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 #: 0729

Measure Title: Optimal Diabetes Care

Measure Steward: MN Community Measurement

Brief Description of Measure: The percentage of patients 18-75 years of age who had a diagnosis of type 1 or type 2 diabetes and whose diabetes was optimally managed during the measurement period as defined by achieving ALL of the following:

- HbA1c less than 8.0 mg/dL
- Blood Pressure less than 140/90 mmHg
- On a statin medication, unless allowed contraindications or exceptions are present
- Non-tobacco user
- Patient with ischemic vascular disease is on daily aspirin or anti-platelets, unless allowed contraindications or exceptions are present

Please note that while the all-or-none composite measure is considered to be the gold standard, reflecting best patient outcomes, the individual components may be measured as well. This is particularly helpful in quality improvement efforts to better understand where opportunities exist in moving the patients toward achieving all of the desired outcomes. Please refer to the additional numerator logic provided for each component.

Developer Rationale: Measure is a composite; please refer to questions 1c.2., 1c.3., and 1c.4.

Numerator Statement: The number of patients in the denominator whose diabetes was optimally managed during the measurement period as defined by achieving ALL of the following:

- The most recent HbA1c in the measurement period has a value less than 8.0 mg/dL
- The most recent Blood Pressure in the measurement period has a systolic value of less than 140 mmHg AND a diastolic value of less than 90 mmHg
- On a statin medication, unless allowed contraindications or exceptions are present
- Patient is not a tobacco user
- Patient with ischemic vascular disease (Ischemic Vascular Disease Value Set) is on daily aspirin or anti-platelets, unless allowed contraindications or exceptions are present

Denominator Statement: Patients ages 18 to 75 with a diagnosis of diabetes (Diabetes Value Set) with any contact during the current or prior measurement period OR had diabetes (Diabetes Value Set) present on an active problem list at any time during the measurement period. Both contacts AND problem list must be queried for diagnosis (Diabetes Value Set).

AND patient has at least one established patient office visit (Established Pt Diabetes & Vasc Value Set) performed or supervised by an eligible provider in an eligible specialty for any reason during the measurement period.

Denominator Exclusions: Valid allowable exclusions include patients who were a permanent resident of a nursing home, pregnant, died or were in hospice or palliative care during the measurement year.

Measure Type: Composite

Data Source: Electronic Health Records, Paper Medical Records

Level of Analysis: Clinician : Group/Practice

IF Endorsement Maintenance – Original Endorsement Date: Mar 28, 2011 **Most Recent Endorsement Date:** Jun 30, 2015

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 health outcome measure include providing empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service; if these data not available, data demonstrating wide variation in performance, assuming the data are from a robust number of providers and results are not subject to systematic bias. For measures derived from patient report, evidence also should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.

Evidence Summary

Composite Component # 4: Patient is Tobacco-Free

- The developer provided a summary of the link between assessment of tobacco status of diabetes patient and maintaining the patient to be tobacco free. The developer also provided clarification that a provider can offer direction and advice to quit tobacco use, counseling, referral for cessation counseling and/or pharmacotherapy.

- The developer cited literature to provided the below literature/data:
  - Cigarette smoking is the leading cause of preventable death in the United States, accounting for more than 480,000 deaths, or one of five deaths, each year (CDC data)
  - The developer cited multiple literature that discussed a multifactorial approach to diabetes care that emphasizes the non-use of tobacco as the five clinical components to maximize health outcomes
- The developer also cited the following guideline:
  - ICSI Diabetes Mellitus in Adults, Type 2; Diagnosis and Management of. July 2014

- Tobacco smoking increases risk of macrovascular complications 4-400% in adults with T2DM and also increases risk of macrovascular complications.
- The guideline (p.32) provides interventions that should be emphasized by clinicians in tobacco cessation such as identifying a tobacco user, counseling and/or pharmacotherapy.

Changes to evidence from last review

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

☐ The developer provided updated evidence for this measure:

Updates:

The developer did not provide updated evidence for Component # 4: Patient is Tobacco-Free.

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*
- *If derived from patient report, does the target population value the measured outcome and finds it meaningful?*

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

Composite Component # 1: A1C is less than 8.0

- The developer provided a summary of the link between assessment of blood sugar control monitored by annual HbA1c lat test and maintaining the patient with an A1C less than 8.0 and reduction of risk of long term complication associated with macro and microvascular complications of hyperglycemia.
- The developer provided the following clinical guidelines
  - Institute for Clinical Systems Improvement (ICSI) Diabetes Mellitus in Adults, Type 2; Diagnosis and Management of. July 2014 Redmon B, Caccamo D, Flavin P, Michels R, Myers C, O'Connor P, Roberts J, Setterlund L, Smith S, Sperl-Hillen J. Institute for Clinical Systems Improvement. Diagnosis and Management of Type 2 Diabetes Mellitus in Adults. Updated July 2014. Page 19.
    - **Glycemic Control and A1c Goals Recommendation** A clinician should personalize goals with patients diagnosed with T2DM to achieve glycemic control with a hemoglobin A1c < 7% to < 8% depending on individual patient factors.
    - GRADE Methodology of Evidence Used;
      - High Quality of Evidence: Further research is very unlikely to change our confidence in the estimate of effect

        - Strength of Recommendation: Strong: The work group is confident that the desirable effects of adhering to this recommendation outweigh the undesirable effects. This is a strong recommendation for or against. This applies to most patients.
    - The developer summarized the Quality, Quantity, and Consistency of the body of evidence associated with the guideline.
    - The guidelines indicated the following harms of glycemic control through intensive pharmacotherapy: less benefit for major CV events; one large trial significantly increased mortality 20%; weight gain and severe hypoglycemia. The long-term cardiovascular safety of agents other than metformin and human insulins has yet to be established.
- The developer provided supplemental information for patients with Type 2 Diabetes Mellitus which promotes an A1C goal of less than 8% being more appropriate than an A1C goal of less than 7% when looking at factors such as CV disease, polypharmacy issues, limited life expectancy, cognitive impairment, comorbid conditions, and inability to recognize and treat hypoglycemia.
- The developer cited literature which supported a multifactorial approach to diabetes care including blood pressure, lips, glucose and aspirin use and non use of tobacco to maximize health outcomes.
- The developer continued to cite studies such as the United Kingdom Prospective Study, ACCORD Trial, ADVANCE trial to support hemoglobin A1C< 8% versus A1C< 7%.

Composite Component # 2: Blood Pressure less than 140 systolic and less than 90 diastolic

- The developer provided a summary of the link between assessment of blood pressure control at each visit and maintaining the patient with a blood pressure less than 140 systolic and 90 diastolic and reduction of risk of long term cardiovascular complications associated with hypertension.
- The developer provided the following clinical guidelines
    - Institute for Clinical Systems Improvement (ICSI) Diabetes Mellitus in Adults, Type 2; Diagnosis and Management of. July 2014 Redmon B, Caccamo D, Flavin P, Michels R, Myers C, O'Connor P, Roberts J, Setterlund L, Smith S, Sperl-Hillen J. Institute for Clinical Systems Improvement. Diagnosis and Management of Type 2 Diabetes Mellitus in Adults. Updated July 2014. Page 33.
        - **Antihypertensive therapy Recommendation.** A clinician should initiate antihypertensive treatment for patients with T2DM with a blood pressure $\geq$ 140/90 mmHG and treat to a goal of < 140/90.
        - GRADE Methodology of Evidence Used;
            - High Quality of Evidence: Further research is very unlikely to change our confidence in the estimate of effect
            - Strength of Recommendation: Strong: The work group is confident that the desirable effects of adhering to this recommendation outweigh the undesirable effects. This is a strong recommendation for or against. This applies to most patients.
            - The developer summarized the Quality, Quantity, and Consistency of the body of evidence associated with the guideline.
            - The guidelines indicated the following harms in diabetic patients who need two or three or more medications to achieve this level of blood pressure control: costly medications, risk of adverse reactions, medication interactions, and potential hypotension. However, guidelines believe benefits of treating hypertension outweigh the risks.

Composite Component # 3: Appropriate statin use for patients with diabetes (based on age, presence of ischemic vascular disease or LDL level greater than 190).

- The developer provided a summary of the link between assessment of diabetes patient variables/risk to determine appropriate statin use (i.e. cardiovascular disease, age, LDL.190) and maintaining the patient with a blood pressure less than 140 systolic and 90 diastolic and reduction of risk of long term cardiovascular complications as
- The developer provided the following clinical guidelines
  - Institute for Clinical Systems Improvement (ICSI) Diabetes Mellitus in Adults, Type 2; Diagnosis and Management of. July 2014 Redmon B, Caccamo D, Flavin P, Michels R, Myers C, O'Connor P, Roberts J, Setterlund L, Smith S, Sperl-Hillen J. Institute for Clinical Systems Improvement. Diagnosis and Management of Type 2 Diabetes Mellitus in Adults. Updated July 2014. Page 35.
    - Recommendations:
      - **ICSI- Statin Therapy (High Risk)**A clinician should recommend high-intensity statin therapy for patients diagnosed with T2DM, between the ages of 40-75 with established ASCVD (strong)
      - **ICSI- Statin Therapy (Moderate Risk)** A clinician should recommend moderate- or high-intensity statin therapy for all patients diagnosed with T2DM between the ages of 40-75 with a LDL $\geq$ 70 mg/dL. (strong)
      - GRADE Methodology of Evidence Used
        - High Quality of Evidence: Further research is very unlikely to change our confidence in the estimate of effect
        - Strength of Recommendation: Strong: The work group is confident that the desirable effects of adhering to this recommendation outweigh the undesirable effects. This is a strong recommendation for or against. This applies to most patients.
      - The developer summarized the Quality, Quantity, and Consistency of the body of evidence associated with the guideline.
      - The guidelines indicated potential for serious adverse events for statin therapy. However, guidelines believe benefits of statin therapy for patients with diabetes and high ASCVD outweigh the risks. Patient preference should be included in decision-making.
  - American College of Cardiology/ American Heart Association Guideline on the Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Risk in Adults. November 2013
    - Recommendations:
      - **ACC/AHA Recommendations for Primary Prevention in Individuals With Diabetes Mellitus and LDL–C 70-189 mg/dL**
        - Moderate-intensity statin therapy should be initiated or continued for adults 40 to 75 years of age with diabetes mellitus. (strong)
        - High-intensity statin therapy is reasonable for adults 40 to 75 years of age with diabetes mellitus with a ≥7.5% estimated 10-year ASCVD risk unless contraindicated. (expert opinion)
        - In adults with diabetes mellitus, who are <40 or >75 years of age, it is reasonable to evaluate the potential for ASCVD benefits and for adverse effects, for drug-drug interactions, and to consider patient preferences when deciding to initiate, continue, or intensify statin therapy. (expert opinion)

- Definition of evidence grading system used: Uses a combination of NHLBI and Class of Recommendation/ Level of Evidence (COR/LOE)
    - Strength of Recommendation: Strong (I A)
        - Class I = Benefits >>> Risks

        Procedure or treatment **should** be performed
        - Level A = multiple populations identified and data derived from multiple random control trials or mega-analysis.
- The developer summarized the Quality, Quantity, and Consistency of the body of evidence associated with the guideline.

Composite Component # 5: Appropriate daily aspirin or antiplatelet use for patients with diabetes with ischemic vascular disease.

- The developer provided a summary of the link between the assessment of diabetes patient for presence of ischemic vascular disease and treatment with aspirin or anti-platelet medication and reduction of risk of subsequent cardiovascular event.
- The developer provided the following clinical guidelines
    - Institute for Clinical Systems Improvement (ICSI) Diabetes Mellitus in Adults, Type 2; Diagnosis and Management of. July 2014 Redmon B, Caccamo D, Flavin P, Michels R, Myers C, O'Connor P, Roberts J, Setterlund L, Smith S, Sperl-Hillen J. Institute for Clinical Systems Improvement. Diagnosis and Management of Type 2 Diabetes Mellitus in Adults. Updated July 2014. Page 36.
        - **Aspirin Therapy Recommendation:** A clinician should recommend aspirin therapy for patients diagnosed with T2DM with established ASCVD and consider aspirin therapy for others where the benefits outweighs the risk in primary prevention.
        - GRADE Methodology of Evidence Used
            - High Quality of Evidence: Further research is very unlikely to change our confidence in the estimate of effect
            - Strength of Recommendation: Strong: The work group is confident that the desirable effects of adhering to this recommendation outweigh the undesirable effects. This is a strong recommendation for or against. This applies to most patients.
        - The developer summarized the Quality, Quantity, and Consistency of the body of evidence associated with the guideline.
        - The guidelines indicated the following harms:
            - Aspirin therapy could increase the risk of clinically significant bleeding and is also associated with medication cost. However, the substantial reduction in recurrent ASCVD events with aspirin therapy in secondary prevention will outweigh the risk of bleeding for patients with established ASCVD and no contraindications to aspirin use.
            - At this time, it is unclear whether adding aspirin therapy to other standard therapy for CV risk factors adds net benefit in patients with T2DM who do not have established ASCVD.

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:

Composite Component # 1: A1C is less than 8.0

- The developer cited a new 2018 study where type 2 diabetic patients who were successful for all five risk factor variables (bp, lipids, glucose, aspirin use and non-use of tobacco) had little or no excess risk of death, MI or stroke compared to general population.
    - Rawshani, A et al Risk Factors, Mortality, and Cardiovascular Outcomes in Patients with Type 2 Diabetes N Engl J Med 2018; 379:633-644 DOI: 10.1056/ NEJMoa1800256

Composite Component # 2: Blood Pressure less than 140 systolic and less than 90 diastolic

- The developer cited new ACC/ AHA Guidelines for the Diagnosis and Management of Hypertension which provides conflicting guidelines and was reviewed by MNCM's multi-stakeholder workgroupin April 2018 and rejected for incorporation into NQF # 0729
    - American College of Cardiology/ American Heart Association Guidelines for the Prevention, Detection, Evaluation and Management of High Blood Pressure in Adults November 13, 2017
        - These guidelines redefined the diagnosis of hypertension moving from > 140/ 90 to a new definition of stage 1 hypertension (130-139/ 80-89)
        - Developer cited specialty groups such as American Academy of Family Practice and American College of Physicians have declined endorsement of this guideline due to concerns with methodology and perceived conflict of interest.
    - The MNCM Workgroup recommended current BP component of targeting BP less than 140 systolic and less than 90 diastolic based on:
        - Guidelines do not address treatment risks (hypotension, kidney function).
        - Evaluation of SPRINT (Systolic Blood Pressure Intervention Trial) demonstrates that for the primary outcome of mortality, there is only a 0.5% difference between the intensive treatment group and the standard treatment group.
        - There is not consensus at this time among the guideline writing groups about the definition of hypertension or appropriate targets for high risk populations therefore not a clear direction for measurement to align with guidelines.

Composite Component # 3 : Appropriate statin use for patients with diabetes (based on age, presence of ischemic vascular disease or LDL level greater than 190). **And Composite Component #5: Appropriate daily aspirin or antiplatelet use for patients with diabetes with ischemic vascular disease.**

- The developer cited the 2018 ADA Standards of Care for cardiovascular disease and risk management, however noted no significant new studies change the statin use and aspirin or antiplatelet use recommendations cited above.

Questions for the Committee:

If the developer provided updated evidence for this measure:

- The evidence provided by the developer is updated, directionally the same compared to that for the previous NQF review. Does the Committee agree there is no need for repeat discussion and vote on Evidence?
- The updated evidence on the blood pressure component provides conflicting recommendations. Does the Committee agree with the developer's rationale for not applying that evidence to the measure.
- For structure, process, and intermediate outcome measures:
    - What is the relationship of this measure to patient outcomes?
    - How strong is the evidence for this relationship?

- Is the evidence directly applicable to the process of care being measured?

Guidance from the Evidence Algorithm

Composite Components #1, 2, 3, 5
Process measure with systematic review (Box 3) → Summary of the QQC provided (Box 4) → Systematic review concludes moderate quality evidence.

The highest possible rating is "High" for Evidence

Composite Component #4
Outcome measure (Box 1) → Empirical data shows relationship between outcome and one healthcare action provided (Box 2) → Empirical data passes for clinical evidence.

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

RATIONALE:

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

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

Maintenance measures – increased emphasis on gap and variation

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

- The developer provided performance gap data at meeting all five components in the composite measure, as well as the individual performance rates of the each of the five components. The developer also was able to show the trends for this measure by reporting year (2006-2018)
- For 2017 dates of service, 44.9% of the patients met all five component targets in the composite measure and considered optimally managed. The trend from 2006 to the 2017 dates of service indicate a gradual upward trend in performance rates (2006-9.5%; 2018-44.9%)
- The developer also provided the performance rates for the individual components:
  - A1c < 8.0- 69%
  - Statin- 88%
  - Blood Pressure < 140/90 -83%
  - Daily Aspirin Use if IVD- 99%
  - Tobacco Non-user -84%

  *Note: Although rate is high for aspirin, similar issue was brought up on last endorsement in 2015 (2015 NQF Endocrine Final Report-Cycle 3)and developer was able to justify that this component is not "topped out" by rates in the literature.
- The developer also indicated variability of performance among clinics. Performance rates varied from 9% to 63.4% with a mean of 44.9%. There were 618 reportable clinics.
- The data provided indicate a performance gap still exists and there is variation in performance amongst clinics.

Disparities

- The developer provided disparities data on race/ethnicity (2014-2018) and age (for year 2014 and 2018). The developer also provide data on type 1 diabetes versus type 2 diabetes (for year 2014 and 2018).

- The data on race/ethnicity indicate lower performance rates for African Americans (33%) and American Indians/Alaska Native (24%) versus Asians (48%) and Whites (47%).
- In addition, the data indicates a higher performance rates with age. In 2018, age 66-75 has a 55.5% performance rate versus age 18 to 25, which has a 29.0% performance rate.

- The developer also indicated there is variability in performance rates between MN Health Care Programs (32.5%) and Other Purchasers (47.6%)
- In addition, the developer cited the following literature that supports the disparities data.
  - Knowler WC. N Engl J Med 346(6):393-403, 2002.
    - African American, Asian or Pacific Islander, American Indian or Hispanic/Latino American populations are at greater risk for developing diabetes, and these populations are also growing. In 2000, roughly one in every eight (12 percent) of Minnesota's nearly five million people were Persons of Color or American Indians; by 2025, that proportion will be 17 percent, or nearly one in every five

Questions for the Committee:

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

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

RATIONALE:

1c. Composite – Quality Construct and Rationale

Maintenance measures – same emphasis on quality construct and rationale as for new measures.

1c. Composite Quality Construct and Rationale. The quality construct and rationale should be explicitly articulated and logical; a description of how the aggregation and weighting of the components is consistent with the quality construct and rationale also should be explicitly articulated and logical.
[Quality construct]

- The composite measure is an "all-or-none" measure
- The quality construct is an all-or-none composite with the goal for the patient to best reduce their overall risk of developing long term complications by targeting all five components (blood pressure control, blood sugar control, tobacco-free patient, statin use and daily aspirin or anti-platelet use as appropriate)
- Numerator is calculated at the patient level and numerator compliance is defined as the patient achieving all five components of the measure. The components are treated equally; there is no weighting. Some of the components have an exception methodology within allowing a "free-pass" on the component if it does not apply to the patient.

Questions for the Committee:

- Are the quality construct and a rationale for the composite explicitly stated and logical?
- Is the method for aggregation and weighting of the components explicitly stated and logical?

Preliminary rating for composite quality construct and rationale:
☒ High ☐ Moderate ☐ Low ☐ Insufficient

RATIONALE:

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

1a. Evidence

Comments:

**Yes the evidence relates well, is applied directly and is current with ADA Standards 2018 and 2019

**new evidence speaks to successful outcomes for patients adherent to the measure

**This composite has both process and outcome measures that relate to improved overall care of patients with diabetes. There is good evidence for each of the outcome measures (A1C control <8%, BP control (<140/90) to improve health of patients with diabetes. Likewise with process measures of statin use, aspirin use if IVC and tobacco cessation, evidence exists to support improvement in outcomes. Not aware of new evidence, but aware of controversy surrounding reduced BP goals.

