of patients that meet the criteria to be considered advanced illness from the denominator. The better indicator in this measure is the "in control" as it is not an inverse measure and more easily understood by those unfamiliar with measure reporting.

**I am surprised by the low (2%) exclusion rate for the old, frail, and institutionalized. Perhaps the definition of this exclusion is much tighter than I imagine, and thus potentially not as much of a threat. Better details in how the health plans used this definition to exclude older patients from this quality measure would be appreciated.

**No

**This measure was reviewed by the Methods Review Panel and rated moderate with respect to Validity.

**no concerns

2b4-7. Threats to Validity

Comments:

**I have no concerns

**Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality? t-tests comparing plans at interquartile range showed statistically significant differences in all three payer types. Missing data/no response: Does missing data constitute a threat to the validity of this measure? No. If data is materially biased (missing data is noted, and can not be rectified, it is not reported. (PASS)

**See my previous comment about missing data, but I don't think it's a "threat"

**I am not convinced that there needs to be both a poor control and an in-control indicator for this measure. I am not concerned about the number of potential data sources, as all require the same data lement, the HbA1c rate and do not see this as a threat to the measure validity. Whether it is the poor control or in-control diabetes indicator, the results or outcomes being reported do indicate meaningful differences in patient outcomes. I do not have concerns with missing data for this measure.

**Measures that are dependent only on claims data have the advantage that "no claim = no money". There is this most powerful incentive to have claims entered and entered correctly. This is not the case with medical record data, although the elements in the medical record to substiate the claim/charges have an incentive of somewhat the same power. The audit of HEDIS data is explained but the impact of the "materially biased" designations needs to be added up and analyzed..

**No concerns with the validity of this measure

**None noted

**no concerns

2b2-3. Other Threats to Validity

2b2. Exclusions

2b3. Risk Adjustment

Comments:

**Social risk factor can have impact on this measure and certain populations may have lower performance on this measure. I agree with the comment made by the SMP committee on this matter.

**Are the exclusions consistent with the evidence? Yes. Based on expert opinion for gestational diabetes and steroid induced diabetes, and hospice/ long term institutional patients. Review of those excluded by NCQA advanced illness and frailty criteria (dementia/neurodegenerative conditions, ESRD, HF, hepatic failure, metastatic cancer, pulmonary fibrosis and respiratory failure) revealed that 2% of patients were removed from denominator counts on average in 25 health plans. Are any patients or patient groups inappropriately excluded from the measure? No (PASS) For risk adjustment, NCQA found no significant performance between lower and higher SES populations. This proxy (SES) may not be an adequate surrogate for risk adjustment, but within the confines of health plan data, is probably adequate.

**I would like to hear more about what was done as it's portrayed as controversial in the working document

**This measure does not include risk adjustment. I continue to question the application of advanced illness as an exclusion for this measure and whether patients with advanced illness would benefit from monitoring their HbA1c control status.

**I agree with the analysis of some of the NQF staff that the issue of risk assessment/adjustment should be discussed.

**As measure is captured by the various plan population there may be a slight tie to accounting for some populations

**Several of the Methods Panel Review members raised concerns over the lack of risk adjustment. The developer did state conduct some analysis along these lines and stated that there was no evidence for the need.

**no concerns

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:

**Required data elements are readily available in electronic form in EHR.

**Feasibility: Which of the required data elements are not routinely generated and used during care delivery? None Which of the required data elements are not available in electronic form (e.g., EHR or other electronic sources)? None What are your concerns about how the data collection strategy can be put into operational use? Data is obtainable from electronic health records, billing databases, and claims based datasets if being used by health systems or practices to improve care, and from available claims data at the health plan level. (PASS)

**It's OK

**No concerns with the feasibility of this measure. Data elements should be available in electronic form during the delivery of care and services for persons diagnosed with diabetes.

**I would like a description from the NCQA on how the data from the various sources was matched up for analysis and of how much effort was involved in a measurement combining claims data with clinical record data, compared to a claims data only measure.

**Components of the datais are found in the EHR however other sources can be utilized to gather the data

**Required data is routinely generated during care delivery (claims/EHR)

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

Criterion 4: Usability and Use

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

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

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

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

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

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

OR

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

Accountability program details

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 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 reports some improvement over time. Performance across all plan types has generally improved (lower is better) over the past three years, with Medicare, Medicaid, and commercial plan performance decreasing 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

- No unexpected findings

Potential harms

- No harms identified

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:

**Performance of this measure has improved quality of health care in

**Use - Accountability and Transparency: How is the measure being publicly reported? Multiple programs, regionally and nationally based, are detailed within the NCQA provided documentation. Are the performance results disclosed and available outside of the organizations or practices whose performance is measured? Yes, by value based programs. For maintenance measures - which accountability applications is the measure being used for? See above Use - Feedback on the measure: Have those being measured been given performance results or data, as well as assistance with interpreting the measure results and data? Yes, and there is assistance offered by NCQA in their support systems. Have those being measured or other users been given an opportunity to provide feedback on the measure performance or implementation? Yes Has this feedback been considered when changes are incorporated into the measure? NCQA reports that minor clarifications have been made due to feedback. (PASS)

**OK

**This measure is used in Medicare, Medicaid, commercial and non-govermental physician value-based or pay-for-performance programs. It is used in the Medicare STAR rating system and in Medicaid scorecards and core measure set reporting.

**Pages 48-50 of this application cover well the multiple ways that the results are publicized and utilized. The process of receiving and acting on feedback is also there.

**Measure used in various programs and results are shared

**This measure is used in many accountability and public reporting programs.

**Publicly reported and used in accountability programs. Feedback from public reported.

4b1. Usability – Improvement

Comments:

**There is a theoretical unintended consequence in terms of bias against patients with A1c > 9, but clinically this is not observed in patient treatment. Furthermore the benefits of identifying poorly controlled diabetes outweighs unintentional theoretical consequences.

**4b1. Usability – Improvement: How can the performance results be used to further the goal of high-quality, efficient healthcare? There has been modest improvements in this measure in all three payer

groups seen in the past 3 years. Healthplan work in concert with health systems or provider groups working together has impacted this important outcome.

4b2. Usability – Benefits vs. harms: Describe any actual unintended consequences and note how you think the benefits of the measure outweigh them. None (PASS)

**OK

**Reported results continue to improve albiet only 1-2% per year. There are no harms associated with the use of the measure.

**There have been no discoveries of unintended benefits or harms from the use of this measure. At the A1C level of >9%, harms from hypoglycemia are unlikely but this is an issue when lower/tighter levels of diabetes control are advocated.

**Continuing to utilize this measure in national programs for various populations will continue to highlight the continued need to lower A1Cs and allows for more collaboration across the industry. No harm noted

**Ther is evidence that performance has improved over time, but there is still additional room for improvement.

**No real harms or unintended consequences.

Criterion 5: Related and Competing Measures

Related or competing measures

- 2607: Diabetes Care for People with Serious Mental Illness: Hemoglobin A1c (HbA1c) Poor Control
- NQF Staff also identified NQF 0575: Comprehensive Diabetes Care: Hemoglobin A1c (HbA1c) Control (<8.0%) as potentially competing. 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 <8.0% during the measurement year. Does the Committee feel there is added benefit associated with having these two measures both endorsed?

Harmonization

The developer states NQF 2607 looks at different populations and is harmonized to the extent possible.

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

5. Related and Competing

Comments:

**A related measure is NQF 0575 HbA1c Control < 8% but this is not a competing measure. I do feel they measure different diabetic patient populations (A1c < 8 identifies patients in control and A1c > 9 identifies patient uncontrolled). Identification of poorly controlled patients is important because they benefit the most from reduction in A1c in terms of diabetic complications. I am in favor of continuing to endorse this measure.

**No directly competing measures

**No additional steps needed

**NQF 0575 and 2607 are related. 2607 is a subset of a population and driven on impacts to HbA1c based on medications used to treat some mental health diagnosis. 0575 is the "in-control" measure, which I prefer since it does not involve an inverse rate and for public reporting, I like to see the positive results or progress to improved results from care and service delivery. It is more understandable to individuals that may look at results that are not heavily involved in care delivery or reporting.

**NQF measure 2607 also looks at the same age range of diabetic patients who have an A1C level of > 9%. But, this measure is for those who have a "serious mental illness" and diabetes. Do we need both or just this one?

**I agree with including 0575 and 2607 as the populations and interventions for these groups could help decrease the population climbing to >9

**There is a related/potentially competing measure, 0575 (A1c Control).

**Can the <8 A1c measure and this be combined or not?

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 1/31/2020

- **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: 0059

Measure Title: Comprehensive Diabetes Care: Hmoglobin A1c (HbA1c) Poor Control (> 9.0%)

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: This measure is almost the opposite of measure #0575, as evidenced by an extremely high negative correlation coefficient (-0.99). It is not clear whether both are needed.

   Panel Member #3: 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.

   Panel Member #5: The measure seems straightforward. The only concern I have is minor and it is more of a question than a concern. In the specifications for the denominator, it is indicated that the patient must have "a diagnosis of DM during the measurement year or the year prior." I wonder, how often is a patient's diagnosis of DM carried forward in EHR/chart documentation? That is, if the provider diagnosed the patient 10 years ago and knows the patient well, is he or she likely to enter DM as a diagnosis annually or even every other year? If so, if this a way to "game the system"? (e.g., If the provider has a patient who has a HgA1C >9 can he/she eliminate this patient from the denominator simply by not carrying the DM diagnosis forward in the record?)

RELIABILITY: TESTING

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

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

   ☐ **Yes** ☐ **No**
6. **Assess the method(s) used for reliability testing**

   Submission document: Testing attachment, section 2a2.2

   Panel Member #1: The 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 #3: Used beta-binominal approach measureing signal to noise. This is acceptable method.

   Panel Member #4: Beta-binomial testing performed – appropriate for measure.

   Panel Member #5: The Beta-binomial model seems aproriate.

   Panel Member #6: Since between vs. within plan variation is of concern for using this measure to discriminate between plan, intraclass correlation coefficients with plotted plan measures and standard errors would have provided more useful information to assess plan level variation. (Insufficient)

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

   Submission document: Testing attachment, section 2a2.3

   Panel Member #1: 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 0.983 for commercial). However, based on the information provided in the table, mathmetaiclly mean cannot be 0.996.