**evidence directly supports the measure

**Agree that there is no need to repeat discussion and vote on evidence. Agree with developer's rationale for not applying conflicting BP information to the measure.

**I am aware of no studies that measure the relationship between the provision of ALL the components of this composite measure and any outcomes whatsoever. The evidence cited by the developer relates to the association of the INDIVIDUAL components of this composite measure with outcomes. As an aside, I believe those components are not as strongly associated with the outcomes cited as claimed in the review.

**Evidence for the five measures has been updated. The selection of sources and hesitation to change the current BP goals seem appropriate at this time.

**the evidence relates directly to the intermediate outcome being measured by this composite measure. The evidence for all five of the indicates that achievement of the goals improve outcome. I am unaware of new studies/information that changes the evidence base that has not already been cited in the submission.

**no need for review

**Individuals with diabetes who achieve all 5 components in this measure are more likely to have a normal life span. The measure applies directly to the outcome being measured. The process which requires meeting the five components of the measure lessens the chance of microvascular and macrovascular complications of diabetes. I am not aware of any new studies that change the evidence base for this measure that has not been cited in the submission.

**There is a direct relationship. Aditionally, I agree with the developers regarding no addition of ASA for patients without established ASCVD, and with continued definition of high blood pressure requiring treatment as 140/90.

**The evidence directly applies to the process and outcome being measured. For some components of the measure, new studies and evidence that supported the focus areas being measures were discussed.

**The evidence provided supports the linkage to outcomes that patients would care about (microvascular and macrovascular complications).

1b. Performance Gap

Comments:

**Yes performance data present and it demonstrated a gap of care with opportunity of improvement and variation. Race/Ethnicity + Age also showed opportunity for improvement and variation.

**the aspirin use seems at a max, despite the evidence they presented. There is significant variation in terms of success on the meaure.

**Aspirin use is very high, but the other measures can be improved. There is noted variability between practices, age and race/ethnicity.

**High variability. Disparities presented by age, and ethnic group.

**yes, performance gap and disparities exist. A national performance measure is needed.

**Performance on a poor measure does not warrant a national performance measure. The finding of disparities provides an important additional argument against the measure, since there may be an inducement for those providing services to discriminate to avoid treating subgroups with lower

achievement of any components of the measure, thereby making it more difficult to achieve the composite measure.

**Performance has improved over the years of this study but a significant gap from the ideal remains. Population subgroup data was provided.

**Current performance data was provided for 618 facilities in Minnesota. Although since 2007, there has been improvement in overall performance from 9% in 2014 to ~40% in 2017,, there is variability in performance among the clinics and even a 40% performance rate leaves much to be improved. Working toward achieving all five goals would greatly improve the health of our population and so a national performance measure is justified. Data on disparities were reported for age and ethnicity. Performance was lowest for Native Americans (including Alaskans) at ~23% and African Americans at ~ 33%, higher for Asians and whites. Regarding Type 2 diabetes, better performance was achieved in older age groups (66 - 75 @ 55.5%) than younger (18 - 25 @ 29%) in 2018

**compliance was 9% in 2006 and is now 44.9%, which is much improved, but still much room for improvement, especially when age disparity is considered (better results in older population)

**Current performance data on the measure was provided. Performance by the 618 clinics showed improvement, i.e., in 2006 9.5% of the patients met target and in 2017 44.9% met target. There was considerable variability among clinics in 2017. The range of patients meeting target was from 9% to 63.4%. So there is still a gap in care. To the extent that a national performance measure would focus attention on the large population of individuals with diabetes such a measure would be useful. Data was provided regarding disparities on population subgroups. Lower performance rates on this measure was noted in African Americans 33% and Native Americans (Indian and Alaskan) 24% vs. Asians 48% and Whites 47%. Disparities also noted with age group: 66-75y 55.5%; 18-25y 29%.

**Current performance data is provided. It demonstrates a substantial gap in care and among women of differences races / ethnic groups.

**There is still a gap in care with some components performing higher than other areas. The composite measure resut for 2017 were 44.9% indicating that there is still a gap in care and an opportunity for improvement (less than optimal care). The data also suggest disparities for African Amerian and American Indians/Alaska Natives compared to Asians and Whites. There were also age disparities with those under age 66 years performing significantly lower than the older age group.

**There is a gap in care, with respect to attainment of the composite

1c. Composite Performance Measure-Quality Construct

Comments:

**Yes they are all evidence by the 2019 ADA Standards of Care

**Yes

**For this measure, each of the areas above are stated and meaningful

**yes.

**yes. adequate construct of this composite performance measure

**I am not aware of a logical construct that supports the argument that this composite measure adds value beyond the value already in place with measures that address the individual components of this composite measure.

**Yes

**As stated all of the concerns are stated in a logical manner

**yes, weighting is equal for all individual components

**All of the above are stated clearly and in a logical manner.

**The quality construct is stated and is logical.

**The quality construct for the diabetes composite measure is high and indicates that targeting all five componenets may reduce oveall risk of developing long-term complications for those diagnosed with diabetes.

**Yes, to all of the above questions

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

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

Reliability

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

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

Validity

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

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

Composite measures only:

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

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

Evaluators: NQF Scientific Methods Panel Subgroup

Methods Panel Review (Combined)

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

Scientific Methods Panel Votes: Measure passes

- Reliability: H-4, M-0, L-0, I-1
- Validity: H-2, M-3, L-0, I-0
- Composite: H-1, M-4, L-0, I-0

This measure was reviewed by the Methods Panel. A summary of the measure is provided below:

Reliability

- Score-level:
    - Reliability testing done at the measure score level using the Adams' formula (signal to noise)
        - Results indicated "very good" reliability score:
        - On 2014 data, the reliability by this methodology was .908. This was repeated in 2018 for measure maintenance and the reliability was .888. The latter was performed on over 300,000 patients in 618 clinics. Reliability correlated with the number of eligible patients at the clinic with a range of .519 for those with 30 patients (the minimum per measure standard) to .994 for the largest clinics.

Validity

- Data element:
    - Data elements were validated with audit and quality checks
        - In 2018, for the diabetes measure, MNCM audited 53 medical groups; 37% of those submitting data. 89% passed the initial audit, 11% required a correction plan and all re-submitted their data and passed the audit with > 90% accuracy.
- Score-level:
    - Measure score validity was done based on clinical supposition that performance on Optimal Vascular Care measure would be similar; results of correlation testing indicated "fairly strong correlation"

Standing Committee Action Item(s):

- Standing Committee to consider if clinical logic of the composite and change in construction are appropriate.

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?

Questions for the Committee regarding composite construction:

- Do you have any concerns regarding the composite construction approach (e.g., do the component measures fit the quality construct and add value to the overall composite? Are the aggregation and weighting rules consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible?)?
- The Scientific Methods Panel is satisfied with the composite construction. Does the Committee think there is a need to discuss and/or vote on the composite construction approach?

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

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

Preliminary rating for composite construction: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Scientific Acceptability: Preliminary Analysis Form

*Note: Completed by multiple Scientific Methods Panel members and therefore multiple responses provided.

Measure Number: 0729

Measure Title: Optimal Diabetes Care

Type of measure:

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

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

Data Source:

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

Level of Analysis:

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

Measure is:

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

RELIABILITY: SPECIFICATIONS

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

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

   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.*

   Panel Member #1: The population is clearly defined. For the numerator of this composite measure, all five subcriteria must be achieved in order to be counted as having met the numerator. Analysis can be performed at the individual criteria level, but it is not the basis of the score. The denominator includes appropriate exclusions and is clearly specified

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

   Panel Member #1: I have no concerns.

   Panel Member #2: None.

   Panel Member #3: This composite measure requires information on 5 different aspects of care for patients with diabetes. This is good because it sets an ambitious target. However, it also leads to detailed and complex specifications, creating room for error, particularly in settings with more limited resources/staffing. As the authors point out, the shift to EHR based reporting is important to making this measure usable over time

RELIABILITY: TESTING

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

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

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

   Submission document: Testing attachment, section 2a2.2

   Panel Member #1: The beta-binomial method was utilized by the developer because the composite score used as the measure numerator is a binary data point, an all-or-none score. On 2014 data, the reliability by this methodology was .908. This was repeated in 2018 for measure maintenance and the reliability was .888. The latter was performed on over 300,000 patients in 618 clinics. Reliability correlated with the number of eligible patients at the clinic with a range of .519 for those with 30 patients (the minimum per measure standard) to .994 for the largest clinics.

   Panel Member #2: Remain concerned that the denominator does not include within patient variance, only within clinic variance. Also n of providers by clinic may add variance to denominator.

   Panel Member #3: The authors used the Adams method of assessing reliability at the provider level. I am assuming provider means the 600+ clinics in the test data, but this is not totally clear in the wright up. This approach seems relevant given the goal of the measure.

   Panel Member #4: Beta binomial (yes/no outcome)

   Panel Member #5: Signal to noise via BETABIN routine. Score high.
   Individual components correlated.

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

   Submission document: Testing attachment, section 2a2.3

   Panel Member #1: The results above indicate an overall high level of reliability to the measure score, especially in larger clinics

   Panel Member #2: See #6.

   Panel Member #3: The reliability score has gone down slightly since the original submission (0.888 versus 0.908), but is still quite high. This makes sense given the high reliance on electronic health records in the test data. It is impressive that so many clinics have structured data elements in their EHR that they can complete the 5 component measures without reading notes.

   Panel Member #4: 0.89 – good reliability

   Panel Member #5: Acceptable level of reliability demonstrated.

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

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

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

Submission document: Testing attachment, section 2a2.2

☒ **Yes**

☐ **No**

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

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

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

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

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

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

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

    Panel Member #1: The reliability has been tested on two different occasions in large populations and clinics and demonstrates not only a high level overall, but also shows practice variation in reliability of the measure score in relation to number of patients reported. Appropriate methodology was utilized for this all-or-nothing score.

    Panel Member #2: Am not clear how between patient level variation across measure was handled.

    Panel Member #3: High reliability score; stable pattern over time (2014 and 2018 samples).

    Would be good to see data for outside of Minnesota – I do wonder if reliability would go down with greater reliance on paper records.

    Panel Member #4: Score level average reliability 0.89. Reliability from 2014 review was 0.91.

    Panel Member #5: No concerns about the reliability. Data basically complete. Find the method of dealing with missing data for individual items acceptable. Correlation of individual score items and overall score acceptable. Measure shows large variance in performance.

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: Measure exclusions are specifically outlined and are appropriate to the measure.

    Panel Member #2: None.

    Panel Member #3: Exclusions make sense.

    Panel Member #5: None. But clinicians on substantive committee should review.

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

    Submission document: Testing attachment, section 2b4.

    Panel Member #1: The testing demonstrates performance differences between practices with high volumes and those with low volumes. Performance on the measure has improved over the last few years but still has not exceeded 40%. The gap between those in the Minnesota Health Care Program and others is noted in the submission.

    Panel Member #2: None.

Panel Member #5: There is large variance in performance on the measure. Ranks relatively stable with and without SES risk adjustment.

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: I have no concerns. The source data is either electronic health records or paper-based data abstraction. Data is not provided about any testing differences between these, as an EMR has been widely adopted across the reporting locations.

Panel Member #3: There is so little paper based information in the test data it is hard to say if there are systematic differences.

Panel Member #5: Almost all clinics in MN reporting from EHR. Use of sample for small number of clinics on paper records is acceptable by past standards.

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

Submission document: Testing attachment, section 2b6.

Panel Member #1: None. Missing data elements are treated as a failure to the measure (they are kept in the denominator). The rate is known for the various components and varies between 0.003% and 0.9%)
Panel Member #5: None.

16. **Risk Adjustment**

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

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

☐ Yes ☐ No ☒ Not applicable

16c. **Social risk adjustment:**

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

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

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

16d.**Risk adjustment summary:**

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

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

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

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

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

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

Panel Member #1: As the TEP pointed out, the most controversial aspect of the risk adjustment model are the use of the deprivation index for socio-economic status, and the use of insurance product for socio-economic status. The former is a continuous variable based on 5-digit zip code and the attainment of SNAP benefits. The latter is stratified as commercial, Medicare, Medicaid, uninsured, and unknown. There was consideration given to the RELO factors, but a conscious decision was made not to use these.

The only point of discussion is about the use of the risk adjustment model. A conscious decision was made about the one chosen and it data provided to demonstrate validity. A correlation was calculated for performance for each of the risk variables, and its impact on measure score and ranking. It is stated that the risk adjustment approach does demonstrate differences between the clinics.

Panel Member #2: Adequate.

Panel Member #3: The model is parsimonious; the deprivation index is a clever way to handle social risk factors; the results showing that the deprivation index raised the expected rates in wealthy communities and lowered it in more rural/inner-city areas were compelling.

Panel Member #4: Developer appears to be using Pearson correlation coefficient for binary data to assess multicollinearity (p 23). Not the appropriate method for testing

Panel Member #5: Risk adjustment uses age, diabetes type, insurance status (a potential measure of access to other services needed to achieve adherence to care or achievement of control), and zip code level social deprivation index (a potential measure access to services and SES related pressures on adherence and success of therapy).

Consideration was given to individual level SES measures, but this was rejected based on expert panel judgment that lack of success due to those factors was intertwined with performance of practice.

No consideration appears to have been given to other individual risk adjusters, notably gender. Would have liked some discussion of this.

Risk adjustment works as expected, generally narrowing differences. This is true for area deprivation index, and there is little movement across classification below, at or above average when this variable is added to risk adjustment model.

Generally satisfied with risk adjustment model.

Need to develop methods to allow census tract rather than zip code level area measures to be used in risk adjustment.

VALIDITY: TESTING

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

    Submission document: Testing attachment, section 2b2.2

    Panel Member #1: Although data is provided about the auditing and training process for data accuracy and "validation", the methodology utilized for validity testing on the measure is another standard, the optimum vascular care measure. The rationale is that they would many complications from poor diabetic control are vascular complications and thus they should be highly correlated. I am not able to assess related as a measure of clinical practice. Correlation analysis between the ODC (this approach in a clinical sense.measure) and the OVC (measure NQF#0076) had an r^2 of 0.629.

    Validity testing at the component level was performed by the CORR procedure. Pearson coefficient values were 0.26253 and 0.77714, with all but one above 0.5. The lowest value was the aspirin or anti-platelet use component.

    Panel Member #2: Correlated Optimal Diabetes Care and its components with Optimal Vascular Care.

    Panel Member #3: The measure developers audit practices for denominator certification, data quality checks, validation audit, and the two-week medical group review period. Between the two years (2014 and 2018), correction action was required in 11-15 percent of provider groups.

    For the composite measure they looked at the correlation between diabetes scores and Optimal Vascular Care – more, correlated measures would have been helpful.

    Individual measures were also validated by looking at the correlations – confirmatory factor analysis or similar technique would have been helpful.

Panel Member #4: Tested Correlation with optimal vascular Care (NQF 0076). No literature to support the clinical supposition. Critical values assessed via audit.

Panel Member #5: Accuracy of data submitted tested. Correlation with optimal vascular Care (NQF 0076). No literature to support the clinical supposition composite measure

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

Submission document: Testing attachment, section 2b2.3

Panel Member #1: The testing of association with another endorsed NQF measure is an interesting approach, but probably valid for diabetes care, given the strong clinical association. The component testing was more standardized with the results above.

Panel Member #2: Lowest association is for aspirin/antiplatelet user =.263 – is this with composite ODC or with OVC? Are values correlations (as stated) or R^2 (variance) as presented?

Panel Member #3: Results are modest – I'm not sure what the correct relationship is between Optimal Vascular Care and Optimal Diabetes Care, but a correlation of 0.629 does not seem very strong. This is partly why additional measures would be helpful (both associated and non-associated measures). Similarly, the individual scores have correlations that range between 0.26 and 0.78 – without more testing, its hard to know if this makes sense.

Panel Member #5: Data capture appears to be highly effective. Correlation with OVC good (r-square 0.6). Examination of scatter plot shows no outliers or unusual patterns.

Panel Member #4: Report r^2 of 0.629

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

Submission document: Testing attachment, section 2b1.

☒ **Yes**

☐ **No**

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

22. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?**

NOTE that data element validation from the literature is acceptable.

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

☒ **Yes**

☐ **No**

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

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

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

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

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

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

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

Panel Member #1 The steering committee pointed out concerns about the measure, but once assumptions are accepted, the methodology is appropriate.

Panel Member #2: Lower sensitivity rates and variation by state is a concern.

Panel Member #3: In general, more testing would help make a stronger case for the validity of this measure. Would like to know more about the relationship between the data elements – do they load on a single factor, for example. As it stands, there is some good evidence suggesting a moderate rating.

Panel Member #5: No, concerns about data quality or construction of the composite for an all or nothing measure. Experts on substantive committee should review clinical logic, particularly the addition of the statin measure and dropping of LDL management.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

25. **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**

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

Panel Member #1: See above.

Panel Member #2: Reflect multi-dimensional nature of diabetes care.

Panel Member #3: As stated above, the correlation pattern among the individual measures is hard to interpret without more information. That said, the authors make a strong clinical/public health case for requiring compliance on all 5 aspects of diabetes care. This is reflected in the data on the steady improvement in rates (approximately 9% to over 30% compliance) over the course of data collection.

Panel Member #4: All but one component (aspirin use) showed high correlation with the composite.

Panel Member #5: Correlation of components

ADDITIONAL RECOMMENDATIONS

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

Panel Member #2: Is the all-or-none strategy the best reflection of quality of diabetes care given that the rates in MN were 44.9%. Would a composite index of the components yield a different distribution of clinics? The contribution of the aspirin component to the ODC is low (r=.26) probably because it has very little variability. Why include it?

Panel Member #5: Substantive experts in diabetes need to review clinical logic of the composite and change in construction.

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

2a1. Reliability-Specification

Comments:

** all data elements clearly defined, codes with descriptors provided. Initially I was concerned about the tobacco use, but after reviewing the sample data they were able show a high percentage of tobacco free goal achieved and therefore a good proportion achieved the composite goals.

** Reliability is very high.

** The specs are well defined, and there are no concerns about implementation as written.

**No concerns

**no concerns

**Although I do not a favor the continuation of this measure for the reasons given, I have no concerns about the likelihood that it can be consistently

**Data elements appear to all be clearly defined and consistent with the current coding system.

**The Scientific Methods Panel voted a "PASS". Reliability assessed by the Adam's formula: in 2014 - .908, in 2018- .888. It was also noted that the reliability correlated with the number of eligible patients at the clinic: .519 for clinics with 30 patients, .994 for the clinics with the largest numbers enrolled. More likely to be consistently implemented when a robust EHR is used

**very good reliability results

**The Scientific Methods Panel voted a "PASS". Reliability assessed by the Adam's formula: in 2014 - .908, in 2018- .888. It was also noted that the reliability correlated with the number of eligible patients at the clinic: .519 for clinics with 30 patients, .994 for the clinics with the largest numbers enrolled. More likely to be consistently implemented when a robust EHR is used

** Specifications: I have no concerns, but I note that one methods panel member is concerned that the denominator does not include within patient variance.. I think there is no need to vote on reliability or validity. I did not see discussion of risk adjustment. The component measures do fit the quality construct

** No concerns with the reliability of the specifications.

** No concerns

2a2. Reliability-Testing

Comments:

** Not at this time

** no

** Reliability is reported and shown to be high. No concerns.

** none

**no concerns

** No

** No

** No

**none

** No

**It is not clear that all EHR practices will have access the needed data elements.

**No

**No concerns

2b1. Validity-Testing

Comments:

**Not at this time

**no

** No - validity audits showed moderately high scores.

**None

**no concerns

**No

** No. Annual audits of a sample of clinics, exclusions analysis, and the comparison of clinics' scores for diabetes care and vascular care are described.

**No

**none

**No

** My only concern is Tested Correlation with optimal vascular Care (NQF 0076). Presumably these should be highly correlated. Is there literature to support the clinical supposition.