   Panel Member #4: Reliability level is acceptable (overall and median > 0.96)

   Panel Member #5: The number and types of health plans used is acceptable (commercial (n=378), Medicare (n=477), Medicaid (n=241). 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.

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

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes Panel Member #1:** with caution as noted above.

   ☒ **No**

   ☐ **Not applicable** (score-level testing was not performed)
9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?

   Submission document: Testing attachment, section 2a2.2

☒ **Yes**

☐ **No**

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

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

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

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

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

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

Note: Reviewer #6 marked this as insufficient but was not counted toward the final vote because this was not submitted on time.

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

Panel Member #3: concern that data element level testing was not conducted since there are multiple sources of data that can be used and if CPT codes ranges are allowed but if medical record a value must be provided, would like to see if these two methods have consistent results when both are available?

Panel Member #4: Reliabity at the plan level is >0.9 – highly reliable measure.

Panel Member #5: The number and types of health plans used is acceptable (commercial (n=378), Medicare (n=241), Medicaid (n=477). 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.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

Submission document: Testing attachment, section 2b2.

Panel Member #1: The exclusion criteria 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 #2: None

Panel Member #3: None.

Panel Member #4: Most exclusions were not formally tested – no concerns with those tested.;

Panel Member #5: The hospice, SNP and frailty make sense to me. Should long hospital stays also be taken into account in exclusions?

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

    Panel Member #2: Substantial variation across plans

    Panel Member #3: None

    Panel Member #4: None

    Panel Member #5: No concerns.

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: 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 #2: Ideally, would like to see comparison of performance rates in manually abstracted data and EMR data, but don't a priori assume bias.
    Panel Member #3: 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 #4: None.
    Panel Member #5: No concerns.

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

    Submission document: Testing attachment, section 2b6.

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

    Panel Member #3: None.

    Panel Member #4: Missing data is considered a 'fail' for this measure. This approach is not supported with analysis – preference would be to exclude.

    Panel Member #5: No concerns.

16. **Risk Adjustment**

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

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

    ☒ Yes ☒ No ☒ Not applicable

    16c. **Social risk adjustment:**

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

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

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

16d.**Risk adjustment summary:**

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

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

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

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

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

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

Panel Member #1: I think it is appropriate that this measure is not risk adjusted**.**

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

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

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

Panel Member #3: 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 #4: Developer argues that there is no conceptual reason for risk adjustment**.** There own analysis shows a difference in measure and reliabitliy across commercial/medicare/medicaid populations. The comparisons are stratified by plan type – although not a risk adjustment method per se, this will moderate the impact of SES differences.

Panel Member #5: I'm not completely convinced that risk assessment is not necessary for this measure.

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

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

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

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

VALIDITY: TESTING

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

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

    ☒ **Face validity**

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

☐ **N/A (score-level testing not conducted)**

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

Submission document: Testing attachment, section 2b2.2

Panel Member #1: The developer correlated the measure score with several other measures that are similarly focused on diabetic patients to estiablish construct validity.

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

Panel Member #3: **? this is statistical results of exclusions is this really the right section to review here? I think 2b1.2 and 2b1.3?** They used construct validity using 5 measures that should be correlated to the CDC measure.

Panel Member #4: Construct validity tested via assessing correlation with other CDC measures.

Panel Member #5: Construct validity was assessed by using a Pearson's correlation to compare the HgA1C measure to a few process measures (A1C Testing, Eye exam performed, Medical Attention for Nepropathy) as well as a few other intermediate clinical outcomes measures (A1C control and BP 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.

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

Submission document: Testing attachment, section 2b2.3

Panel Member #1: 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 poor control is highly and negatively correlated with HbA1c good 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.

Panel Member #3: **? this is about exclusions testing?** 2b1.3: The results for several of the correlations were poor (<.5 absolute value) especially for commercial health plans, but quite strong for the opposite measure HbAic control.

Panel Member #4: All correlations statistically and practically significant.

Panel Member #5: No concerns.

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: 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 #2: Treatment of missing data not clearly described. Lack of risk adjustment is questionable.

    Panel Member #3: Some concern about low correlations with measure of HbA1c testing (only -.32 for commercial plans). If there was low rates of testing, one would expect a high rate of patients with poor control due to missing values or no testing resulting in numerator compliance. Correlation was higher for Medicaid and Medicare but would expect it to be higher.

    Panel Member #4: None

    Panel Member #5: Construct validity was assessed by using a Pearson's correlation to compare the HgA1C measure to a few process measures (A1C Testing, Eye exam performed, Medical Attention for Nepropathy) as well as a few other intermediate clinical outcomes measures (A1C control and BP control). I have only a few concerns: 1) the rationale for comparing to the process measures is not clear to me; 2) Pearson correlations should be used for interval/ratio level data and, in my opinion, the process measures are not interval/ratio level.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

    ☐ **High**

    ☐ **Moderate**

    ☐ **Low**

    ☐ **Insufficient**

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

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

    Panel Member #3: Concern that appropriate testing for SES was not conducted. I believe there is a substantial body of literature pointing to low SES resulting in worse outcomes in diabetes due to lack of access to appropriate diet, ability to afford medications, and other social risk.

    Panel Member #4: Details for composite construct not supplied.