**No concerns

2b4-7. Threats to Validity

2b.4. Meaningful Differences

Comments:

** They changed the denominator inclusion criteria and switched to a better more inclusive denominator identification and showed the results of the test.

** There is both large variation, and all with significant room for improvement. It speaks to the challenge of achieving success on this measure, which does weaken the validity of the measure.

** Majority of data is from electronic records -data source issue should be of minimal concern. There are meaningful differences in quality identified, which may be narrowing with time per data presented. Missing data, treated as failure to measure/complete is small and should not be a validity threat

** no concerns

**No concerns

** No identified threat to validity of the individual components of the measure. I am aware of no measures of validity of the composite measure, although there is no reason to question validity since the composite measure is based on components with no identified important threats to validity.

** No significant threats to validity seem apparent.

** 2b4. higher performance ratio, better survival with fewer complications . 2b5 there are multiple sets of specifications as there are 5 components of this composite measure. Although to meet the standard all 5 components must be achieved there is value in looking at performance for each individually. Use of aspirin in patients with vascular disease has the highest performance score. 2b6. Missing data certainly affects the validity.

** apparently slight concern of differences in results between EHR query versus manual extraction versus paper records, but not sufficient to be of real concern

** No. For this performance measure, missing data percent is 0.03 to 0.9%. Also missing data for a component of this measure = negative and remains to be included in the denominator.

** 2b.6 There was no distinction between a negative and missing data which may constitute a threat to validity.A lack of EHR documentation at one of the three testing sites raises concerns with the measure's validity and feasibility.

** No concerns with the validity or threats to validity. Because some of the data comes from chart review, moving it towards an electronic version would increase the validity and the value of the measure.

** No concerns

2b2-3. Other Threats to Validity

2b2. Exclusions

2b3. Risk Adjustment

Comments:

** Exclusions are consistent with evidence.

** no.

** Risk adjustment is satisfactory

** no concerns

** no concerns

** Unsure how this might apply -- I need to look into this more.

** Exclusions are reasonable and limited. It appears that the study leadership has struggled appropriately to try to legimitately study socioeconomic factors as applied to health outcomes. Credit to them to continue to work on this aspect of the study.

** 2b2. exclusions are consistent with evidence. No groups are inappropriately excluded. 2b3. the risk-adjustment variablesw ere present at the start of care. Social risk factor variables were available and have been analyzed with regard to the conceptual description provided by the developers.

** no concern

** Exclusions are appropriate. There is a conceptual relationship between poential social isk factor variables and the measure focus. The Social risk factor variables that were analyzed align with the conceptual description provided. The risk-adjustment variables were present at the start of care. There was some discussion of the development of the case-mix adjustment among the Scientific Methods Panel members. From my point of view, the risk adjustment was appropriately developed and tested. The analyses show acceptable results. An appropriate risk-adjustment strategy is included in the measure.

** Exclusions are appropriate. Panel members comments are helpful: the reliance on the use of the deprivation index for socio-economic status is convenient but may not provide the most accurate results. I do not have a better suggestion.

** The risk adjustment was tested and appeared to be appropriate, though limited in the number of clinics/providers.

** No concerns

2c. Composite Performance Measure - Composite Analysis

Comments:

**yes

** Unclear if value is added, given the challenges in meeting this all or none measure.

** This composite captures important components of diabetes care. The aspirin process measure has little variability, and may have little to add to the measure

** My only question is does aspirin use contribute enough to continue as an element.

** yes. Fits the quality construct and rationale for a composite measure

** I am aware of no evidence that the composite measure adds value beyond that of the components individually.

** The composite analysis in addition to the analysis of each individual goal appears to be sensible and helpful in this setting.

** Yes to all parts of question 11

** component measures all appropriate

** The authors make a strong case for requiring compliance on (at least 4, other than ASA) components of the measure.

** Yes and Yes.

** Yes for both questions

Criterion 3. Feasibility

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

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

Data Specifications and Elements

- The measure is constructed using Electronic Medical Records (EMR) or Paper Medical Records. The developer provided the following information from 2018:
    - 71% (476) clinics had an EMR and pulled all data via query
    - 26% (176) clinics had an EMR and used a combination of query and manual look up for data collection
    - 2.2% (15) clinics had an EMR and looked up all data manually
    - 0.15% (1) clinic had a hybrid EMR and paper record system
    - 0.15% (1) clinic had paper records only

    *Feasibility Note: 71% of practices can extract all of the information needed via query.
- Some data elements are in defined fields in electronic sources.
    - Although most data field are available in EHR, the developer noted that some data fields could be built into the EHR but may not be readily available unless a discrete field is built (e.g. some exceptions to statin such as breast feeding or women of child bearing years not actively taking birth control or aspirin/antiplatelet use).
- This measure is not currently an eMeasure. The developer would be interested to specify the measure as an eMeasure if funding available.
- There are no fees associated with participation and submitting data for this measure. However the developer noted there are costs to the medical groups in terms of extract programs or abstraction to submit patient level clinical information for rate calculation.

Questions for the Committee:

- Are the required data elements routinely generated and used during care delivery?
- Are the required data elements available in electronic form, e.g., EHR or other electronic sources?
- Is the data collection strategy ready to be put into operational use?

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

RATIONALE:

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. Feasibility

Comments:

** documentation of tobacco use may or may not be check off in an EHR. Yes and no answer as there is not a standardized process for free form notes if tobacco cessation is provided by an ancillary health provider.

** The statin exceptions are more challenging to note, and would be cumbersome to build in to easily measure in an EMR. The other pieces are readily available in EMR.

** The composite measures should be easily obtained via paper chart review or emr extraction.

** Majority of elements in ehr.

** some data elements-if not available in EHR may add additional burden to collect.

** No concerns.

** The adoption of EHRs by Minnesota medical care providers and the ability of all their systems to supply most of the data elements needed in this project is impressive.

** This measure is widely used in minnesota. The required data elements are routinely generated and used during care delivery and for almost all clinics available in electronic form

** none; in some cases, EHR may need to add specific data fields which are consistent with the measure

** All of the required data elements are routinely generated and used during care delivery. All data elements are available in electronic form. This measure is currently in operational use

** Can we anticipate that all clinics will have as many of the data elements available in the EHR?

** Is currently a feasible measure, that will improve as use of electronic health records and data extraction becomes more consistent and reliable.

** No concerns

Criterion 4: Usability and Use

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

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

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

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

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

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

OR

Planned use in an accountability program? ☐ Yes ☐ No

Accountability program details

The measure is in four accountability programs:
1. MN Bridges to Excellence
Pay for Performance Program for top performers and those attaining improvement goals/ benchmarks.
2. MN Community Measurement- MN HealthScores Website
Public Reporting consumer-facing website for all primary care and endocrinology clinics in Minnesota (mandatory) and bordering communities (voluntary)
3. MN Community Measurement- Health Care Quality Report
Public Reporting: Hard-copy report (pdf) highlighting top performers, most improved
4. MN Department of Health- Statewide Quality Reporting and Measurement System
This program requires mandatory submission of data from Minnesota physician clinics that have provider specialties that are applicable to the measured population.

- For the Optimal Diabetes Measure, the provider specialties are: family medicine, general practice, internal medicine, geriatric medicine and endocrinology.

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

- Performance results are provided to all medical groups who submit data for this state-wide measure via several options:
    - Preliminary measure rates are provided immediately after file upload to HIPAA secure data portal
    - A two-week review process is conducted to allow groups to review and potentially appeal prior to public reporting of rates
    - Rates are reported by medical group and clinic level on public website MN Healthscores
    - Rates including all historical rates can be obtained from the MNCM data portal
- The 2018 Medical Group Survey (this is an annual survey conducted) provided the following feedback on the value of the measure and if it was easy or difficult to obtain data needed for submission for this measure. See results below which indicate clinics find the measure to have high to moderate value. However, almost 40% found it difficult to obtain data needed.
    - To what degree does your medical group find value in the measure? (n = 112)
        - High Value 45.5% (51)
        - Moderate Value 33.0% (37)
        - Minimal Value 14.3% (16)
        - No Value 7.1% (8)
    - How easy or difficult is it to obtain the data needed for DDS submission for this measure? (n = 104)
        - Very Easy 19.2% (20)
        - Easy 44.2% (46)

- Difficult 29.8% (31)
- Very Difficult 8.7% (9)

Additional Feedback:

- The developer shared addition feedback processes [feedback loops]
  - MNCM's Measure Review Committee (MRC) is tasked with the annual review of all publicly reported measures on MN Healthscores, including this measure. In 2018, the Committee voted unanimously to continue the measure without changes.
  - The MN Department of Health conducts several comment periods during its annual rule making process.
  - MNCM provides a year-round staffed support through a helpline via phone or email.
  - MNCM conducts an annual medical group survey for all clinics in the state

- The Optimal Diabetes Care measure was on the 2017 Measures Under Consideration list for CMS and it was conditionally supported for use by NQF Measures Application Partnership (MAP) Clinician and MAP Rural Health Workgroups.
  - In the 2018 MAP Clinician Final Report, MAP Clinician Workgroup stressed the importance of this composite measure but also acknowledged the utility of the individual subcomponents of the measure to drive quality improvemen. MAP conditionally supported MUC17-181 based on the condition that there are no competing measures in the program and that the measure is updated to the most current clinical guidelines. In addition, MAP recommended risk stratification be appropriately applied to the measure.
  - In the 2018 MAP Rural Final Report, MAP Rural Workgroup recommended that the measure only be used for quality or population health improvement and not for payment adjustment.

***Questions for the Committee**:*

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

Preliminary rating for Use: ☒ **Pass** ☐ **No Pass**

RATIONALE:

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

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

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

Improvement results

- The developer provided performance gap data at meeting all five components in the composite measure, as well as the individual performance rates of the each of the five components. The developer also was able to show the trends for this measure by reporting year (2006-2018)
  - For 2017 dates of service, 44.9% of the patients met all five component targets in the composite measure and considered optimally managed. The trend from 2006 to the 2017 dates of service indicate a gradual upward trend in performance rates (2006-9.5%; 2018-44.9%)
- The developer also provided the performance rates for the individual components:

- A1c < 8.0- 69%
- Statin- 88%
- Blood Pressure < 140/90 -83%
- Daily Aspirin Use if IVD- 99%
- Tobacco Non-user -84%

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

- Developer indicated that adults age 65 and older have better outcome rates than their younger counterparts with diabetes. The developer noted this could be related to compliance with providers orders.
- Developer noted that the Minnesota statewide averages of A1c values are drifting upward, which is a trend also indicated by the American Diabetes Association.
- A positive finding of the measure is that there is a statewide focus in Minnesota for this measure which is used by many health systems.

Potential harms

- No additional harms noted apart from the potential clinicial adverse events noted in the Evidence for the individual components of the composite measure.

Additional Feedback:

N/A

***Questions for the Committee**:*

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

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

RATIONALE:

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

4a1. Use-Accountability and Transparency

Comments:

** We reviewed their initial composite measure application they have made a great deal of good changes and addressed the fixes and changes in the ADA standards and Cholesterol guidelines.

** It is being used, with most favoring its value, as demonstrated by its continued use.

** Feedback is provided to participating practices, and also to the public. Feedback is solicited from public and surveyed medical groups.

** Results are publicly reported and feedback of results to those being assessed. the majority of whom find it useful.

** no concerns

** Not sure.

** Reporting of results to participating clinics and to the public are well organized. Feedback is solicted from all stakeholders.

** 4a1. The measure is publicly reported in four ways in Minnesota. The performance rates are available to the public and to the providing clinics. The information is used in some pay for performance settings. One of the reporting sites not only give performance rates but also notes improvement. There is a mechanism for informing the clinicians of their results in a timely fashion and there is opportunity for advice and remediation. Feedback has been considered by the developers when changes are incorporated into the measure. The state of Minnesota requires that all clinics, physician groups who represent specialties that care for people with diabetes participate in this measure

** seems to be used widely across MN with consideration in CMS and NQF programs

** The measure is publicly reported in Minnesota in 4 ways: 1) Pay for performance for top performers. 2) On the Minnesota Health Scores Website. 3) Healthcare Quality Report is available as a pdf listing the top performers and the most improved. 4) Minnesota statewide report for all physician groups and clinics that have provider specialties applicable to the measured population. With regard to 4a2. Use. All questions can be answered by YES.

** The measure is currently being used in 4 accountability programs, but all 4 are in MN. Feedback consistently supports use of the measure, though some clinics find it difficult to obtain data.

** Yes. There is accountability and transparency and those involved in reporting the measure have been provided an opportunit for feedback.

** The results are publicly disclosed

4b1. Usability-Improvement

Comments:

**My concern is that with a composite that all or none you do not give positive feedback to the professional that are working toward goals with their patients. With applying SDOH there may be groups of their population under-reporting or not reporting tobacco (or other types of smoking i.e. vaping and other substances

**As with other measures that are outcome based, it runs the risk of adverse patient selection, especially against smokers.

**Difficult to foresee unintended consequences, but benefits are significant for improving each of the measures.

**no concerns

**no concerns

**I know of no credible rationale to support the argument that the composite furthers the goals noted beyond the value of the individual components. I am concerned that the composite measure might further disparate treatment of those with an undue burden related to social determinants of health whereby providers of medical services might be less inclined to serve those groups for whom social determinants make achievement of the composite measure more challenging. Given no identified additional benefit provided by the composite measure and the possibility of unintended harm, I believe the potential harms outweigh the benefits. I therefore do not support the continuation of the composite measure.

**Improvements in the Minnesota diabetic population's health have been published. This project may have contributed to the improvements. No perceived harms.

**4b1. Clinicians should seek improved performance scores, patients with diabetes will likely seek care from providers with better scores. 4b2. Harms related to actual unintended consequences may include cost of more medication, side effects from prescribed drugs, bleeding from aspirin use. Benefits however outweigh

these risks. Clinicians must continue to use good clinical judgement in their efforts to achieve the benefit of meeting the components of this measure

**no substantial harms; seems reasonable to use individual component scores as well as the 'all or none' composite score;

**4b1 Usability- high rates of performance will likely result in better population health regarding diabetes and avoidance of complications. 4b2- benefits of this measure outweigh the potential harms. Healthcare professionals will need to modify application of the goals of this measure depending on medication tolerance, cost, lifestyle etc. but that is what constitutes responsible healthcare.

**The measure appears to be useful in those clinics currently employing it. Again, I am concerned that there is no use outside MN.

**Not necessarily unintended consequences, however, some of the medications used do have consequences or less than optimal results. However, the benefit appears to outweigh the risks. Optimal diabetes care achives the goal of high-quality, efficient healthcare for individuals and overall for those diagnosed with diabetes.

**The performance results can be used to better inform care coordinate across various clinicians and sites of care

Criterion 5: Related and Competing Measures

Related or competing measures

0061 : Comprehensive Diabetes Care: Blood Pressure Control (<140/90 mm Hg)

0575 : Comprehensive Diabetes Care: Hemoglobin A1c (HbA1c) Control (<8.0%)

2712 : Statin Use in Persons with Diabetes

Harmonization

Per developer, the measures are not harmonized dud to differences in data sources, different composite measure construct and input from measure development workgroups.

- #0061 and 0575 are part of another composite measure by NCQA, which has a different measure construct than #0729. NCQA composite is calculated at the physician level whereas 0729 is at the patient level.
    - There are also differences in the data sources. #0061 an #0575 use claims, whereas #0729 uses EMR data. There are also variation in the denominator value sets for the measures.
- #2712 is a related (not competing) measure on statin use in diabetes patients however #2712 uses pharmacy claims data source. Thus #2712 identifies only patients on diabetic drug that should be on statins, which is different from #0729 which also targets diabetes patients who are not on diabetes medications. #2712 also does not include the exception that #0729 includes for statin use.

*#0545: Adherence to Statins for Individuals with Diabetes Mellitus is a statin measure for diabetes patients stewarded by CMS. Per developer, there is no exception noted for #0545 as with the statin component of #0729. However, NQF staff noted this measure is no longer NQF endorsed so not identified as related or competing measure.

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

5. Related and Competing

Comments:

**We have other measures that are looking at each of these measure separately. MN will most likely only use their composite in their population.

**There are certainly related measures, but there is no immediate concern for harmonization.

**There are related measures, two of which are components in this composite measure.

**There are some differences between this measure and related ones.

**no concerns

**In a sense the measures that assess the individual components of this composite measure are competing. Harmonization can easily be achieved by withdrawing this composite measure, also fostering the goal of reducing the burden of measurement when harm exceeds benefit.

**Three NQF measures are related as they address components (BP, A1C, statin use) of this composite measure. These do not appear to be competing measures.

**#2712 is related but not competing as it is based on pharmacy claims data for patients on a diabetes drug who ought to be on a statin. #0061 and #0575 are part of another composite measure developed by NCQA. The developer of #0729 notes that these measures are not harmonized because of difference measure construct, different data sources, and variation in data set.

**yes, related measures; no additional steps

**#2712 is related but not competing as it is based on pharmacy claims data for patients on a diabetes drug who ought to be on a statin. #0061 and #0575 are part of another composite measure developed by NCQA. The developer of #0729 notes that these measures are not harmonized because of difference measure construct, different data sources, and variation in data set.

**Related measures use different data sources of different levels of analysis.

**Yes, comprehensive diabetes care (blood pressure control and HbA1C control and statin use in persons with diabetes. The measures are not harmonized.

**No concerns here.

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: January/25/2019

Public Comment

**The American Medical Association (AMA) appreciates the opportunity to comment on Measure 729: Optimal Diabetes Care prior to the Standing Committee's evaluation. The AMA is concerned that the composite does not adequately address the guideline recommendations from the Institute for Clinical Systems Improvement (ICSI) cited in the evidence form as well as the American College of Physicians' guidance statement update on hemoglobin A1c (HbA1c) targets (Qasseem, 2018). Both organizations call for patient-centered individualized HbA1c goals, which are not adequately addressed in the measure specifications or the risk adjustment model (e.g., accounting for comorbidities, hospice). These same concerns also apply to the blood pressure control as it does not balance achievement of these targets with the patient's risk tolerance and clinical factors such as advanced cognitive impairment and multiple co-morbidities (e.g., acute kidney injury or failure). As a result, the AMA asks the Standing Committee to consider whether the measure as specified meets the NQF criteria of evidence and scientific acceptability or whether further refinements are needed prior to re-endorsement.

Reference:

Qaseem A, Wilt TJ, Kansagara D, et al. Hemoglobin A1c targets for glycemic control with pharmacologic therapy for nonpregnant adults with type 2 diabetes mellitus: a guidance statement update from the American College of Physicians. *Ann Intern Med*. 2018;168(8):569-576. doi:10.7326/M17-0939

Support/Non-Support

- One NQF Member did not support as of this date.

Brief Measure Information

NQF #: 0729

Corresponding Measures:

De.2. Measure Title: Optimal Diabetes Care

Co.1.1. Measure Steward: MN Community Measurement

De.3. Brief Description of Measure: The percentage of patients 18-75 years of age who had a diagnosis of type 1 or type 2 diabetes and whose diabetes was optimally managed during the measurement period as defined by achieving ALL of the following:

- HbA1c less than 8.0 mg/dL
- Blood Pressure less than 140/90 mmHg
- On a statin medication, unless allowed contraindications or exceptions are present
- Non-tobacco user
- Patient with ischemic vascular disease is on daily aspirin or anti-platelets, unless allowed contraindications or exceptions are present

Please note that while the all-or-none composite measure is considered to be the gold standard, reflecting best patient outcomes, the individual components may be measured as well. This is particularly helpful in quality improvement efforts to better understand where opportunities exist in moving the patients toward achieving all of the desired outcomes. Please refer to the additional numerator logic provided for each component.

1b.1. Developer Rationale: Measure is a composite; please refer to questions 1c.2., 1c.3., and 1c.4.

S.4. Numerator Statement: The number of patients in the denominator whose diabetes was optimally managed during the measurement period as defined by achieving ALL of the following:

- The most recent HbA1c in the measurement period has a value less than 8.0 mg/dL
- The most recent Blood Pressure in the measurement period has a systolic value of less than 140 mmHg AND a diastolic value of less than 90 mmHg
- On a statin medication, unless allowed contraindications or exceptions are present
- Patient is not a tobacco user
- Patient with ischemic vascular disease (Ischemic Vascular Disease Value Set) is on daily aspirin or anti-platelets, unless allowed contraindications or exceptions are present

S.6. Denominator Statement: Patients ages 18 to 75 with a diagnosis of diabetes (Diabetes Value Set) with any contact during the current or prior measurement period OR had diabetes (Diabetes Value Set) present on an active problem list at any time during the measurement period. Both contacts AND problem list must be queried for diagnosis (Diabetes Value Set).

AND patient has at least one established patient office visit (Established Pt Diabetes & Vasc Value Set) performed or supervised by an eligible provider in an eligible specialty for any reason during the measurement period.

S.8. Denominator Exclusions: Valid allowable exclusions include patients who were a permanent resident of a nursing home, pregnant, died or were in hospice or palliative care during the measurement year.

De.1. Measure Type: Composite

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

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

IF Endorsement Maintenance – Original Endorsement Date: Mar 28, 2011 **Most Recent Endorsement Date:** Jun 30, 2015

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

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

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? This is a composite "all or none" measure calculated at the patient level, each individual patient needs to meet all five component targets to be considered in the numerator. All components are contained within this measure and the measure is not paired with another measure.

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

0729_MNCM_2018_Optimal_Diabetes_Care_evidence_template_7-1.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.

Yes

1a. Evidence (subcriterion 1a)

Component # 1 A1c

Measure Number (*if previously endorsed*)**:** 0729

Measure Title: Optimal Diabetes Care - A1c Control Component

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

Date of Submission: <u>10/30/2018</u>

Instructions

- *Complete 1a.1 and 1a.2 for all measures. If instrument-based measure, complete 1a.3.*
- *Complete **EITHER 1a.2, 1a.3 or 1a.4** as applicable for the type of measure and evidence.*
- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.

- If you are unable to check a box, please highlight or shade the box for your response.
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

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

1a. Evidence to Support the Measure Focus

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

- Outcome: [3] Empirical data demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service. If not available, wide variation in performance can be used as evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.
- For measures derived from patient reports, evidence should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.
- Process measures incorporating Appropriate Use Criteria: See NQF's guidance for evidence for measures, in general; guidance for measures specifically based on clinical practice guidelines apply as well.

Notes

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

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

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

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

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

Outcome

☐ Outcome: Click here to name the health outcome

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

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

☒ Intermediate clinical outcome (*e.g., lab value*): A1c is less than 8.0

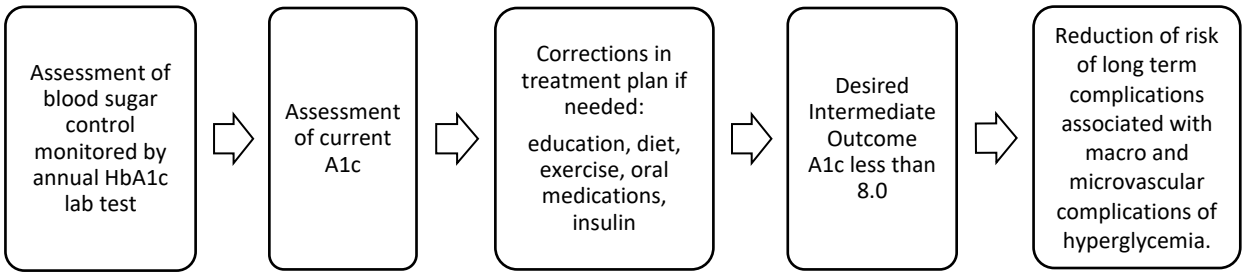
☐ Process: Click here to name what is being measured

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

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

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



Long term complications: blindness, renal failure, amputation

Macrovascular complications: coronary artery disease, peripheral artery disease, stroke

Microvascular complications: diabetic nephropathy, neuropathy and retinopathy

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

Measure component is not derived from patient report.

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

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

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

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

☒ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	**Institute for Clinical Systems Improvement (ICSI) Diabetes Mellitus in Adults, Type 2; Diagnosis and Management of. July 2014** **Redmon B, Caccamo D, Flavin P, Michels R, Myers C, O'Connor P, Roberts J, Setterlund L, Smith S, Sperl-Hillen J. Institute for Clinical Systems Improvement. Diagnosis and Management of Type 2 Diabetes Mellitus in Adults. Updated July 2014. Page 19.**
https://www.icsi.org/guidelines__more/catalog_guidelines_and_more/catalog_guidelines/catalog_endocrine_guidelines/diabetes/	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	**Glycemic Control and A1c Goals Recommendation** A clinician should personalize goals with patients diagnosed with T2DM to achieve glycemic control with a hemoglobin A1c < 7% to < 8% depending on individual patient factors.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	Quality of Evidence: High
Provide all other grades and definitions from the evidence grading system	GRADE Methodology of Evidence Used: High Quality of Evidence ▪ Further research is very unlikely to change our confidence in the estimate of effect Moderate Quality Evidence ▪ Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate ▪ Low Quality Evidence Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate or any estimate of effect is very uncertain
Grade assigned to the **recommendation** with definition of the grade	Strength of Recommendation: Strong
Provide all other grades and definitions from the recommendation grading system	GRADE classifies recommendations as strong or weak. Strength of recommendation is determined by the balance between desirable and undesirable consequences of alternative management strategies, quality of evidence, variability in values and preferences, and resource use. Strong Recommendation ▪ The work group is confident that the desirable effects of adhering to this recommendation outweigh the undesirable effects. This is a strong recommendation for or against. This applies to most patients.

Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	Randomized control trial - 14 Systematic review – 1 Relevant Resources: Hemmingsen, 2013; Callaghan, 2012; Anderson, 2011; Ismail-Beigi, 2010; Abraira, 2009; Duckworth, 2009; NICE-SUGAR Study Investigators, The, 2009; Ray, 2009; Turnbull, 2009; ACCORD Study Group, The, 2008; ADVANCE Collaborative Group, The, 2008; Gaede, 2008; Holman, 2008
Estimates of benefit and consistency across studies	Benefits: Achieving near-normal glycemic control lowers risk of diabetes microvascular complications such as retinopathy, nephropathy and amputations. Achieving A1c of 6.9 to 7.9% may also significantly reduce macrovascular complications based on Steno-2 and UKPDS data. Follow-up data from the United Kingdom Prospective Diabetes Study of newly diagnosed patients with T2DM confirm major macrovascular and microvascular benefits of achieving A1c in the 7.1 to 7.3% range, versus A1c of about 8% in the comparison groups (Holman, 2008a). The United Kingdom Prospective Diabetes Study main trial included 3,867 newly diagnosed T2DM patients and showed over a 10-year period a 25% decrease in microvascular outcomes with a policy using insulin and sulfonylureas that achieved a median A1c of 7.1%, compared to 7.9%. A subgroup of obese patients (n=1,704) treated with metformin and achieving a median A1c of 7.3% showed greater advantages over conventional treatment: a 32% reduction of diabetes-related end points (P=0.002), a 42% reduction of diabetes-related deaths (P=0.017), and a 36% reduction of all-cause mortality (P=0.011) (UK Prospective Diabetes Study Group, 1998b).

What harms were identified?	Harms: Near-normal glycemic control (A1c around 6.4 to 6.5%) achieved through intensive pharmacotherapy appears to have less benefit for major CV events (ACCORD ADVANCE VADT) and in one large trial significantly increased mortality 20% (ACCORD). In some patients, aggressive pharmacotherapy with insulin, sulfonylureas or certain other agents may lead to weight gain and severe hypoglycemia. The long-term cardiovascular safety of agents other than metformin and human insulins has yet to be established. Benefits-Harms Assessment: Therefore, to optimize the balance between benefits and harms for a given patient, personalization of glycated hemoglobin (A1c) goals in the range of < 7% to < 8% is recommended.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	2018 ADA Standards of Care for glycemic targets reference section (n = 87) reviewed for new studies of A1c targets; no significant new studies change the conclusion of the SR.

Supplemental Information

For patients with T2DM, an A1c goal of less than 8% may be more appropriate than an A1c goal of less than 7%, when including the following factors:

- Known cardiovascular disease or high cardiovascular risk, and may be determined by the Framingham or ACC/AHA Cardiovascular Risk Calculator, or alternatively as having two or more cardiovascular risks (BMI > 30, hypertension, dyslipidemia, smoking and microalbuminuria)
- Inability to recognize and treat hypoglycemia, including a history of severe hypoglycemia requiring assistance
- Inability to comply with standard goals, such as polypharmacy issues
- Limited life expectancy or estimated survival of less than 10 years.
- Cognitive impairment.
- Extensive comorbid conditions such as renal failure, liver failure and end-stage disease complications.

Multifactorial approach

A multifactorial approach to diabetes care that includes emphasis on blood pressure, lipids, glucose, aspirin use and non-use of tobacco will maximize health outcomes far more than a strategy that is limited to just one or two of these clinical domains *(American Diabetes Association, 2014; Duckworth, 2009; Gaede, 2008; Holman, 2008a)*.

The benefits of a multifactorial approach to diabetes care are supported by the results of the Steno-2 Study of 160 patients with T2DM and microalbuminuria. Multifactorial interventions achieved a 50% reduction in mortality and significant reduction in microvascular complications five years after ending a 7.8-year multifactorial intervention that achieved A1c of 7.8%, low-density lipoprotein 83 mg/dL, blood pressure 131/73, compared to a conventional group that achieved A1c 9%, low-density lipoprotein 126 mg/ dL and blood pressure 146/78 *(Gaede, 2008)*. Results of this study are consistent with the need for reasonable blood glucose control with emphasis on blood pressure and lipid management.

In 2018, a study of over 200,000 patients with type 2 diabetes found that excess risk of cardiovascular event outcomes decreased in a stepwise fashion for each risk factor variable that was within the target range. Patients who were successful in achieving targets for all five risk factors had little or no excess risk of death, myocardial infarction or stoke as compared to the general population. *[Rawshani, A et al Risk Factors, Mortality, and Cardiovascular Outcomes in Patients with Type 2 Diabetes N Engl J Med 2018; 379:633-644 DOI: 10.1056/ NEJMoa1800256]*

Microvascular/macrovascular complications

Follow-up data from the United Kingdom Prospective Diabetes Study of newly diagnosed patients with T2DM confirm major macrovascular and microvascular benefits of achieving A1c in the 7.1 to 7.3% range, versus A1c of about 8% in the comparison groups *(Holman, 2008a)*. The United Kingdom Prospective Diabetes Study main trial included 3,867 newly diagnosed T2DM patients and showed over a 10-year period a 25% decrease in microvascular outcomes with a policy using insulin and sulfonylureas that achieved a median A1c of 7.1%, compared to 7.9%. A subgroup of obese patients (n=1,704) treated with metformin and achieving a median A1c of 7.3% showed greater advantages over conventional treatment: a 32% reduction of diabetes-related end points (P=0.002), a 42% reduction of diabetes-related deaths (P=0.017), and a 36% reduction of all-cause mortality (P=0.011) *(UK Prospective Diabetes Study Group, 1998b)*.

Several reported clinical trials have evaluated the impact of A1c less than 7% on macrovascular and micro-vascular complications of T2DM. These studies – the Action to Control Cardiovascular Risk in Diabetes (ACCORD), the Action in Diabetes and Vascular Disease: Preferax and Diamcron Modified Release Controlled Evaluation (ADVANCE), and VADT Trials – are the first that have ever achieved and maintained A1c less than 7% in his/her intensive treatment patients.

Cardiovascular risk

In the ACCORD Trial, excess mortality in the intensive group (A1c mean 6.4% vs. standard group A1c 7.5%) forced the safety board to discontinue the intensive treatment arm earlier than planned *(Action to Control Cardiovascular Risk in Diabetes Study Group, The, 2008)*. There was one excess death for every 90 patients in the intensive group over a 3.5-year period of time. In the ADVANCE trial, intensive group patients achieved A1c 6.5% (vs. 7.5% in standard group) but had no reduction in cardiovascular complications or events. In the VADT trial, intensive group patients achieved A1c of 6.9% but had no significant reduction in cardiovascular events or microvascular complications compared to standard group patients who achieved A1c of 8.4%. However, the VADT Trial was underpowered for its main hypothesis tests *(Duckworth, 2009)*. In the ADVANCE trial, intensive group patients had less progression to proteinuria (one less patient advancing to proteinuria for every 100 people in the intensive group over a five-year period of time), but no fewer eye complications in the intensive group than in the standard group. ACCORD analysis showed lower rates of early stage microvascular complications in the intensively treated group. Some patients, especially those with little comorbidity and long life expectancy, may benefit from more intensive glycemic goals as long as hypoglycemia does not become a barrier. However, the risk of lower glycemic targets may outweigh the potential benefits on microvascular complications for many patients *(ACCORD, 2010b; Ismail-Beigi, 2010)*.

A meta-analysis analyzed five randomized controlled trials (UKPDS, PROactive, ADVANCE, VADT and ACCORD) for the effect of intensive glucose control on cardiovascular outcomes. Overall, this meta-analysis concluded that more intensive glucose control significantly reduced non-fatal myocardial infarct events and coronary heart disease events (non-fatal myocardial infarct and all-cardiac mortality) with no evidence of either a benefit or adverse effect on all-cause mortality. Heterogeneity among studies was noted with regard to all-cause mortality, suggesting that the impact of glycemic reduction on all-cause mortality may differ among different populations *(Ray, 2009)*. A subset analysis from ACCORD, ADVANCE and VADT suggested that intensive glucose lowering has a modest (9%) but statistically significant reduction in major CVD outcomes, primarily non-fatal MI, with no significant effect on mortality. However, a pre-specified subgroup analysis suggested that major cardiovascular disease outcome reduction occurred in patients without known cardiovascular disease at baseline *(Turnbull, 2009)*.

1a.4 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

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

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

Component # 2 Blood Pressure

Measure Number (*if previously endorsed*)**:** 0729

Measure Title: Optimal Diabetes Care – BP Control Component

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

Date of Submission: 10/30/2018

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

Outcome

☐ Outcome: Click here to name the health outcome

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

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

☒ Intermediate clinical outcome (*e.g., lab value*): Blood pressure is less than 140 systolic AND less than 90 diastolic

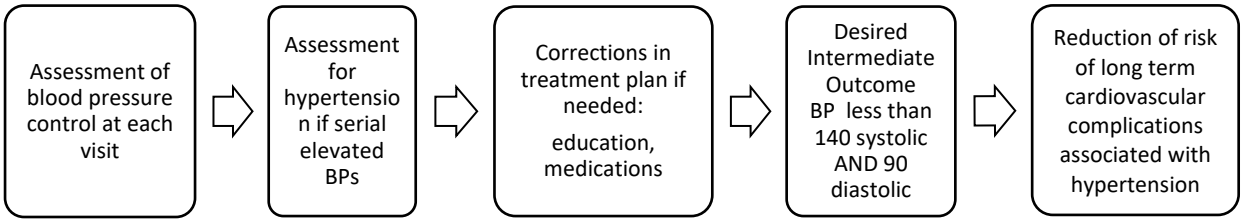
☐ Process: Click here to name what is being measured

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

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

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



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

Measure component is not derived from patient report.

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

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

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

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

☒ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	**Institute for Clinical Systems Improvement (ICSI) Diabetes Mellitus in Adults, Type 2; Diagnosis and Management of. July 2014** **Redmon B, Caccamo D, Flavin P, Michels R, Myers C, O'Connor P, Roberts J, Setterlund L, Smith S, Sperl-Hillen J. Institute for Clinical Systems Improvement. Diagnosis and Management of Type 2 Diabetes Mellitus in Adults. Updated July 2014. Page 33.**
https://www.icsi.org/guidelines__more/catalog_guidelines_and_more/catalog_guidelines/catalog_endocrine_guidelines/diabetes/	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	**Antihypertensive therapy** A clinician should initiate antihypertensive treatment for patients with T2DM with a blood pressure ≥ 140/90 mmHG and treat to a goal of < 140/90.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	Quality of Evidence: High

Provide all other grades and definitions from the evidence grading system	GRADE Methodology of Evidence Used: High Quality of Evidence ▪ Further research is very unlikely to change our confidence in the estimate of effect Moderate Quality Evidence ▪ Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate Low Quality Evidence ▪ Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate or any estimate of effect is very uncertain
Grade assigned to the **recommendation** with definition of the grade	Strength of Recommendation: Strong
Provide all other grades and definitions from the recommendation grading system	GRADE classifies recommendations as strong or weak. Strength of recommendation is determined by the balance between desirable and undesirable consequences of alternative management strategies, quality of evidence, variability in values and preferences, and resource use. Strong Recommendation The work group is confident that the desirable effects of adhering to this recommendation outweigh the undesirable effects. This is a strong recommendation for or against. This applies to most patients.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	Randomized control trial – 6 Systematic review – 2 Meta-analysis of RCT's- 1 Cohort study- 1 Relevant Resources: Arguedas, 2013; Bangalore, 2011; Nilsson, 2011; ACCORD Study Group, 2010a; ADVANCE Collaborative Group, 2008; Howard, 2008; Estacio, 2006; Wing, 2003; ALLHAT, 2002; UKPDS, 1998; Hansson, 1998

Estimates of benefit and consistency across studies	Benefits: Uncontrolled hypertension is a major risk factor for ASVCD events. Multiple large studies (UKPDS, HOT, ADVANCE) have shown improved cardiovascular outcomes with treatment of blood pressure to this range in patients with diabetes. The UKPDS, HOT, ADVANCE and ACCORD trials are all large randomized clinical trials that allow comparison of more stringent to less stringent blood pressure levels on major cardiovascular outcomes (ACCORD Study Group, The, 2010a; ADVANCE Collaborative Group, 2008; Hansson, 1998; United Kingdom Prospective Diabetes Study Group [UKPDS], 1993). The UKPDS, HOT and ADVANCE trials all found reduced cardiovascular outcomes with lower achieved blood pressure levels. However, none of these trials achieved average systolic blood pressure levels below 130 mmHg. The ACCORD trial found no difference in major cardiovascular outcomes between a more intensive blood pressure intervention targeting systolic blood pressure < 120 mmHg compared to a more standard intervention targeting systolic blood pressure between 130 and 139 mmHg. The more intensive blood pressure regimen was associated with a small reduction in the rate of stroke, greater medication use and more serious adverse events (ACCORD Study Group 2010). The general recommendation from The 2014 Evidence-Based Guideline for the Management of High Blood Pressure in Adults: Report From the Panel Members Appointed to the Eighth Joint National Committee (JNC 8) to treat to a goal of a blood pressure < 140/90 mmHg does not preclude setting individual patient goals lower than that based on patient characteristics, comorbidities, risks or the preference of an informed patient (James, 2014).

What harms were identified?	Harms: In many patients with diabetes, two or three or more medications are required to achieve this level of blood pressure control. Medications may be costly, and there are risks of adverse reactions, medication interactions and overtreatment causing hypotension. Benefits-Harms Assessment: Considering the high level of ASCVD risk and the significant benefits for primary and secondary prevention of cardiovascular events in treating hypertension, along with the low cost generic status of the vast majority of antihypertensive medications, it is believed that the benefits of treating hypertension to this goal outweigh the risks. Careful attention should be given to monitoring for side effects, medication interactions and avoiding overtreatment.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	See 2018 discussion of new ACC/ AHA guidelines below in red font. Based on workgroups review of evidence and in light of conflicting and somewhat controversial guidelines that have been rejected by several specialty societies.

New ACC/ AHA Guidelines for the Diagnosis and Management of Hypertension-

Review by MNCM Measure Development Workgroup

Background:

In November of 2017, the American Academy of Cardiology and American Heart Association released new guidelines for the prevention, diagnosis and management of hypertension in adults.[1] These guidelines redefined the diagnosis of hypertension moving from ≥ 140/ 90 to a new definition of stage 1 hypertension (130-139/ 80-89). With new definition, it is estimated that 46% of Americans will now be considered to have hypertension, up from 32% with a definition of ≥ 140/90. The release of the guidelines is not without controversy, and while most agree that a lower blood pressure is better, it is within the context of a patient's individualized goal. Less than 130/80 may not be an appropriate target for every patient. The American College of Physicians and the American Academy of Family Practice has declined endorsement of the new guidelines. They cite concerns with the methodology used in making recommendations and perceived conflict of interest. They are recommending reliance on 2014 JNC8 and ACP/AAFP guidelines for older adults.