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.**

Brief Measure Information

NQF #: 0059

Corresponding Measures:

De.2. Measure Title: Comprehensive Diabetes Care: Hemoglobin A1c (HbA1c) Poor Control (>9.0%)

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

De.3. Brief Description of Measure: The percentage of patients 18-75 years of age with diabetes (type 1 and type 2) whose most recent HbA1c level is >9.0% during the measurement year.

1b.1. Developer Rationale: This measure assesses poor HbA1c control among diabetics. The improvement in quality envisioned by the use of this measure is to have fewer diabetic adults 18-75 years of age with HbA1c levels above 9.0%. This measure is critically important for clinical diabetes management, because the largest improvement in outcomes occurs by reduction of sugar levels in those patients with the highest glycohemoglobin levels.

S.4. Numerator Statement: Patients whose most recent HbA1c level is greater than 9.0% or is missing a result, or for whom an HbA1c test was not done during the measurement year.

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

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

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

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

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

S.20. Level of Analysis: Health Plan

IF Endorsement Maintenance – Original Endorsement Date: Aug 10, 2009 **Most Recent Endorsement Date:** Sep 02, 2014

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

0731:Comprehensive Diabetes Care

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

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

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

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of

healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***

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

xxxxxxxxxx.docx

1a.1 <u>For Maintenance of Endorsement:</u> 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.

1a. Evidence (subcriterion 1a)

Poor_Cntrl_Evidence_Form_-59-.docx

1b. Performance Gap

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

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

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

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

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

This measure assesses poor HbA1c control among diabetics. The improvement in quality envisioned by the use of this measure is to have fewer diabetic adults 18-75 years of age with HbA1c levels above 9.0%. This measure is critically important for clinical diabetes management, because the largest improvement in outcomes occurs by reduction of sugar levels in those patients with the highest glycohemoglobin levels.

1b.2. Provide performance scores on the measure as specified (<u>current and over time</u>) at the specified level of analysis. *(<u>This is required for maintenance of endorsement</u>. 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 poor HbA1c control.

Comprehensive Diabetes Care: Hemoglobin A1c (HbA1c) Poor Control (>9.0%)

N = Number of Health Plans

YEAR = Measurement Year

Commercial

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

2016|395|38%|17%|12%|22%|27%|34%|42%|56%|99%|15%

2017|383|36%|17%|11%|22%|26%|32%|41%|57%|98%|15%

2018|378|34%|16%|15%|21%|25%|30%|36%|49%|99%|11%

Medicaid

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

2016|267|43%|14%|15%|29%|36%|41%|49%|59%|100%|13%

2017|261|41%|12%|18%|30%|33%|38%|47%|53%|100%|14%

2018|241|41%|14%|20%|28%|33%|38%|47%|58%|100%|14%

Medicare

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

2016|472|26%|15%|2%|12%|15%|22%|33%|43%|100%|18%

2017|475|25%|15%|6%|12%|15%|20%|28%|42%|100%|13%

2018|477|22%|12%|4%|11%|14%|18%|25%|36%|93%|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. <u>If this is an eMeasure</u>, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)

This is not an eMeasure **Attachment:**

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

Attachment **Attachment:** 0059_CDC_HbA1c_Poor_Control_Value_Sets_Fall_2019-637088187123576417.xlsx

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

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

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

Not an instrument-based measure

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

Yes

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

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

NCQA added a hospice exclusion to most HEDIS measures in 2016. The focus on 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 is inappropriate.

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

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

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

Patients whose most recent HbA1c level is greater than 9.0% or is missing a result, or for whom an HbA1c test was not done 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 patient is numerator compliant if the most recent HbA1c level is >9.0% or is missing a result, or if an HbA1c test was not done during the measurement year. The patient is not numerator compliant if the result for the most recent HbA1c test during the measurement year is =9.0%.

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 / Not compliant

HbA1c Level 7.0-9.0 Value Set / Not compliant

HbA1c Level Greater Than 9.0 Value Set / Compliant

MEDICAL RECORD REVIEW

The most recent HbA1c level (performed during the measurement year) is >9.0% or is missing, or was not done during the measurement year, as documented through 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 patient is numerator compliant if the result for the most recent HbA1c level during the measurement year is >9.0% or is missing, or if an HbA1c test was not done during the measurement year. The patient is not numerator compliant if the most recent HbA1c level during the measurement year is =9.0%.

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

Note: A lower rate indicates better performance for this indicator (i.e., low rates of poor control indicate better care).

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

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

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

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

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

CLAIM/ENCOUNTER DATA

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

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

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

1. Identify all acute and nonacute inpatient stays.

2. Exclude nonacute inpatient stays.

3. Identify the discharge date for the stay.

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

1. Identify all acute and nonacute inpatient stays.

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

3. Identify the discharge date for the stay.

-- Only include nonacute inpatient encounters without telehealth.

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

See attached code value sets.

PHARMACY DATA

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

PRESCRIPTIONS TO IDENTIFY MEMBERS WITH DIABETES

DESCRIPTION / PRESCRIPTION

Alpha-glucosidase inhibitors / Acarbose, Miglitol

Amylin analogs / Pramlintide

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

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

Meglitinides / Nateglinide, Repaglinide

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

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

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

Thiazolidinediones / Pioglitazone, Rosiglitazone

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

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

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

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

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

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

ADMINISTRATIVE CLAIMS

Exclude patients who use hospice services or elect to use a hospice benefit any time during the measurement year, regardless of when the 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. Members 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 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 = Lower 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 >9.0%, a missing result or if no HbA1c test was done during the measurement year then that patient is numerator compliant. If the most recent result is instead with an HbA1c level </=9.0% then the number is not in the numerator.