Patients with diabetes and cardiovascular disease represent two very high-risk subgroups; in an effort to reduce their modifiable risk factors, the blood pressure component target of the Optimal Diabetes Care (ODC) and Optimal Vascular Care (OVC) measures has reflected a goal that is below the hypertension definition cut-point.

In similar measure development activities, the National Committee for Quality Assurance (NCQA) convened three expert panels (diabetes, cardiovascular and geriatric) for their evaluation of blood pressure targets for the HEDIS Controlling High Blood Pressure measure and concluded that for patients with hypertension ages 18 - 85 the blood pressure target is < 140/90.

[1] American College of Cardiology/ American Heart Association Guidelines for the Prevention, Detection, Evaluation and Management of High Blood Pressure in Adults November 13, 2017

An MNCM convened measure development multi-stakeholder group met in April of 2018 to evaluate and discuss recent changes in guidelines and evidence surrounding blood pressure targets for patients with diabetes and vascular disease. Based on this evaluation, determine the best BP component targets for the composite measures.

Considerable time was devoted to the discussion of the evidence supporting guidelines, applicability of research studies into clinical practice, risk-harm benefit and need for individualized patient goals. After thoughtful and thorough discussion of current guidelines, evidence, and real-world practice implications, the work group gained consensus on the best BP targets for patients with diabetes and vascular disease.

Key considerations included:

- Evaluation of SPRINT (Systolic Blood Pressure Intervention Trial) demonstrates that for the primary outcome of mortality, there is only a 0.5% difference between the intensive treatment group and the standard treatment group. Generalization of the SPRINT results to every day practice raised the issues of:
    1. SPRINT study design called for the withdrawal of treatment in asymptomatic patients in the conservative treatment arm, which does not match clinical practice
    2. Average systolic BP achieved was 121
    3. Best practice methods to obtain BP in a study (auto-BP machine, quiet setting, and resting 10 min) do not match current clinic practice.

The SPRINT study excluded patients with diabetes, so its results are not transferable when there is direct evidence from the ACCORD study that is applicable. ACCORD (Action to Control Cardiovascular Disease in Diabetes) showed very little benefit for BP targets < 140/90.

- Guidelines do not address treatment risks (hypotension, kidney function). The main concern of the workgroup was that in setting a lower target for all patients to strive for, knowing that providers will want to meet that target and may be accountable for hitting that target, may put some patients at risk for serious and costly side effects of intensive treatment. The workgroup would like to encourage individualized targets, knowing that a lower blood pressure is better for the patient, but only if it can be achieved safely.
- There is not consensus at this time among the guideline writing groups about the definition of hypertension or appropriate targets for high risk populations like patients with diabetes or ischemic vascular disease, therefore not a clear direction for measurement to align with guidelines.

Measure Development Workgroup Recommendation:

- **Blood Pressure Targets Remain at < 140/90 for ODC and OVC (unchanged)**
    - Encourage individualized goals for those patients who may benefit from BP target < 130/80

MNCM Diabetes and Vascular Measure Development Work Group Members

Name	Member Type	Organization
Beth Averbeck, MD	Clinical Provider; Internal Medicine; Chair	HealthPartners
Joseph Bianco, MD	Clinical Provider; Family Medicine and MARC	Essentia Health- Ely
Andrew Greenland, MD	Clinical Provider; Internal Medicine	Mayo Clinic
Christopher Fallert, MD	Clinical Provider; Family Medicine	University of Minnesota
Christian Anderson, MD	Clinical Provider; Family Medicine	Entira Family Clinics
Steven Bradley, MD MPH	Clinical Provider; Cardiology	Minneapolis Heart
David Homans, MD	Clinical Provider; Cardiology	Park Nicollet

Name	Member Type	Organization
Jesse Wheeler, MD	Clinical Provider; Nephrology and MARC	Park Nicollet
Nicole Paterson, PharmD	Clinical Provider; Pharmacist	Fairview Health Services
Karen Margolis, MD MPH	Data Analyst	HealthPartners
Cindy Ferrara, RN	Quality Improvement	Essentia Health- Duluth
Patrick Schultz, ACNS-BC	Clinic Administrator	Sanford
James Peacock, PhD MPH	State Agency	MN Dept. of Health
Cynthia Toher, MD	Health Plan/ Cardiologist	Blue Cross/Blue Shield MN
David Klocke, MD	Health Plan/Hospital Medicine and EM	Blue Cross/Blue Shield MN
Christine Norton	Consumer and MARC	Retired
Deb Krause	Employer	MN Health Action Group

_____

1a.4 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

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

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

Component # 3 Cholesterol Statin Use

Measure Number (*if previously endorsed*)**:** 0729

Measure Title: Optimal Diabetes Care- Cholesterol Statin Use Component

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

Date of Submission: 10/30/2018

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

Outcome

☐ Outcome: Click here to name the health outcome

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

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

☒ Intermediate clinical outcome (*e.g., lab value*): Appropriate statin use for patients with diabetes (based on age, presence of ischemic vascular disease or LDL level greater than 190).

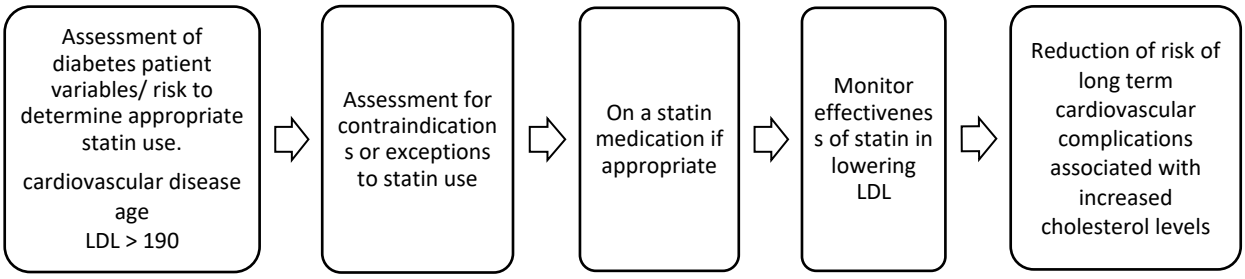
☐ Process: Click here to name what is being measured

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

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

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



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

Measure component is not derived from patient report.

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

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

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

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

☒ Clinical Practice Guideline recommendation (with evidence review)

☐US Preventive Services Task Force Recommendation

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

☐Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	**Institute for Clinical Systems Improvement (ICSI) Diabetes Mellitus in Adults, Type 2; Diagnosis and Management of. July 2014** **Redmon B, Caccamo D, Flavin P, Michels R, Myers C, O'Connor P, Roberts J, Setterlund L, Smith S, Sperl-Hillen J. Institute for Clinical Systems Improvement. Diagnosis and Management of Type 2 Diabetes Mellitus in Adults. Updated July 2014. Page 35.** **American College of Cardiology/ American Heart Association** **Guideline on the Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Risk in Adults. November 2013**
https://www.icsi.org/guidelines__more/catalog_guidelines_and_more/catalog_guidelines/catalog_endocrine_guidelines/diabetes/ https://www.ahajournals.org/cms/attachment/069caaec-0338-4b71-8182-9e841cb8c509/01.cir.0000437738.63853.7av1.pdf	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	**ICSI- Statin Therapy (High Risk)** A clinician should recommend high-intensity statin therapy for patients diagnosed with T2DM, between the ages of 40-75 with established ASCVD (strong) **ICSI- Statin Therapy (Moderate Risk)** A clinician should recommend moderate- or high-intensity statin therapy for all patients diagnosed with T2DM between the ages of 40-75 with a LDL $\geq$ 70 mg/dL. (strong) **ACC/AHA Recommendations for Primary Prevention in Individuals With Diabetes Mellitus and LDL–C 70-189 mg/dL** Moderate-intensity statin therapy should be initiated or continued for adults 40 to 75 years of age with diabetes mellitus. (strong) High-intensity statin therapy is reasonable for adults 40 to 75 years of age with diabetes mellitus with a ≥7.5% estimated 10-year ASCVD risk unless contraindicated. (expert opinion) In adults with diabetes mellitus, who are <40 or >75 years of age, it is reasonable to evaluate the potential for ASCVD benefits and for adverse effects, for drug-drug interactions, and to consider patient preferences when deciding to initiate, continue, or intensify statin therapy. (expert opinion)
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	**ICSI**- Quality of Evidence: High **ACC/AHA-** I and A

Provide all other grades and definitions from the evidence grading system	**ICSI**- GRADE Methodology of Evidence Used: High Quality of Evidence ▪ Further research is very unlikely to change our confidence in the estimate of effect Moderate Quality Evidence ▪ Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate Low Quality Evidence Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate or any estimate of effect is very uncertain **ACC/AHA-** please see page 5 or URL for explanation of combined method Class I = Benefits >>> Risks Procedure or treatment **should** be performed Level A = multiple populations identified and data derived from multiple random control trials or mega-analysis.
Grade assigned to the **recommendation** with definition of the grade	**ICSI-** Strength of Recommendation: Strong **ACC/AHA-** Strength of Recommendation: Strong (I A)

Provide all other grades and definitions from the recommendation grading system	**ICSI-** GRADE classifies recommendations as strong or weak. Strength of recommendation is determined by the balance between desirable and undesirable consequences of alternative management strategies, quality of evidence, variability in values and preferences, and resource use. Strong Recommendation The work group is confident that the desirable effects of adhering to this recommendation outweigh the undesirable effects. This is a strong recommendation for or against. This applies to most patients. **ACC/ AHA** Uses a combination of NHLBI and Class of Recommendation/ Level of Evidence (COR/LOE) High Quality of Evidence ▪ Well-designed, well-executed† RCTs that adequately represent populations to which the results are applied and directly assess effects on health outcomes. ▪ MAs of such studies. ▪ Highly certain about the estimate of effect. Further research is unlikely to change our confidence in the estimate of effect. Moderate Quality of Evidence ▪ RCTs with minor limitations‡ affecting confidence in, or applicability of, the results. ▪ Well-designed, well-executed nonrandomized controlled studies§ and well designed, well-executed observational studies ▪ MAs of such studies. ▪ Moderately certain about the estimate of effect. Further research may have an impact on our confidence in the estimate of effect and may change the estimate.

Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	**ICSI-** Randomized control trial – 60 Systematic review – 1 Meta-analysis of RCT's- 1 Relevant Resources: Taylor, 2013; CTT, 2010; Cannon, 2004; Heart Protection Study Collaboration Group, 2002, Macchia, 2012; AIM High, 2011; ACCORD Study Group, 2010; CTT, 2008; Heart Protection Study, 2005; Cannon, 2004 **ACC/AHA-** High and Moderate Quality of Evidence cited in Recommendations table for Diabetics (5 High, 4 Moderate all related to RCTs)
Estimates of benefit and consistency across studies	**ICSI-** Benefits: A high-intensity statin reduces the relative risk of ASCVD events more than moderate-intensity statin in patients with and without diabetes, and in primary and secondary prevention in those with diabetes. The use of at least moderate-intensity statin therapy in persons of this age and an elevated LDL level with a diagnosis of diabetes has been shown to be effective. The only trial of high-intensity therapy in primary prevention was performed in a population without diabetes. High-intensity statin therapy reduces the relative risk of ASCVD events more than moderate-intensity statin therapy in patients with ASCVD. Because individuals with diabetes are at substantially increased lifetime risk for ASCVD events and death, similar to those who have had a previous ASCVD event, persons with diabetes with high estimated 10-year ASCVD risk are likely to benefit similarly from high-intensity therapy.

What harms were identified?	**ICSI-** Harms: Harms: Serious adverse events such as myopathy and rhabdomyolysis are rare, but patient characteristics that may influence statin safety and be cause for not recommending high-intensity statin therapy include multiple concomitant comorbidities, impaired renal or hepatic function, a history of previous statin intolerance or muscle disorders, concomitant use of drugs known to affect statin metabolism, a history of hemorrhagic stroke and age > 75. **ICSI-** Benefits-Harms Assessment: The benefits of high-intensity statin therapy for patients with diabetes and high ASCVD risk usually outweigh potential harm, but side effects and individual patient characteristics that predispose patients to statin toxicity can influence the risk/harm balance. Patient preference should be included in decision-making.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	No new guidelines published. 2018 ADA Standards of Care for cardiovascular disease and risk management reference section (n = 146) reviewed for new studies of statin use; no significant new studies change the conclusion of the SR.

1a.4 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

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

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

Component # 4 Tobacco Free

Measure Number (*if previously endorsed*)**:** 0729

Measure Title: Optimal Diabetes Care- Tobacco-Free Component

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

Date of Submission: 10/30/2018

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

Outcome

☒ Outcome: Patient is tobacco-free

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

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

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

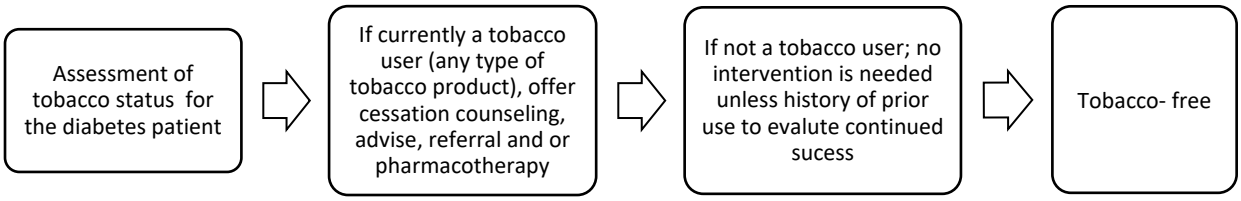
☐ Process: Click here to name what is being measured

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

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

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



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

Measure component is not derived from patient report.

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

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

According to the Centers for Disease Control, cigarette smoking is the most important preventable cause of premature death in the United States. Cigarette smoking is the leading cause of preventable death in the United States, accounting for more than 480,000 deaths, or one of five deaths, each year.

A multifactorial approach to diabetes care that includes emphasis on blood pressure, lipids, glucose, aspirin use and non-use of tobacco will maximize health outcomes far more than a strategy that is limited to just one or two of these clinical domains *(American Diabetes Association, 2014; Duckworth, 2009; Gaede, 2008; Holman, 2008a)*.

Tobacco smoking increases risk of macrovascular complications 4-400% in adults with T2DM and also increases risk of macrovascular complications. Tobacco cessation is very likely to be the single most beneficial intervention that is available, and it should be emphasized by clinicians.

ICSI Diabetes Mellitus in Adults, Type 2; Diagnosis and Management of. July 2014

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

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

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

1a.4 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

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

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

Component # 5 Daily Aspirin/ Anti-platelet

Measure Number (*if previously endorsed*)**:** 0729

Measure Title: Optimal Diabetes Care - Aspirin Anti-platelet Use Component

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

Date of Submission: 10/30/2018

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

Outcome

☐ Outcome: Click here to name the health outcome

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

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

☒ Intermediate clinical outcome (*e.g., lab value*): Appropriate daily aspirin or antiplatelet use for patients with diabetes with ischemic vascular disease.

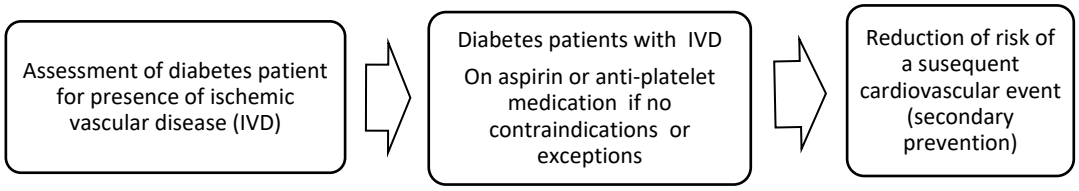
☐ Process: Click here to name what is being measured

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

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

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



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

Measure component is not derived from patient report.

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

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

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

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

☒ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	**Institute for Clinical Systems Improvement (ICSI) Diabetes Mellitus in Adults, Type 2; Diagnosis and Management of. July 2014** **Redmon B, Caccamo D, Flavin P, Michels R, Myers C, O'Connor P, Roberts J, Setterlund L, Smith S, Sperl-Hillen J. Institute for Clinical Systems Improvement. Diagnosis and Management of Type 2 Diabetes Mellitus in Adults. Updated July 2014. Page 36.**
https://www.icsi.org/guidelines__more/catalog_guidelines_and_more/catalog_guidelines/catalog_endocrine_guidelines/diabetes/	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	**Aspirin Therapy** A clinician should recommend aspirin therapy for patients diagnosed with T2DM with established ASCVD and consider aspirin therapy for others where the benefits outweighs the risk in primary prevention.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	Quality of Evidence: High

Provide all other grades and definitions from the evidence grading system	GRADE Methodology of Evidence Used: High Quality of Evidence ▪ Further research is very unlikely to change our confidence in the estimate of effect Moderate Quality Evidence ▪ Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate ▪ Low Quality Evidence Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate or any estimate of effect is very uncertain
Grade assigned to the **recommendation** with definition of the grade	Strength of Recommendation: Strong
Provide all other grades and definitions from the recommendation grading system	GRADE classifies recommendations as strong or weak. Strength of recommendation is determined by the balance between desirable and undesirable consequences of alternative management strategies, quality of evidence, variability in values and preferences, and resource use. Strong Recommendation The work group is confident that the desirable effects of adhering to this recommendation outweigh the undesirable effects. This is a strong recommendation for or against. This applies to most patients.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	Randomized control trial – 5 Systematic review – 2 Meta-analysis of RCT's- 1 Relevant Resources: Rosiak, 2013; Macchia, 2012; Soejima, 2012; Valentine, 2012; Antithrombotic Trialists' (ATT) Collaboration, 2009; Belch, 2008; Ogawa, 2008; Campbell, 2007; Pignone, 2006
Estimates of benefit and consistency across studies	Benefits: Patients with established ASCVD are at high risk for recurrent events, and aspirin therapy for secondary prevention has been shown to reduce the rate of future events to a clinically meaningful degree. As T2DM is an independent risk factor for ASCVD, patients with T2DM might be expected to benefit from aspirin therapy even before they manifest evidence of ASCVD.

What harms were identified?	Harms: Aspirin therapy could increase the risk of clinically significant bleeding and is also associated with medication cost. Benefits-Harms Assessment: The substantial reduction in recurrent ASCVD events with aspirin therapy in secondary prevention will outweigh the risk of bleeding for patients with established ASCVD and no contraindications to aspirin use. In patients with T2DM where aspirin is considered for primary prevention, while the risk of clinically significant bleeding is low, it is still likely increased relative to no therapy. At this time, it is unclear whether adding aspirin therapy to other standard therapy for CV risk factors adds net benefit in patients with T2DM who do not have established ASCVD.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	2018 ADA Standards of Care for cardiovascular disease and risk management reference section (n = 146) reviewed for new studies of aspirin or antiplatelet use; no significant new studies change the conclusion of the SR.

Supplemental Information

Patients with T2DM are at a significantly increased risk for development of heart disease (American Diabetes Association, 2014). Recent trials of aspirin use in diabetes have shown less benefit than older trials for primary prevention, perhaps due to better background A1c, blood pressure, and low-density lipoprotein control and lower smoking rates in recent trials (Rosiak, 2013; Macchia, 2012; Belch, 2008; Ogawa, 2008).

_____

1a.4 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

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

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

1b. Performance Gap

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

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

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

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

Measure is a composite; please refer to questions 1c.2., 1c.3., and 1c.4.

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.*

For 2017 dates of service, 44.9% of the patients met all five component targets in the composite measure and considered optimally managed. This rate is a weighted average of the total population of patients for clinics (618) submitting data (Total Population = 307,661 Submitted = 307,158).