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

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

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

Plans may report this measure using a systematic sample of 411 members. Plans are instructed to list and sort all eligible members for a measure. NCQA then provides plans with a Random Number Table that is released towards the end of the measurement year. The Random Number Table lists a value that is used to determine which members from the eligible populations (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 Management Organizations and Preferred Provider Organizations 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

Poor_Cntrl_Testing_Form_-59-.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.

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

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

Measure Number (*if previously endorsed*)**:** 0059
Measure Title: Comprehensive Diabetes Care: HbA1c Poor Control (>9.0%)
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: (***must be consistent with data sources entered in S.17***)	**Measure Tested with Data From:**
☒ abstracted from paper record	☒ abstracted from paper record
☒ claims	☒ claims
☐ registry	☐ registry
☒ abstracted from electronic health record	☒ abstracted from electronic health record

☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other:	☐ other:

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

N/A

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

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

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

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

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

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 greater than 9.0%. Therefore, testing was done at the health-plan level, which is appropriate for the level of reporting for this measure.

Note: A lower rate indicates better performance for this measure (i.e., low rates of poor control indicate better care).

We calculated the measure score reliability and construct validity from HEDIS data that included 378 commercial health plans, 477 Medicare health plans, and 241 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 Poor Control (>9.0%), 2018.

Product Type	Number of Plans	Median Denominator Size/Plan
Commercial	378	411
Medicaid	241	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	0.996	0.831	0.980	0.982	0.983	0.987	0.992	1.000
Medicaid	0.983	0.627	0.915	0.955	0.966	0.973	0.977	1.000
Medicare	0.980	0.792	0.970	0.975	0.977	0.982	0.985	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 Poor Control (>9.0%) 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 Control (<8.0%): The percentage of adults 18-75 with diabetes whose most recent HbA1c level is <8% 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 value sets within the measure that included ICD-9 codes. We used General Equivalence Mapping (GEM) to identify ICD-10 codes that map to ICD-9 codes and reviewed GEM mapping in both directions (ICD-9 to ICD-10 and ICD-10 to ICD-9) to identify potential trending issues.
2. NCQA then searched for additional codes (not identified by GEM mapping step) that should be considered due to the expansion of concepts in ICD-10. Using ICD-10 tabular list and ICD-10 Index, searches by diagnosis or procedure name were conducted to identify appropriate codes.
3. NCQA HEDIS Expert Coding Panel review: Updated value set recommendations were presented for expert review and feedback.
4. NCQA RMAP clinical review: Due to increase specificity in ICD-10, new codes and definitions require review to confirm the diagnosis or procedure is consistent and appropriate given the scope of the measure.
5. New value sets containing ICD-10 code recommendations were posted for public review and comment in 2014 and updated in 2015. Comments received were reconciled with additional feedback from HEDIS Expert Coding Panel and MAPs as needed.
6. NCQA staff finalized value sets containing ICD-10 codes for publication in 2015.

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

Expert Participation

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

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

The results from construct validity testing of the health plan level measure are presented by product line in Tables 3a, 3b, and 3c below.

Note: A lower rate indicates better performance for this measure (i.e., low rates of poor control indicate better care), which explains the negative relationship with the other measures, for which a higher rate indicates better performance.

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

	Pearson Correlation Coefficients				
	HbA1c Testing	HbA1c Control (<8.0%)	Eye Exams	Medical Attention for Diabetic Nephropathy	Blood Pressure Control <140/90
HbA1c Poor Control >9.0%)	-0.3266	-0.9896	-0.4370	-0.3419	-0.8986

Note: All correlations are significant at p<0.0001

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

	Pearson Correlation Coefficients				
	HbA1c Testing	HbA1c Control (<8.0%)	Eye Exams	Medical Attention for Diabetic Nephropathy	Blood Pressure Control <140/90
HbA1c Poor Control >9.0%)	-0.6663	-0.9868	-0.6275	-0.3286	-0.7820

Note: All correlations are significant at p<0.0001

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

	Pearson Correlation Coefficients				
	HbA1c Testing	HbA1c Control (<8.0%)	Eye Exams	Medical Attention for Diabetic Nephropathy	Blood Pressure Control <140/90
HbA1c Poor Control (>9.0%)	-0.5984	-0.9657	-0.6049	-0.4761	-0.5914

Note: All correlations are significant at p<0.0001

2b1.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)
Across all product lines, the correlations are moderate to very strong and statistically significant. These results confirm the hypothesis that plan performance on these diabetes measures are correlated with each other. Coefficients with absolute value of less than .3 are generally considered indicative of weak associations. Absolute values of .3 to .59 are considered moderate associations, absolute values of .6 to .69 indicate a strong positive relationship, and absolute values of .7 or higher indicate a very strong positive relationship. These correlation results suggest that this plan level measure has sufficient validity.

Note: The correlation values are all negative because the HbA1c Poor Control measure is a "lower is better quality" measure, while the other measures are "higher is better". This indicates that plans that have low rates on this measure will have high rates on the others.

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 Poor Control (>9.0%) 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-10 codes that were relevant to each of the conditions to create value sets for testing. To identify those with dementia, NCQA also included drug codes for medications such as donepezil hydrochloride and galantamine hydrobromide, to capture those who may not carry a diagnosis of dementia but are prescribed a drug for treatment.

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

---

[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.

bed confinement, palliative care and age-related physical debility. Members met the frailty proxy criteria if they had a claim for any of the codes included in the frailty code set in the measurement year.

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

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

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

Table 4 shows the results of applying the exclusion of adults 66 and older with advanced illness and frailty to the Comprehensive Diabetes Care: HbA1c Poor Control (>9.0%) 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 Poor Control (>9.0%) 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 _risk factors**

☐ **Stratification by _risk categories**

☐ **Other,**

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

N/A

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

NCQA recognizes that there is a growing body of literature that might support risk adjustment or stratification of intermediate outcome measures. However, at this time, NCQA does not currently risk adjust this measure given the potential to mask poor performance and disparities in care.

NCQA conducted a study on 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 this measure.

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 performance of the two plans differs significantly.

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)

Note that for this measure, a lower rate equals better performance so the values in the 10th percentile are representative of the top performers and the values in the 90th percentile are representative of the poor performers.

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

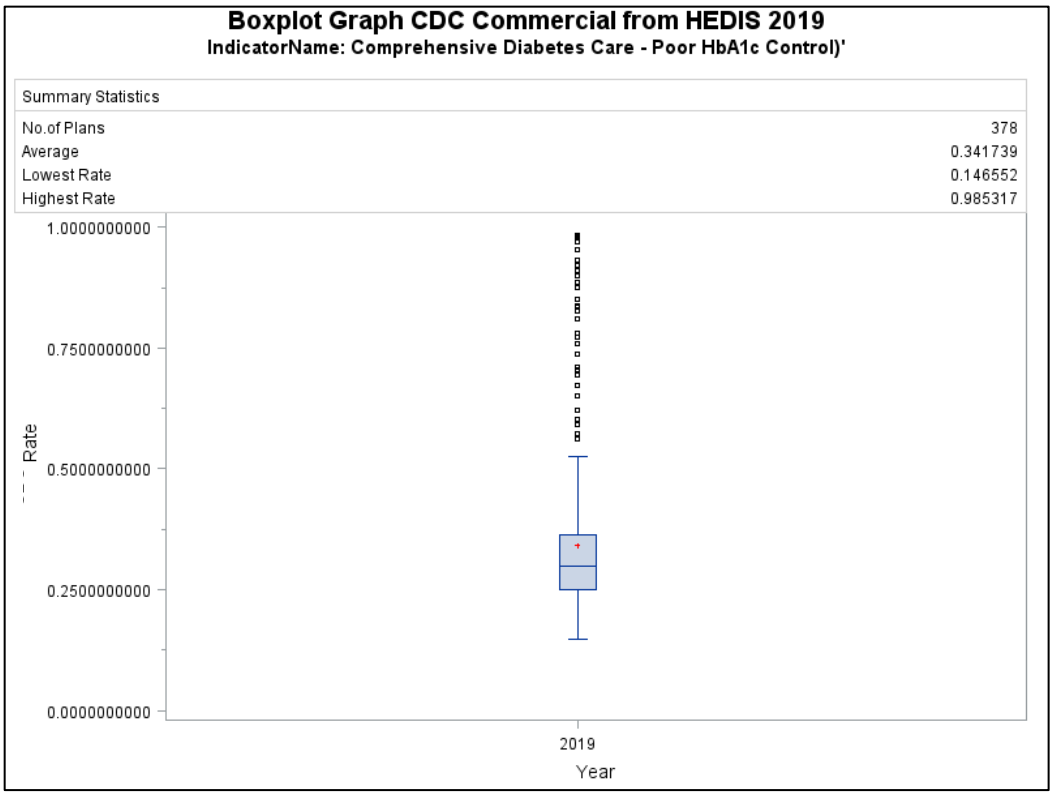
Plan Type	N	Average (%)	St Dev (%)	10th (%)	25th (%)	50th (%)	75th (%)	90th (%)	IQR (%)	p-value
Commercial	378	34.17	16.20	21.17	25.06	29.93	36.48	49.31	11.42	<0.0001
Medicaid	241	41.15	13.52	27.98	32.85	38.44	46.53	57.91	13.69	<0.0001
Medicare	477	21.86	12.24	11.39	13.87	18.25	25.35	36.40	11.48	<0.0001