There was a wide range of variability with the lowest scoring clinic at 9% and the highest scoring clinic at 63.8%

The trends for this measure by reporting year:

Year	Rate	Patients (Den)	Numerator	Eligible	% submit/elig
2006	9.5%	8,401	798	41,831	20.1%
2007*	13.5%	58,911	8,297	85,225	69.1%
2008	17.1%	83,034	15,772	130,019	63.9%
2009	18.9%	112,819	23,470	178,748	63.1%
2010	25.1%	140,945	40,078	216,290	65.2%
2011	37.0%	158,770	61,930	209,479	75.8%
2012	38.2%	184,234	73,037	212,077	86.9%
2013	37.7%	208,809	80,190	223,036	93.6%
2014	38.9%	230,818	90,499	237,354	97.2%
2015+	53.5%	245,241	131,847	249,878	98.1%
2016	46.3%	257,078	119,194	260,197	98.8%
2017	44.8%	295,049	132,106	295,199	99.9%
2018	44.9%	307,158	137,985	307,661	99.8%**

* New direct data submission process. Increasing rates indicative of EMR implementation and submission of full population data, which is a MN Health Reform requirement for all clinics that had an EMR implemented for the full year prior to the measurement year.

** Only 3 clinics in the state of MN submitted a sample for RY 2018

+ Measure component change. Cholesterol component previously specified as LDL < 100 prior to report year 2015, was suppressed during measure redesign. RY 2015 reflects artificially higher composite rate with only 4 components. RY 2016 reflects newly redesigned cholesterol component for appropriate statin use.

Consumer facing website MN HealthScores Displays the top 15 best performers in addition to rates for all clinics in MN:

Rating%	Clinic	Location
64%	CentraCare River Campus-Internal Medicine	St. Cloud, MN
63%	Fairview Rogers Clinic	Rogers, MN
62%	Park Nicollet Clinic- Golden Valley	Golden Valley, MN
62%	Entira Family Clinics- White Bear Lake	White Bear Lake, MN
61%	Mayo Clinic Health System St. Peter	St. Peter, MN

60% | Fairview Rosemount Clinic | Rosemount, MN

60% | Entira Family Clinics- West St. Paul | West St. Paul, MN

60% | Fairview Eagan Clinic | Eagan, MN

58% | Fairview Lakeville Clinic | Lakeville, MN

58% | Allina Health- Uptown | Minneapolis, MN

58% | CentraCare Clinic- Northway | St. Cloud, MN

57% | Fairview Pine City Clinic | Pine City, MN

57% | Fairview Uptown Clinic | Minneapolis, MN

56% | Entira Family Clinics- Como/Roseville | St. Paul, MN

52% | Sanford Health Vermillion Clinic | Vermillion, SD

51% | Fairview Bloomington Lake Minneapolis | Minneapolis, MN

42% | Southside Medical Clinic | Minneapolis, MN

Individual rates of the components are as follows:

A1c < 8.0 | 69%

Statin | 88%

Blood Pressure < 140/90 | 83%

Daily Aspirin Use if IVD| 99%

Tobacco Non-user | 84%

Optimal Diabetes Care Rates by Clinic

Number of Medical Groups: 100

Number of Reportable Clinics (n is greater than or equal to 30): 618

Rate Range Clinics: 9% to 63.4%

Ranking of Clinics:

Above 25.6% (158)

Average 48.1% (297)

Below 26.4% (163)

Distribution of Rates Among Clinics:

Range of Rate | # of Clinics | % of Clinics

0%-9.9% | 1 | 0.2%

10%-19.9% | 17 | 2.8%

20%-29.9% | 52 | 8.4%

30%-39.9% | 118 | 19.1%

40%-49.9% | 260 | 42.1%

50%-59.9% | 158 | 25.6%

60%-69.9% | 12 | 1.9%

Weighted Mean: 44.9%

Mean: 43.1%

Median: 45.0%

Standard Deviation: 0.102

Min: 9.0%

Max: 63.0%

Clinic Rates by Decile:

Percentile	Rate
10th	28.1%
20th	35.0%
30th	39.5%
40th	42.5%
50th	45.0%
60th	46.8%
70th	49.0%
80th	51.8%
90th	55.2%

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.

Not applicable; opportunity for improvement and variation demonstrated.

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.*

Race/ Ethnicity	2014	2015	2016	2017	2018
Asian	44%	59%	53%	48%	48%
White	41%	56%	48%	47%	47%
Native Hawaiian/Pacific Isl	36%	48%	46%	44%	41%
Multi-Racial	30%	48%	43%	38%	36%
Black or African American	27%	41%	35%	33%	33%
American Indian/Alaska Native	25%	31%	27%	24%	24%
Some Other Race	42%	53%	47%	44%	43%
Unknown	24%	49%	34%	28%	32%
Chose not to disclose	42%	53%	46%	44%	46%
Race not Reported	31%	46%	37%	37%	40%
Statewide Average	39%	53%	46%	44%	45%

Age Band	2014	2018
18 to 25	15.0%	29.0%
26 to 50	27.2%	33.8%
51 to 65	38.6%	43.2%
66 to 75	49.2%	55.5%

Grand Total | 39.2% | 44.9%

Type of Diabetes | 2014 | 2018

Type I DM | 28.3% | 34.7%

Type 2 DM | 40.1% | 45.8%

Type Unknown DM | 26.6% | 32.2%

Grand Total | 39.2% | 44.9%

Of note, our rates for this measure compare similarity with literature supporting increased risk of diabetes for African Americans, Asian Americans, and American Indians.

Age, family history and a previous history of gestational diabetes are indicators of increased risk for diabetes, along with being African American, Asian American, Hispanic/Latino or American Indian. Potentially modifiable risks for developing diabetes include: obesity, inactivity, high blood pressure and abnormal cholesterol levels. Studies show that people at high risk for type 2 diabetes can prevent or delay the onset of the disease by maintaining a healthy diet and regular exercise. Knowler WC. N Engl J Med 346(6):393-403, 2002.

The risk of diabetes increases with age. According to projections from the Minnesota State Demographic Center, the population aged 65 years and older will rise sharply in the coming decades: In 2000, one in every eight Minnesotans were 65 years of age or older; by 2030, that ratio will be one in five. Increases in the elderly population are likely to contribute significantly to the burden of diabetes in Minnesota in the future. African American, Asian or Pacific Islander, American Indian or Hispanic/Latino American populations are at greater risk for developing diabetes, and these populations are also growing. In 2000, roughly one in every eight (12 percent) of Minnesota's nearly five million people were Persons of Color or American Indians; by 2025, that proportion will

be 17 percent, or nearly one in every five.

There continues to be a significant gap in rates for this measure between MN Health Care Programs- MHCP (25.3%) and Other Purchasers.

Year | 2013 | 2014 | 2015 | 2016 | 2017

MN Health Programs- MHCP | 25.3% | 32.1% | 43.1% | 33.6% | 32.5%

Other Purchasers | 40.2% | 41.9% | 56.7% | 48.9% | 47.6%

Health Care Disparities reports can be accessed at http://mncm.org/reports-and-websites/reports-and-data/health-care-disparities-report/ ?

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

Not applicable; opportunity for improvement and variation demonstrated within disparate health categories.

1c. Composite Quality Construct and Rationale

1c.1. A composite performance measure is a combination of two or more component measures, each of which individually reflects quality of care, into a single performance measure with a single score.

For purposes of NQF measure submission, evaluation, and endorsement, the following will be considered composites:

- Measures with two or more individual performance measure scores combined into one score for an accountable entity.
- Measures with two or more individual component measures assessed separately for each patient and then aggregated into one score for an accountable entity:

- all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient);

1c.1. Please identify the composite measure construction: all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient)

1c.2. Describe the quality construct, including:

- the overall area of quality
- included component measures and
- the relationship of the component measures to the overall composite and to each other.

This composite measure is a patient level all-or-none composite in which the desired goal is for the patient is to achieve multiple intermediate physiological clinical outcome and medication use targets to best reduce their overall risk of developing long term complications (acute MI, cardiovascular and peripheral vascular disease, kidney damage and failure, loss of vision, amputation, etc.) Reducing modifiable risks was the reason why this measure was developed. The components of this measure include blood sugar and blood pressure control, being tobacco-free, appropriate use of statins and daily aspirin or anti-platelet use if ischemic vascular disease.

1. HbA1c less than 8.0
2. Blood pressure less than 140 systolic and less than 90 diastolic
3. Statin use if no contraindications/ exceptions
4. Tobacco-free
5. Daily aspirin or anti-platelets if has ischemic vascular disease and no contraindications/ exceptions

Numerator is calculated at the patient level and numerator compliance is defined as the patient achieving all five components of the measure. The components are treated equally.

1c.3. Describe the rationale for constructing a composite measure, including how the composite provides a distinctive or additive value over the component measures individually.

Achieving the intermediate physiological outcome targets related to blood pressure and glycemic control in addition being tobacco free and use of daily aspirin and statins where appropriate are the diabetic patient's best mechanisms of avoiding or postponing long term complications associated with this chronic condition which affects millions of Americans. Measuring providers separately on individual targets is not as patient centric as a measure that seeks to reduce multiple risk factors for each patient. Diabetic patients are more likely to reduce their overall risk and maximize health outcomes by achieving several intermediate physiological targets.

In 2018, a study of over 200,000 patients with type 2 diabetes found that excess risk of cardiovascular event outcomes decreased in a stepwise fashion for each risk factor variable that was within the target range. Patients who were successful in achieving targets for all five risk factors had little or no excess risk of death, myocardial infarction or stoke as compared to the general population. [Rawshani, A et al Risk Factors, Mortality, and Cardiovascular Outcomes in Patients with Type 2 Diabetes N Engl J Med 2018; 379:633-644 DOI: 10.1056/ NEJMoa1800256]

1c.4. Describe how the aggregation and weighting of the component measures are consistent with the stated quality construct and rationale.

Numerator is calculated at the patient level and numerator compliance is defined as the patient achieving all five components of the measure. The components are treated equally; there is no weighting. Some of the components have an exception methodology within allowing a "free-pass" on the component if it does not apply to the patient.

Most recent HbA1c in the measurement period is less than 8.0 (applies to all denominator patients)

AND

Most recent blood pressure in the measurement period is less than 140 systolic AND less than 90 diastolic (applies to all denominator patients)

AND

Statin Use if appropriate and no contraindications/ exceptions

Diabetic age 18-20 "free-pass"

Diabetic Age 21 to 75 and ischemic vascular disease? on statin unless LDL < 40 or contraindications/ exceptions

Diabetic Age 21 to 39 and LDL greater than or equal to 190? on statin or contraindications/ exceptions. If in this age group and LDL less than 190 is a "free-pass"

Diabetic Age 40 to 75 ? on statin unless LDL < 70 or contraindications/exceptions.

AND

Patient's tobacco status (documented within the last 2 years) is tobacco free (applies to all denominator patients)

AND

Daily aspirin or anti-platelet use if patient has ischemic vascular disease. If the patient does not have ischemic vascular disease, this component is a "free-pass"

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):*

Person-and Family-Centered Care

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.)*

http://mncm.org/wp-content/uploads/2017/10/Optimal-Diabetes-Care-2018-Data-Collection-Guide-FINAL-v1.pdf

S.2a. If this is an eMeasure, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)

This is not an eMeasure **Attachment:**

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

Attachment **Attachment:** MNCM_Diabetes_Measure_Data_Dictionary_and_Risk_Adj__10-19-2018.xlsx

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

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

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

Not an instrument-based measure

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

Yes

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

For denominator inclusion, changed from a visit counting methodology (two diabetes visits in 2 years) to a method that utilizes both diagnoses coded to a contact and/or diagnosis on active problem list. For inclusiveness and accuracy, both must be queried. This is used in combination with an established patient CPT office visit code during the measurement period. For this measure and the denominator population, the impact on the denominator was relatively small but clinically accurate (~ 15% increase).

After pilot testing this specification change, hypothesis was confirmed that the previous visit counting method was eliminating eligible patients who did indeed have the chronic disease of interest/ measure focus.

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

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

The number of patients in the denominator whose diabetes was optimally managed during the measurement period as defined by achieving ALL of the following:

- The most recent HbA1c in the measurement period has a value less than 8.0 mg/dL
- The most recent Blood Pressure in the measurement period has a systolic value of less than 140 mmHg AND a diastolic value of less than 90 mmHg
- On a statin medication, unless allowed contraindications or exceptions are present
- Patient is not a tobacco user
- Patient with ischemic vascular disease (Ischemic Vascular Disease Value Set) is on daily aspirin or anti-platelets, unless allowed contraindications or exceptions are present

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).

Please note that while the all-or-none composite measure is considered to be the gold standard, reflecting best patient outcomes, the individual components may be measured as well. This is particularly helpful in quality improvement efforts to better understand where opportunities exist in moving the patients toward achieving all of the desired outcomes. Please refer to the additional numerator logic provided for each

component and note that all of the denominator criteria apply to the numerator as well, but are not repeated in the numerator codes/ descriptions.

HbA1c Date [Date (mm/dd/yyyy)] AND

HbA1c Value [Numeric]

Numerator component calculation: numerator component compliant is HbA1c during the last 12 months (measurement year) AND most recent HbA1c value is less than 8.0.

Enter the date of the most recent HbA1c test during the measurement period.

Enter the value of the most recent HbA1c test during the measurement period.

Leave BLANK if an HbA1c was never performed.

- A test result from a provider outside of the reporting medical group is allowed if the result is documented in the reporting medical group's patient record and is the most recent test result during the measurement period.
- If the HbA1c result is too high to calculate, still enter the HbA1c test date if it is the most recent test result during the measurement period.

Blood Pressure Date [Date (mm/dd/yyyy)] AND

BP Systolic [Numeric] AND

BP Diastolic [Numeric]

Numerator component calculation: numerator component compliant is BP during the measurement year AND Systolic < 140 AND Diastolic < 90.

Enter the date of the most recent blood pressure result during the measurement period.

Leave BLANK if a blood pressure was not obtained during the measurement period.

- A test result from a provider outside of the reporting medical group is allowed if the result is documented in the reporting medical group's patient record and is the most recent test result during the measurement period.
- Do not include BP readings:
  - Taken during an acute inpatient stay or an ED visit.
  - Taken during an outpatient visit which was for the sole purpose of having a diagnostic test or surgical procedure performed (e.g., sigmoidoscopy, removal of a mole).
  - Obtained the same day as a major diagnostic or surgical procedure (e.g., EKG/ECG, stress test, administration of IV contrast for a radiology procedure, endoscopy).
  - Reported by or taken by the patient.

BP Systolic

Enter the value of the most recent systolic blood pressure result during the measurement period.

If more than one value is recorded on the most recent date, the lowest systolic value from multiple readings on the same date may be submitted.

NOTE: The systolic blood pressure is the upper number in the recorded fraction. For example, the systolic value for a blood pressure of 124/72 mmHg is 124.

BP Diastolic

Enter the value of the most recent diastolic blood pressure result during the measurement period.

If more than one value is recorded on the most recent date, the lowest diastolic value from multiple readings on the same date may be submitted.

- NOTE: The diastolic blood pressure is the lower number in the recorded fraction. For example, the diastolic value for a blood pressure of 124/72 mmHg is 72.

LDL Date [Date (mm/dd/yyyy)] AND

LDL Value [Numeric]

Numerator component calculation: Is used for the cholesterol component for statin use; patients with low untreated LDL values may not be appropriate for the initiation of statin medication.

Enter the date of the most recent LDL test on or prior to the end of the measurement period.

Leave BLANK if an LDL was never performed.

- A test result from a provider outside of the reporting medical group is allowed if the result is documented in the reporting medical group's patient record and is the most recent test result within the allowable time period.
- If the LDL result is too high to calculate, still enter the LDL test date if it is the most recent test result within the allowable time period.

LDL values within the last five years will be used to calculate potential exceptions to being on a statin medication. Leave BLANK if an LDL test was not performed between 01/01/201x and 12/31/201x (five-year increments).

Statin Medication [Numeric] AND

Statin Medication Date [Date (mm/dd/yyyy)] AND/OR

Station Medication Exception [Numeric] AND

Station Medication Exception Date [Date (mm/dd/yyyy)]

Numerator component calculation: numerator component compliant if on a statin (prescribed/ ordered) or low LDL value (see above) or documented contraindication/exception is present.

Statin Medication:

Enter the code that corresponds to whether the patient was prescribed a statin medication or if a statin medication was active on the patient's medication list during the measurement period.

Please refer to Appendix C for a list of statin medications.

1 = Yes, patient was prescribed a statin medication or a statin medication was indicated as active on the patient's medication list during the measurement period.

2 = No, patient was not prescribed a statin medication and a statin medication was not indicated as active on the patient's medication list during the measurement period.

The following exceptions to statin medication use will be identified by the Data Portal based on the submitted LDL values:

- Patients with ischemic vascular disease aged 21 to 75 years and an LDL result less than 40 mg/dL
- Patients aged 40 – 75 years with an LDL result less than 70 mg/dL
- Patients aged 21 – 39 years with an LDL less than 190 mg/dL

Statin Medication Date:

Enter the most recent date of a statin prescription, order or review of active medications list during the measurement period.

If no statin prescribed, ordered, or reviewed as an active medication during the measurement period, leave blank

Statin Medication Exception:

If the patient was NOT prescribed or did not have a statin medication active on their medication list during the measurement period, enter the value that corresponds to any of the following contraindications or exceptions:

1 = Pregnancy at any time during the measurement period

2 = Active liver disease (liver failure, cirrhosis, hepatitis)

3 = Rhabdomyolysis

4 = End stage renal disease on dialysis

5 = Heart failure

6 = Other provider documented reason: breastfeeding during the measurement period

7 = Other provider documented reason: woman of childbearing age not actively taking birth control during the measurement period

8 = Other provider documented reason: allergy to statin

9 = Drug interaction with a listed medication taken during the measurement period (valid drug-drug interactions include HIV protease inhibitors, nefazodone, cyclosporine, gemfibrozil, and danazol).

10 = Other provider documented reason: intolerance (with supporting documentation of trying a statin at least once within the last five years). Additionally, Myopathy and Myositis (CHOL-05) Value Set may be used to document intolerance to statins.

If none of the above contraindications or exceptions are documented, leave BLANK. NOTE: Items 1 – 5 above can be defined by diagnosis codes that may be used in data collection. Value Sets include: Pregnancy V/Z Codes (PREG-01), Pregnancy Diagnosis Codes (PREG-02), Liver Disease (CHOL-01), Rhabdomyolysis (CHOL-02), ESRD on Dialysis (CHOL-03), and Heart Failure (CHOL-04)

Statin Medication Exception Date:

If the patient has a documented contraindication or exception enter the date of the contraindication or exception. If only the month and year are known, enter the first day of the month.

Tobacco Status Documentation Date [Date (mm/dd/yyyy)] AND

Tobacco Status [Numeric]

Numerator component calculation: numerator component compliant if tobacco status within the last two years and status is tobacco-free.

Tobacco Status Documentation Date:

Enter the most recent date that the patient's tobacco status was documented during the measurement period or year prior.

• If the patient's tobacco status is not documented or the date of documentation cannot be determined, leave BLANK

Tobacco Status:

Enter the code that corresponds to the patient's most recent tobacco status during the measurement period or year prior.

1 = Tobacco free (patient does not use tobacco; patient was a former user and is not a current user)

2 = No documentation

3 = Current tobacco user (tobacco includes any amount of cigarettes, cigars, pipes or smokeless tobacco)

- If the date of the tobacco status documentation is not documented in the patient record, enter 2
- E-cigarettes are not considered tobacco products.

Aspirin or Anti-platelet Medication [Numeric] AND

Aspirin or Anti-platelet Date [Date (mm/dd/yyyy)] AND/OR

Aspirin or Anti-platelet Exception [Numeric] AND

Aspirin or Anti-platelet Exception Date [Date (mm/dd/yyyy)]

Numerator component calculation: Calculation applied only if patient has ischemic vascular disease (IVD); if no IVD indicated, is a numerator component "free-pass". For patients with IVD, numerator component compliant if indicated on daily aspirin or anti-platelet medication (prescribed/ ordered) or documented contraindication/exception is present.