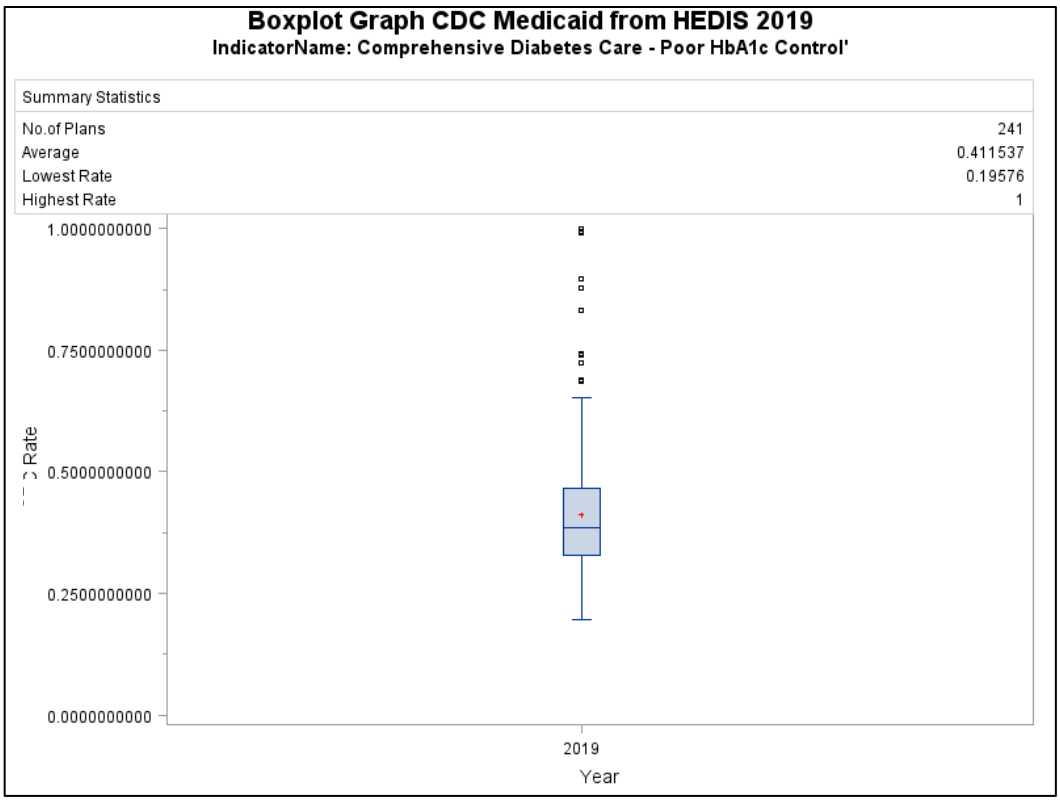
N = total number of plans reporting data

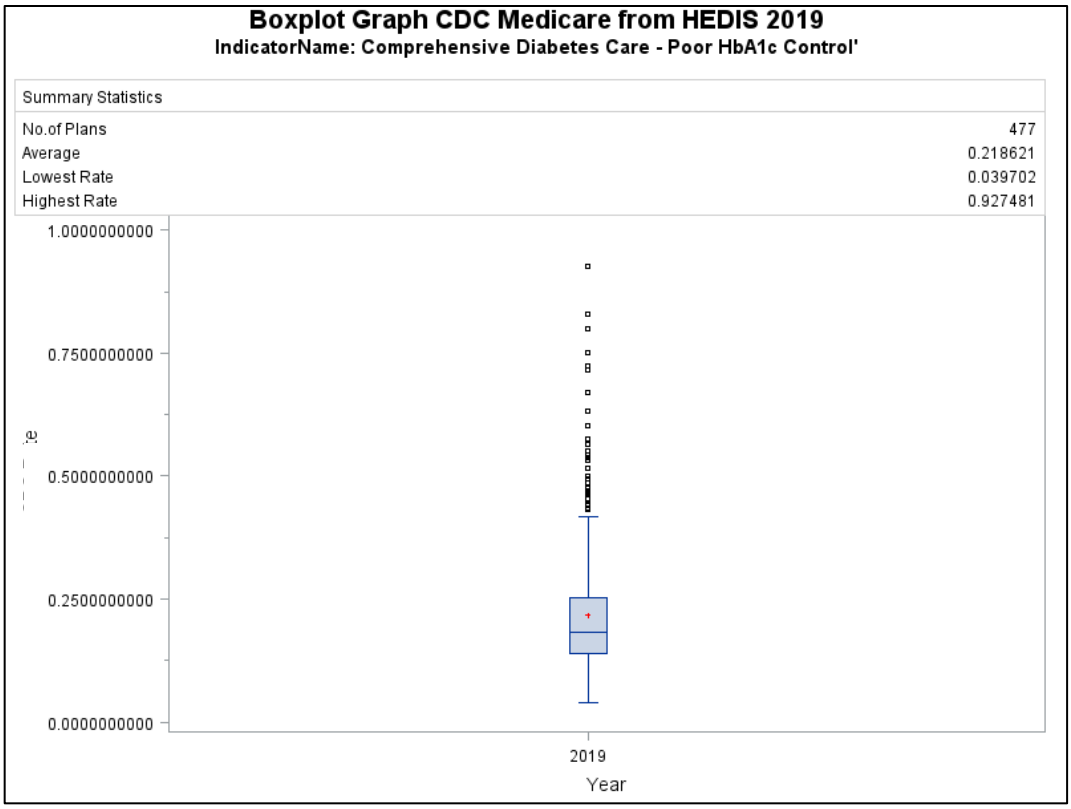
IQR: Interquartile range

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 (better performance) and 75th percentile 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 Poor Control (>9.0%) 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 Rating http://reportcard.ncqa.org/plan/external/plansearch.aspx Report Cards https://reportcards.ncqa.org/#/health-plans/list Health Plan Rating http://reportcard.ncqa.org/plan/external/plansearch.aspx Report Cards https://reportcards.ncqa.org/#/health-plans/list Payment Program IHA California Pay for Performance http://www.iha.org/manuals_operations_2014.html CMS Quality Payment Program https://qpp.cms.gov/ CMS Medicare Star Rating Program https://www.medicare.gov/find-a-plan/questions/home.aspx CMS Medicaid Adult Core Set https://www.medicaid.gov/medicaid/quality-of-care/performance-measurement/adult-core-set/index.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/ Quality Improvement (Internal to the specific organization)

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

CALIFORNIA VALUE BASED PAY FOR PERFORMANCE PROGRAM: 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 ten health plans representing 9 million insured persons. IHA reports results on approximately 35,000 physicians in 200 physician organizations. CMS MEDICARE ADVANTAGE STAR RATING PROGRAM: This measure is included in the composite Medicare Advantage Star Rating. CMS calculates a Star Rating (1-5) for all Medicare Advantage health plans based on 53

performance measures. Medicare beneficiaries can view the star rating and individual measure scores on the CMS Plan Compare website. The Star Rating is also used to calculate bonus payments to health plans with excellent performance. The Medicare Advantage Plan Rating program covers 11.5 million Medicare beneficiaries in 455 health plans across all 50 states.
CMS MEDICAID ADULT CORE SET: There are a core set of health quality measures for Medicaid-enrolled adults. The Medicaid Adult Core Set was identified by the Centers of Medicare & Medicaid (CMS) in partnership with the Agency for Healthcare Research and Quality (AHRQ). The data collected from these measures will help CMS to better understand the quality of health care that adults enrolled in Medicaid receive nationally. Beginning in January 2014 and every three years thereafter, the Secretary is required to report to Congress on the quality of care received by adults enrolled in Medicaid. Additionally, beginning in September 2014, state data on the adult quality measures will become part of the Secretary's annual report on the quality of care for adults enrolled in Medicaid.
CMS QUALITY PAYMENT PROGRAM: This measure is used in the Quality Payment Program (QPP) which is a reporting program that uses a combination of incentive payments and payment adjustments to promote reporting of quality information by eligible professionals (EPs).
DIABETES RECOGNITION PROGRAM: This measure is used in NCQAs 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 2017, a total of 184 Medicare Advantage health plans were accredited using this measure among others covering 9.2 million Medicare beneficiaries; 451 commercial health plans covering 113 million lives; and 125 Medicaid health plans covering 35 million lives. Health plans are scored based on performance compared to benchmarks.
HEALTH PLAN RATING/REPORT CARDS: This measure is used to calculate health plan rankings which are reported on the NCQA website. These rankings are based on performance on HEDIS measures among other factors. In 2019, a total of 255 Medicare health plans, 515 commercial health plans and 188 Medicaid health plans across 50 states were included in the rankings.
QUALITY COMPASS: This measure is used in Quality Compass which is an indispensable tool used for selecting a health plan, conducting competitor analysis, examining quality improvement and benchmarking plan performance. Provided in this tool is the ability to generate custom reports by selecting plans, measures, and benchmarks (averages and percentiles) for up to three trended years. Results in table and graph formats offer simple comparison of plans' performance against competitors or benchmarks.
STATE OF HEALTH CARE ANNUAL REPORT: This measure is publicly reported nationally and by geographic regions in the NCQA State of Health Care annual report. This annual report published by NCQA summarizes findings on quality of care. In 2018, the report included results from calendar year 2017 for health plans covering a record 136 million people, or 43 percent of the U.S. population.

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

N/A

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

N/A

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

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

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

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

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

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

Describe how feedback was obtained.

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

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

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 Support System have generally centered around clarification on optional exclusions in relation to the other Comprehensive Care Diabetes measures (HbA1c Control <8 or <7), guidelines supporting the age ranges for the measure, and methods used to convert units fort 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 Ratings.

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 (lower is better) over the past three years, with Medicare, Medicaid, and commercial plan performance decreasing each year by about 1-2%. We are encouraged by the continued improvement across health plans. Current average performance (MY 2018) is best in Medicare plans (22%), followed by commercial plans (34%), and then Medicaid plans (41%).

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)

2607 : Diabetes Care for People with Serious Mental Illness: Hemoglobin A1c (HbA1c) Poor Control (>9.0%)

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

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

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

Are the measure specifications harmonized to the extent possible?

Yes

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

There is another related measure that assesses poor control of HbA1c (>9%) but it is focused on a different population than NQF 0059. Measure 2607 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 >9.0%. HARMONIZED MEASURE ELEMENTS: Both measures focus on an HbA1c target of >9% for adults age 18-75 and are collected using administrative claims and/or medical record review using health plan reported data. DIFFERENCES: - Population Focus: NQF 0059 focuses on the general population of people with diabetes while NQF 2607 focuses on people with a serious mental illness and diabetes. IMPACT ON INTERPRETABILITY?AND DATA COLLECTION BURDEN:? The differences between measures 0575 and 2607 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: 1999

Ad.3 Month and Year of most recent revision: 07, 2019

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

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

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

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

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

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

www.ncqa.org/publications.

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

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

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