Aspirin or Anti-platelet Medication:

For patients with Ischemic Vascular Disease (IVD), enter the code that corresponds to whether the patient is prescribed a daily aspirin product or antiplatelet medication or if an aspirin product or anti-platelet medication was active on the patient's medication list during the measurement period.

Please see Appendix D for methods to identify appropriate aspirin products or antiplatelet medications.

1 = Yes, patient was prescribed a daily aspirin product or antiplatelet medication, or one was indicated as active on the patient's medication list during the measurement period.

2 = No, patient was not prescribed a daily aspirin product or antiplatelet medication and one was not indicated as active on the patient's medication list during the measurement period.

Aspirin/narcotic combination medications do not qualify as a daily aspirin product.

Aspirin or Anti-platelet Date:

For patients with IVD, enter the date of the most recent daily aspirin product or anti-platelet medication prescription, order or review of an active medication list that included a daily aspirin product or anti-platelet medication during the measurement period.

If a daily aspirin product or anti-platelet medication was not prescribed, ordered or reviewed as an active medication during the measurement period leave blank

Aspirin or Anti-platelet Medication Exception:

For patients with IVD who were not prescribed or taking a daily aspirin product or anti-platelet medication during the measurement period, enter the code that corresponds to any of the following contraindications or exceptions:

1 = Prescribed anti-coagulant medication during the measurement period

2 = History of gastrointestinal bleeding

3 = History of intracranial bleeding

4 = Bleeding disorder

5 = Other provider documented reason: allergy to aspirin or anti-platelets

6 = Other provider documented reason: use of non-steroidal anti-inflammatory agents

7 = Other provider documented reason: documented risk for drug interaction with a medication taken during the measurement period.

8 = Other provider documented reason: uncontrolled hypertension (systolic blood pressure greater than 180 mmHg and/or diastolic blood pressure greater than 110 mmHg)

9 = Other provider documented reason: gastroesophageal reflux disease (GERD)

If none of the above contraindications or exceptions are documented, leave BLANK.

NOTE: Items 2 and 3 above can be defined by diagnosis codes that may be used in data collection. Value Sets include: GI Bleed (ASA-01) and Intracranial Bleed (ASA-02).

Aspirin or Anti-platelet Medication Exception Date:

If the patient has a documented aspirin product or anti-platelet medication exception enter the date of the contraindication or exception.

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

Patients ages 18 to 75 with a diagnosis of diabetes (Diabetes Value Set) with any contact during the current or prior measurement period OR had diabetes (Diabetes Value Set) present on an active problem list at any time during the measurement period. Both contacts AND problem list must be queried for diagnosis (Diabetes Value Set).

AND patient has at least one established patient office visit (Established Pt Diabetes & Vasc Value Set) performed or supervised by an eligible provider in an eligible specialty for any reason during the measurement period.

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).

Please also refer to all code lists included in the data dictionary attached in S.2b.

• 18 years or older at the start of the measurement period AND less than 76 years at the end of the measurement period

• Patient had a diagnosis of diabetes (Diabetes Value Set) with any contact during the current or prior measurement period OR had diabetes (Diabetes Value Set) present on an active problem list at any time during the measurement period. Both contacts AND the active problem list must be queried for diagnosis (Diabetes Value Set).

• At least one established patient office visit (Established Pt Diabetes & Vasc Value Set) performed or supervised by an eligible provider in an eligible specialty for any reason during the measurement period

Eligible specialties: Family Medicine, Internal Medicine, Geriatric Medicine, Endocrinology

Eligible providers: Medical Doctor (MD), Doctor of Osteopathy (DO), Physician Assistant (PA), Advanced Practice Registered Nurses (APRN)

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

Valid allowable exclusions include patients who were a permanent resident of a nursing home, pregnant, died or were in hospice or palliative care during the measurement year.

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.)*

• Patient was pregnant during measurement period (ICD-10 O24.011, O24.012, O24.013, O24.019, O24.02, O24.03, O24.111, O24.112, O24.113, O24.119, O24.12, O24.13, O24.311, O24.312, O24.313, O24.319, O24.32, O24.33, O24.811, O24.812, O24.813, O24.819, O24.82, O24.83, O24.911, O24.912, O24.913, O24.919, O24.92, O24.93

• Patient was a permanent nursing home resident during the measurement period

• Patient was in hospice or palliative care at any time during the measurement period,

• Patient died prior to the end of the measurement period

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.)*

The diabetes population is not currently stratified when publicly reported on our consumer website, MN HealthScores. The data is, however, stratified by public (MN Health Care Programs- Prepaid Medical Assistance including dual eligibles, MinnesotaCare, and General Assistance Medical Care) and private purchasers for our 2017 Health Care Disparities Report. This report notes a gap in outcomes of fifteen percentage points between diabetic patients in public programs and other purchasers. http://mncm.org/wp-content/uploads/2018/03/2017-Disparities-Report-FINAL-3.26.2018.pdf

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

Statistical risk model

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Higher score

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

This measure is calculated by submitting a file of individual patient values (e.g. blood pressure, A1c value, etc.) to a HIPAA secure data portal. Programming within the data portal determines if each patient is a numerator case and then a rate is calculated for each clinic site. Please also refer to the measure calculation algorithms submitted within the data dictionary for this measure.

If any component of the numerator is noncompliant for any one of the five components, then the patient is numerator noncompliant for the composite patient level all-or none optimal diabetes care measure.

Numerator logic is as follows:

A1c Component:

Is the HbA1c date in the measurement period? If no, is numerator noncompliant for this component. If yes, assess next variable.

Is the HbA1c value less than 8.0? If yes, is numerator compliant for this component. If no, is numerator noncompliant for this component.

Note: A1c needs to occur during the measurement year AND most recent value less than 8.0

Assess next component.

Blood Pressure Component:

Is Blood Pressure date in the measurement period? If no, is numerator noncompliant for this component. If yes, assess next variable.

BP Systolic < 140? If no, is numerator noncompliant for this component. If yes, assess next variable.

BP Diastolic < 90? If yes, is numerator compliant for this component. If no, is numerator noncompliant for this component.

Note: BP needs to occur during the measurement year AND most recent BP systolic less than 140 AND BP diastolic less than 90

Assess next component.

Cholesterol Statin Use Component:

Is the patient on a statin medication? If yes, and most recent date is in the measurement year, is numerator compliant for this component. If no, assess next variable.

For patients not on a statin the following variables are used to assess numerator compliance related to contraindications or exceptions to statin use:

Is the patient age 18 to 20? If yes, numerator compliant (free-pass), if no, assess next variable.

Is the patient age 21 to 75? Do they have ischemic vascular disease (IVD)?

If Yes IVD, is their most recent LDL in the last five years less than 40? If Yes, numerator compliant (free-pass), if no, assess next variable.

Does the patient have a valid contraindication/ exception to statin use defined as one of the following: pregnancy, active liver disease, rhabdomyolysis, ends stage renal disease on dialysis, heart failure, breastfeeding, allergy to statin, drug-drug interaction with statin, or intolerance with documentation of trying a statin at least once in the last 5 years)? If yes, is numerator compliant for this component. If no, fail this numerator component and remains in the denominator.

If No IVD, is the patient age 21 to 39 and is their most recent LDL in the last 5 years greater than or equal to 190? If No, numerator compliant (free-pass).

If Yes LDL greater than or equal to 190, does the patient have a valid contraindication/ exception to statin use defined as one of the following: pregnancy, active liver disease, rhabdomyolysis, ends stage renal disease on dialysis, heart failure, breastfeeding, allergy to statin, drug-drug interaction with statin, or intolerance with documentation of trying a statin at least once in the last 5 years)? If yes, is numerator compliant for this component. If no, fail this numerator component and remains in the denominator.

If No IVD, no LDL greater than or equal to 190 for patients ages 40 to 70, is their most recent LDL in the last five years less than 70? If Yes, numerator compliant (free-pass), if no, assess next variable.

Does the patient have a valid contraindication/ exception to statin use defined as one of the following: pregnancy, active liver disease, rhabdomyolysis, ends stage renal disease on dialysis, heart failure, breastfeeding, allergy to statin, drug-drug interaction with statin, or intolerance with documentation of trying a statin at least once in the last 5 years)? If yes, is numerator compliant for this component. If no, fail this numerator component and remains in the denominator.

Note: Patient is either on a statin (prescribed/ ordered) during the measurement year or has a valid exception either by age, presence or absence of ischemic vascular disease, low untreated LDL or valid contraindication/ exception.

Assess next component.

Tobacco-Free Component:

Is Tobacco Status = 1 (Tobacco Free) and Tobacco Assessment Date a valid date? If yes, is numerator compliant for this component. If no, is numerator noncompliant for this component. Assess next component.

Daily Aspirin/ Anti-platelet Component:

Does the patient have cardiovascular/ ischemic vascular disease? If no, is numerator compliant (free-pass), if yes assess next variable.

Is the patient on daily aspirin or an antiplatelet? If yes, and date of most recent aspirin/ anti-platelet is in the measurement year is numerator compliant, if no, assess next variable.

Does the patient have a valid contraindication/ exception to aspirin anti-platelet use defined as one of the following: anti-coagulant medication, history of gastrointestinal bleed, history of intracranial bleed, allergy, or physician documented reasons related to: risk of drug interaction, use of NSAIDS, uncontrolled HTN or gastro-intestinal reflux disease. If yes, is numerator compliant for this component. If no, fail this numerator component and remains in the denominator.

Note: Patients with ischemic vascular disease are either on daily aspirin (indicated/ prescribed/ ordered) or an anti-platelet prescribed/ ordered) during the measurement year or has a valid contraindication/ exception.

If all of the above numerator components are in compliance, then the patient calculated as a numerator case for the optimal diabetes care measure.

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.

Medical groups are encouraged to submit their full population of patients when possible (EMR) however clinics who are on a paper chart system are allowed to create a random sample of no less than 60 patients per clinic site. MNCM recommends that medical groups submit total population for each measure. By submitting total population, the confidence interval around the rate narrows, indicating a higher confidence that the rate accurately reflects the clinics' performance. If total population is not an option for a medical group, MNCM encourages medical groups to submit a large sample. The minimum required sample is 60 patients per clinic site, per measure. If a clinic site has less than 60 patients in the total population for the measure, the entire population must be submitted.

For 2017 dates of service 99.5% (615 clinics) submitted total population, 0.5% (3 clinics) submitted a sample.

Excel's Random Number Generator Instructions:

For lists generated in Excel, use the "RAND" function to assign a random number to each record (please also see Microsoft Excel Help, topic RAND for more information):

1. Insert a blank column on the leftmost side of the spreadsheet

2. Label new column "RAND"

3. Place cursor in the first blank cell (A2) and type =RAND()

4. Press enter (a number like 0.793958 will appear)

5. Place the cursor back into this cell; resting over the corner to have the pointer change to a black cross, double click or drag the formula down to the last row/patient

6. Highlight the whole column and click Edit, Copy, Paste Special = Values to freeze the random number (otherwise it will change with every click on the spreadsheet)

7. Sort entire patient population by this new random number

8. Work down the list row by row, starting with row 1 until the number of records in the sample is met for submission (at least 60 patients per clinic, per measure)

9. If a patient meets one of the accepted exclusions, keep working down the list and use oversamples that are after the number of records in the sample. For example, if 60 records will be submitted and 2 exclusions were found, include patient rows 61 and 62 to replace the excluded records.

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

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

Not applicable; measure is not based on a survey or instrument.

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

If other, please describe in S.18.

Electronic Health Records, 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.

An excel template with formatted columns for data fields is provided. Almost all medical groups in MN (99.5%) extract the information from their EMR. Paper abstraction forms are provided for those clinics who wish to use them as an interim step to create their data file. All data is uploaded in electronic format (.csv file) to a HIPAA secure, encrypted and password protected data portal. We capture information from the clinics about how their data is obtained. In 2018:

• 71% (476) clinics had an EMR and pulled all data via query

• 26% (176) clinics had an EMR and used a combination of query and manual look up for data collection

• 2.2% (15) clinics had an EMR and looked up all data manually

• 0.15% (1) clinic had a hybrid EMR and paper record system

• 0.15% (1) clinic had paper records only

Feasibility Note: 71% of practices can extract all of the information needed via query.

Please note that all fields are defined and included in the data dictionary [Tab = Data Field Dictionary] and also included in the data collection guide URL provided in S.1.

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)*

Available at measure-specific web page URL identified in S.1

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

Clinician : Group/Practice

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.)*

Calculation of the patient level all-or-none composite measure algorithms are provided in the data dictionary. The individual components are treated equally (not weighted). Please note that while the all-or-none composite measure is considered to be the gold standard, reflecting best patient outcomes, the individual components may be measured as well. This is particularly helpful in quality improvement efforts to better understand where opportunities exist in moving the patients toward achieving all the desired outcomes. Individual component logic is included below:

Denominator is the same for calculating individual component rates as the patient level all-or-none composite measure:

Patients ages 18 to 75 with a diagnosis of diabetes (Diabetes Value Set) with any contact during the current or prior measurement period OR had diabetes (Diabetes Value Set) present on an active problem list at any time during the measurement period. Both contacts AND problem list must be queried for diagnosis (Diabetes Value Set).

AND patient has at least one established patient office visit (Established Pt Diabetes & Vasc Value Set) performed or supervised by an eligible provider in an eligible specialty for any reason during the measurement period.

Component for HbA1c Control:

Is the HbA1c date in the measurement year? If No, fails the numerator. If Yes, assess next variable.

Is the most recent HbA1c value less than 8.0? If Yes, is in the numerator for this component.

Expressed as a rate:

Patients with most recent A1c during the measurement year is less than 8.0/

Eligible patients with diabetes

Component for Blood Pressure Control:

Is the BP date in the measurement year? If No, fails the numerator. If Yes, assess next variable.

Is the most recent BP value less than 140 systolic AND less than 90 diastolic? If Yes, is in the numerator for this component.

Expressed as a rate:

Patients with most recent BP during the measurement year is less than 140 systolic AND 90 diastolic/

Eligible patients with diabetes

Component for Cholesterol/ Statin Use:

Is the patient on a statin medication? If yes, and most recent date is in the measurement year, is in the numerator for this component.

For patients not on a statin the following variables are used to assess numerator compliance related to contraindications or exceptions to statin use:

Is the patient age 18 to 20? If yes, in the numerator (free-pass), if no, assess next variable.

Is the patient age 21 to 75? Do they have ischemic vascular disease (IVD)?

If Yes IVD, is their most recent LDL in the last five years less than 40? If Yes, in the numerator(free-pass), if no, assess next variable.

Does the patient have a valid contraindication/ exception to statin use defined as one of the following: pregnancy, active liver disease, rhabdomyolysis, ends stage renal disease on dialysis, heart failure, breastfeeding, allergy to statin, drug-drug interaction with statin, or intolerance with documentation of trying a statin at least once in the last 5 years)? If yes, is in the numerator. If no, fail this numerator component and remains in the denominator.

If No IVD, is the patient age 21 to 39 and is their most recent LDL in the last 5 years greater than or equal to 190? If No, is in the numerator (free-pass).

If Yes LDL greater than or equal to 190, does the patient have a valid contraindication/ exception to statin use defined as one of the following: pregnancy, active liver disease, rhabdomyolysis, ends stage renal disease on dialysis, heart failure, breastfeeding, allergy to statin, drug-drug interaction with statin, or intolerance with documentation of trying a statin at least once in the last 5 years)? If yes, is in the numerator. If no, fail this numerator component and remains in the denominator.

If No IVD, no LDL greater than or equal to 190 for patients ages 40 to 70, is their most recent LDL in the last five years less than 70? If Yes, is in the numerator (free-pass), if no, assess next variable.

Does the patient have a valid contraindication/ exception to statin use defined as one of the following: pregnancy, active liver disease, rhabdomyolysis, ends stage renal disease on dialysis, heart failure, breastfeeding, allergy to statin, drug-drug interaction with statin, or intolerance with documentation of trying a statin at least once in the last 5 years)? If yes, is in the numerator. If no, fail this numerator component and remains in the denominator.

Expressed as a rate:

Patients with statin use unless with contraindications/ exceptions/

Eligible patients with diabetes

Component for Tobacco-Free:

Is the date of smoking status in the measurement year or the year prior? If No, fails the numerator. If Yes, assess next variable.

Is the patient's tobacco status noted as tobacco-free ? If Yes, is in the numerator.

Expressed as a rate:

Patients with most recent tobacco status during the measurement year or the year prior is free of all tobacco products (tobacco free)/

Eligible patients with diabetes

Component for Daily Aspirin/ Anti-platelet Component:

Does the patient have cardiovascular/ ischemic vascular disease? If No, is in the numerator (free-pass)

Is the patient on daily aspirin or an antiplatelet? If yes, and date of most recent aspirin/ anti-platelet is in the measurement year is numerator compliant, if no, assess next variable.

Does the patient have a valid contraindication/ exception to aspirin anti-platelet use defined as one of the following: anti-coagulant medication, history of gastrointestinal bleed, history of intracranial bleed, allergy, or physician documented reasons related to: risk of drug interaction, use of NSAIDS, uncontrolled HTN or gastro-intestinal reflux disease. If yes, is numerator compliant for this component. If no, fail this numerator component and remains in the denominator.

Expressed as a rate:

Patients with ischemic vascular disease with daily aspirin/ anti-platelet use unless with contraindications/ exceptions/Eligible patients with diabetes

2. Validity – See attached Measure Testing Submission Form

composite_testing_attachment_2017_0729_mncm_8-16-2018.docx

2.1 <u>For maintenance of endorsement</u>

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 <u>For maintenance of endorsement</u>

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 <u>For maintenance of endorsement</u>

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.

Yes - Updated information is included

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

Measure Number (*if previously endorsed*)**:** 0729

Composite Measure Title: Optimal Diabetes Care
Date of Submission: 8/1/2018

Composite Construction:

☐ Two or more individual performance measure scores combined into one score

☒ All-or-none measures (e.g., all essential care processes received or outcomes experienced by each patient)

Instructions: Please contact NQF staff before you begin.

- If a component measure is submitted as an individual performance measure, the non-composite measure testing form must also be completed and attached to the individual measure submission.
- 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.
- **Sections 1, 2a2, 2b1, 2b2, and 2b4 must be completed.**
- **For composites with 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) and composites (2c) 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. and the 2017 Measure Evaluation Criteria and Guidance.

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 decision making) 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 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 non-responders) and how the specified handling of missing data minimizes bias.

2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:

2c1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and

2c2.the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.

(*if not conducted or results not adequate, justification must be submitted and accepted)*

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 different components in the composite, indicate the component after the checkbox. 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*).

[2014] Existing data set was used. Patient level data was submitted from 118 medical groups representing 580 clinics in Minnesota and bordering communities. Data submission is an annual process; for 2013 dates of service reported in 2014, 230,818 patients were submitted for rate calculation and this represents 97.2% of all eligible diabetic patients in MN. Sampling is allowed for those clinics with paper records or those who have not had their EMR in place for one year prior to the measurement period; in MN there are very few clinics not yet on an EMR. For 2013 dates of service 91% submitted total population, 7% submitted a sample, and 2% submitted a mix of total and sample. Data submission for this measure is mandatory in the state of MN by 2008 health reform legislation and the MN Department of Health Statewide Quality and Reporting Measurement System.

Patients: 230,818

Medical Groups: 118

Individual Clinics: 580

Patient level data files are submitted by medical groups to a HIPAA secure data portal for rate calculation.

Types of fields included in the submission for 2013/ 2017 dates of service include the following:

Clinic ID ▪ Patient ID ▪ Patient Date of Birth ▪ Patient Gender Zip Code, Primary Residence ▪ Race/Ethnicity1 ▪ Race/Ethnicity2 ▪ Race/Ethnicity3 ▪ Race/Ethnicity4 ▪ Race/Ethnicity5 ▪ Country of Origin Code ▪ Country of Origin "Other" Description ▪ Preferred Language Code ▪ Preferred Language "Other" Description ▪ Provider

NPI ▪ Provider Specialty Code ▪ Insurance Coverage Code ▪ Insurance Coverage "Other" Description ▪ Insurance Plan Member ID ▪ Patient Has IVD? ▪ Patient Has Depression? ▪ Type 1 or Type 2 Diabetes? ▪ HbA1c Date ▪ HbA1c Value ▪ LDL Date ▪ LDL Value ▪ BP Date ▪ BP Systolic ▪ BP Diastolic ▪ Statin Medication Date ▪ Statin Medication Exception ▪ Statin Medication Exception Date ▪ Aspirin or Antiplatelet Medication ▪ Aspirin or Antiplatelet Medication Date ▪ Aspirin or Antiplatlet Medication Exception ▪ Aspirin or Antiplatelet Medication Exception Date ▪ Tobacco Status Documentation Date ▪ Tobacco Status

2018 Update

Due to change in evidence and guidelines for cholesterol management, one component of this measure was redesigned in 2014. At the time of this measure's last maintenance endorsement the plan to redesign the component from LDL < 100 to appropriate use of statin was in place, specs completed but MNCM did not yet have data for the redesigned component. Update in testing is provided and does reflect the composite measure with the redesigned component.

Patient level data was submitted from 103 medical groups representing 618 reportable clinics (n ≥ 30) in Minnesota and bordering communities. Data submission is an annual process; for 2017 dates of service reported in 2018, 307,158 patients were submitted for rate calculation and this represents 99.8% of all eligible diabetic patients in MN.

Sampling is allowed for those clinics with paper records or those who have not had their EMR in place for one year prior to the measurement period; in MN there are very few clinics not yet on an EMR. For 2017 dates of service 99.5% submitted total population, 0.5% submitted a sample.

1.3. What are the dates of the data used in testing? 1/1/2017 to 12/31/2017

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*)

[2014] Patients: 230,818

Medical Groups: 118

Individual Clinics with ≥ 30 eligible: 580 clinic sites and 229,806 patients

Individual clinics with < 30 eligible: 77 clinic sites with 1,012 patients

Includes all primary care and endocrinology clinics in MN; rates for public reporting and associated reliability and clinic level statistics only include those clinic sites with 30 or more eligible diabetic patients meeting denominator criteria; in 2013 this was 580 clinics representing over 229,000 patients. All patients submitted are used to calculate statewide averages, risk adjustment models and aggregate descriptive statistics.

2018 Update

Patients: 307,158

Medical Groups: 103

Clinics: 669

Clinics with ≥ 30 eligible: 618 clinic sites and 306,509 patients

Clinics with < 30 eligible: 51 clinic sites with 649 patients

Includes all primary care and endocrinology clinics in MN; rates for public reporting and associated reliability and clinic level statistics only include those clinic sites with 30 or more eligible diabetic patients meeting denominator criteria; in 2017 dates of service/ 2018 report year this was 618 clinics representing over 306,000 patients. All patients submitted are used to calculate statewide averages, risk adjustment models and aggregate descriptive statistics.

Level of analysis is the clinic site level and the data source is patient level discrete fields submitted from the clinic's electronic medical records systems.

- 476 clinics had an EMR and pulled all data via query
- 176 clinics had an EMR and used a combination of query and manual look up for data collection
- 15 clinics had an EMR and looked up all data manually
- 1 clinic had a hybrid EMR and paper record system
- 1 clinic had paper records only

71% of practices are able to extract all of the information needed via query.

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*)

[2014] Patients: 230,818

Medical Groups: 118

Individual Clinics with ≥ 30 eligible: 580 clinic sites and 229,806 patients

Individual clinics with < 30 eligible: 77 clinic sites with 1,012 patients

Includes all primary care and endocrinology clinics in MN; rates for public reporting and associated reliability and clinic level statistics only include those clinic sites with 30 or more eligible diabetic patients meeting denominator criteria; in 2013 this was 580 clinics representing over 229,000 patients. All patients submitted are used to calculate statewide averages, risk adjustment models and aggregate descriptive statistics.

2018 Update

Patients: 307,158

Medical Groups: 103

Clinics: 669

Clinics with ≥ 30 eligible: 618 clinic sites and 306,509 patients

Clinics with < 30 eligible: 51 clinic sites with 649 patients

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.

Clinic level analysis of rates is conducted only for clinics with ≥ 30 patients (n = 618 clinics)

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.

The social risk factors that were available and analyzed include insurance product type and deprivation index as proxies for socioeconomic status and race, Hispanic ethnicity, preferred language and country of origin (RELO) data.

- Health insurance coverage information is included in the patient level file that is submitted from the medical group and is translated to specific insurance product type (commercial, Medicare, Medicaid, uninsured and unknown). Insurance product type has demonstrated properties for inclusion in risk adjustment models (p-values < 0.01 to 0.02).
- The deprivation index is a calculation based on US Census Data at the patient's zip code level that considers the percentage of people in that zip code with supplemental nutrition assistance program (SNAP) benefits, in poverty, unemployed, on public assistance and single females with children. The five census variables are centered to zero and are run through a factor analysis to create a single deprivation index for each patient.
- MNCM considered the inclusion of race, Hispanic ethnicity, preferred language and country of origin (RELO) as potential risk adjustment variables. While there are significant differences in outcome when segmenting the data by RELO variables, these variables are not used in risk adjustment because it is impossible to separate the disparity in outcome between the patient's environment and the clinic's contribution to the disparity specific biases from healthcare providers that influence their interactions with patients.
    - MNCM convened a panel of social science researchers from the University of Minnesota in August 2016 to understand if RELO variables were confounded by the clinic's contribution. For every measure and for every factor, as a group, the researchers were hard pressed to find a conceptual relationship that was not confounded by the clinic's contribution. This is confirmed in the social science literature available on the topic.
    - The panel recommended use of geography instead, specifically neighborhood characteristics which lead to the development of the deprivation index mentioned above.

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*)

Note*: Current guidance for composite measure evaluation states that reliability must be demonstrated for the composite performance measure score.*

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. 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*)

[2014] Used paper "Reliability in Provider Profiling" by John L. Adams, Ph.D as a reference

The BETABIN macro was used on each measure (SAS).

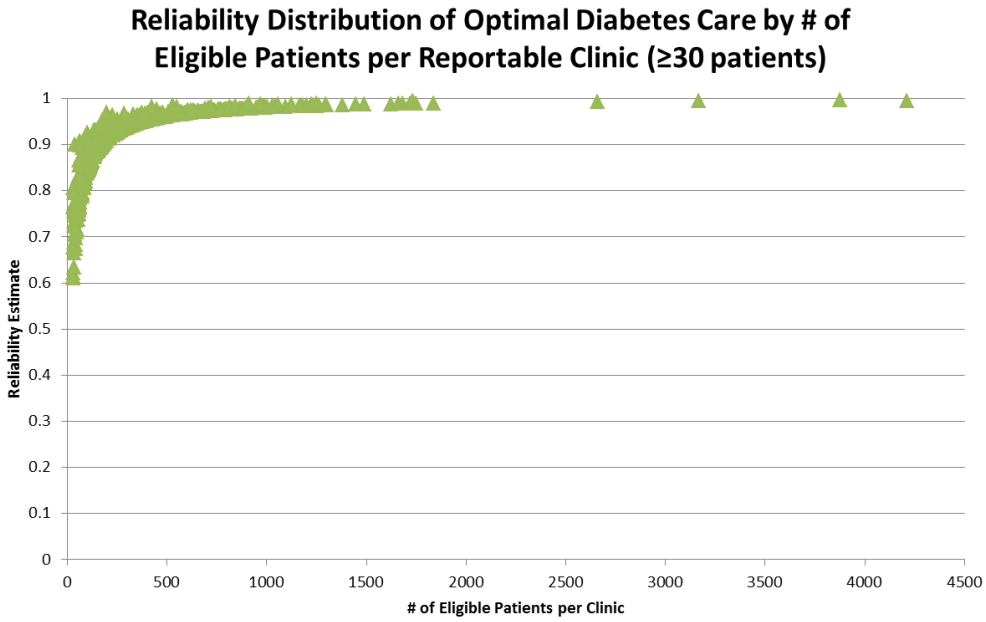
- First, we need to find the provider-to-provider variance:
    - $\sigma^2 = (\alpha\ \beta) / (\alpha + \beta + 1)(\alpha + \beta)^2$
    - $= (6.3684 * 11.1559) / (6.3684 + 11.1559 + 1)(6.3684 + 11.1559)^2$
    - = **0.0125** (plug this value into the reliability equation)
- Reliability = $\sigma^2 / (\sigma^2 + (p(1 - p)/n))$
    - p = rate
    - n = number of eligible patients
- Determine reliability rate for each provider.

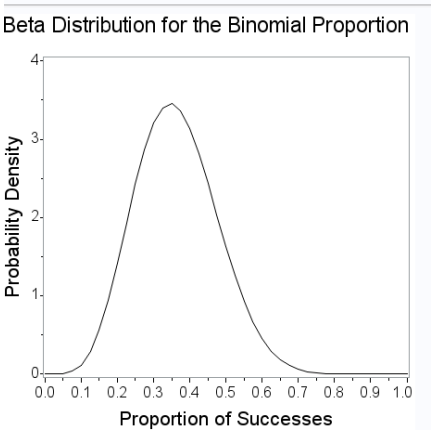
- Average the reliability rate.

Beta-binomial model: 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.

2a2.3. 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*)

[2014] Reliability = 0.908





BETABIN Macro: Simple Binomial Model

parameter	Estimate	Standard Error	t Value	Prob > |t|	Alpha	Lower	Upper
mu	0.3925	0.0010	385.34	<.0001	0.05	0.3905	0.3945
|mu-0.5|	0.1075	0.0010	105.51	<.0001	0.05	0.1055	0.1095

BETABIN Macro: Beta-Binomial Model Parameters

parameter	Estimate	Standard Error	t Value	Pr > |t|	Alpha	Lower	Upper
mu	0.3634	0.004878	74.50	<.0001	0.05	0.3538	0.3730
alpha	6.3684	0.4155	15.33	<.0001	0.05	5.5540	7.1828
beta	11.1559	0.7266	15.35	<.0001	0.05	9.7317	12.5800
gamma	0.05398	0.003288	16.42	<.0001	0.05	0.04754	0.06043
theta	0.05706	0.003674	15.53	<.0001	0.05	0.04986	0.06426
|mu-0.5|	0.1366	0.004878	28.00	<.0001	0.05	0.1270	0.1462

BETABIN Macro: Variance-Covariance Matrix of Estimated Parameters

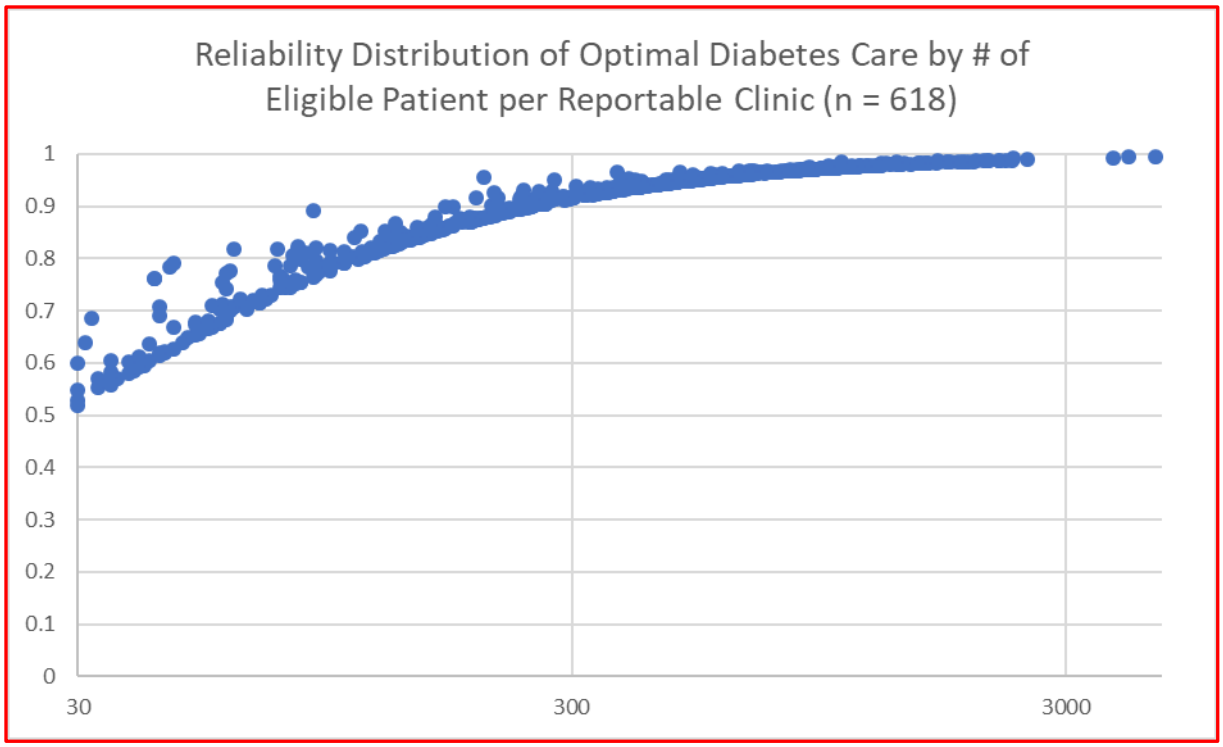
Label	mu	alpha	beta	gamma	theta
mu	0.000024	0.000338	-0.00055	6.298E-7	7.037E-7
alpha	0.000338	0.1727	0.2861	-0.00134	-0.00149
beta	-0.00055	0.2861	0.5280	-0.00237	-0.00265
gamma	6.298E-7	-0.00134	-0.00237	0.000011	0.000012
theta	7.037E-7	-0.00149	-0.00265	0.000012	0.000013

2018 Updated Reliability Statistics

618 clinics with 306,509 patients

Logarithmic Scale

Average Reliability: 0.888



Descriptive Statistics for Reliability Scores 618 observations (clinics)

Mean	0.887818
Median	0.929769
Standard Deviation	0.102883
Minimum	0.519081
Maximum	0.993923

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?*)

[2014] In terms of understanding reliability in detecting signal to noise, a reliability score of 0.70 or greater is considered acceptable for drawing conclusions about groups. This data analysis, along with precise specifications and excellent validation results of critical data elements, demonstrates this measure construct to be reliable and detect meaningful differences among provider groups.

_____

2b1. VALIDITY TESTING

Note*: Current guidance for composite measure evaluation states that validity should be demonstrated for the composite performance measure score. If not feasible for initial endorsement, acceptable alternatives include assessment of content or face validity of the composite OR demonstration of validity for each component. Empirical validity testing of the composite measure score is expected by the time of endorsement maintenance.*

2b1.1. What level of validity testing was conducted?

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

☒ **Composite 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.

☒ **Validity testing for component measures** *(check all that apply)*

Note*: applies to ALL component measures, unless already endorsed or are being submitted for individual endorsement.*

☐ **Endorsed (or submitted) as individual performance measures**

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

☒ **Empirical validity testing of the component measure score(s)**

☐ **Systematic assessment of face validity of component measure score(s) 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*)

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)*

[2014]

Critical Data Elements

Validating the submitted data via the direct data submission process is completed in four steps: denominator certification, data quality checks, validation audit, and the two-week medical group review period.

Denominator certification prior to data collection and extraction/ abstraction ensures that all medical groups apply the denominator criteria correctly and in a consistent manner. MNCM staff review the documentation to verify all criteria were applied correctly, prior to approval for data submission.

Denominator certification documentation for this measure includes:

- Date of Birth (ranges)
- Date of Service (ranges)
- ICD-9 /ICD-10 Codes used

- Eligible specialties and provider types
- Exclusions to the measure and attest to mechanism for exclusions
- Attestations related to changes in medical record or billing systems
- Supplying all query code for review

Common areas of correction in denominator for this measure included missing query code, incorrect date of birth ranges, incorrect dates for counting visits, missing ICD-9 codes or incomplete attestation. All were corrected prior to data submission.

Following data submission to the MNCM Data Portal, there are additional data quality checks in place for evaluating the accuracy of data submitted. During file upload, program checks for valid dates, codes and values and presents users with errors and warnings. Additionally, MNCM staff review population counts (denominator) and outcome rates for any significant variance from the previous year's submission and may prompt further clarification from the medical group.

Validation audits verify that the clinical data submitted for the numerator component of the measure matched the data in the patient record. Other data elements are also audited to verify the patient was included in the denominator correctly (e.g., diagnosis of depression).

In 2014, for the diabetes measure, MNCM audited 128 medical groups; 76% of those submitting data. 85% passed the initial audit, 15% required a correction plan and all re-submitted their data and passed the audit with $\geq$ 90% accuracy. Types of discrepancies noted on audit included: not including most recent values, aspirin/ anti-platelet dates not in the measurement year and incorrect tobacco status.

[2018] ICD-10 Coding

In 2014 maintenance submission for this measure the ICD-9 and ICD-10 code lists were provided as a part of our data dictionary attachment S.2b. This process was achieved by 1) reviewing the code cross walk systems available at that time (American Medical Association and ICD-10Data.com) and 2) clinical review of ICD-10 codes that best fit the intended denominator (patients with type 1 and type 2 diabetes). In our experience, cross walks are not always effective and accurate in translating ICD-9 code definitions to ICD-10. The crosswalks tend to only go in one direction from ICD-9 to ICD-10 mapping one code to many in the new coding system, however the reverse cross-walk (ICD-10 to ICD-9) doesn't always give the same results. The codes selected were reviewed by a clinical (RN) measure developer for best fit with measure intent. Additionally, apart from excluding pregnant diabetes patients, our ICD-10 value set aligns with another major measure developer (NCQA)

Volumes for this measure:	**Patients**	**# Clinics**	**Coding System**
2012 dates of service/ 2013 report year:	208,809	613 clinics	ICD-9
2013 dates of service/ 2014 report year:	230,818	657 clinics	ICD-9
2014 dates of service/ 2015 report year:	245,241	664 clinics	3/4 ICD-9; 1/4 ICD-10
2015 dates of service/ 2016 report year:	257,078	684 clinics	ICD-10

Analysis of empirical data demonstrates appropriate increases in denominator based on increasing numbers of clinics reporting and support the continuity of definition over the change in coding systems.

Additionally, as part of our annual validation process described above and below, the auditors verify the diagnosis of diabetes for each patient record audited and there have been no errors identified related to denominator identification by data submitters.

[2018] Validation Audit Results

In 2018, for the diabetes measure, MNCM audited 53 medical groups; 37% of those submitting data. 89% passed the initial audit, 11% required a correction plan and all re-submitted their data and passed the audit with ≥ 90% accuracy. Please note that all data elements comprising each component of the composite measure are reviewed against the medical record.

Types of discrepancies and # with issue noted on audit: blood pressure date or values (5), incorrect diabetes type (2), incorrect tobacco status (2), statin medication or date (2), HbA1c date or values, (2) LDL date or values (2), IVD co-morbidity diagnosis (1), and exception for statin or aspirin (1)

Composite Performance Measure Score

Validity was tested for the computed composite score by testing the correlation of medical group performance with their performance on the Optimal Vascular Care measure (NQF# 0076). Ischemic vascular disease and diabetes are chronic conditions that require ongoing management of multiple risk factors to reduce a patient's overall risk of developing long term complications. It is expected that the quality of care provided by a medical group to patients with diabetes would be of similar quality as the care provided to patients with ischemic vascular disease, and the respective performance measure scores should demonstrate this.

Validity Testing Component Measures

Validity was tested for the individual components of the composite measure with correlation analysis (CORR Procedure) and the analysis of rates of all possible combinations of components (e.g. each individual component, aspirin + blood pressure, tobacco + A1c, etc.)

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

Critical Data Elements

[2014/2018] 100% of groups achieved the desired > 90% data accuracy when submitted data was compared to medical record data (EMR or paper) of the patient.

Composite Performance Measure Score

2018 Report Year/ 2017 Dates of Service

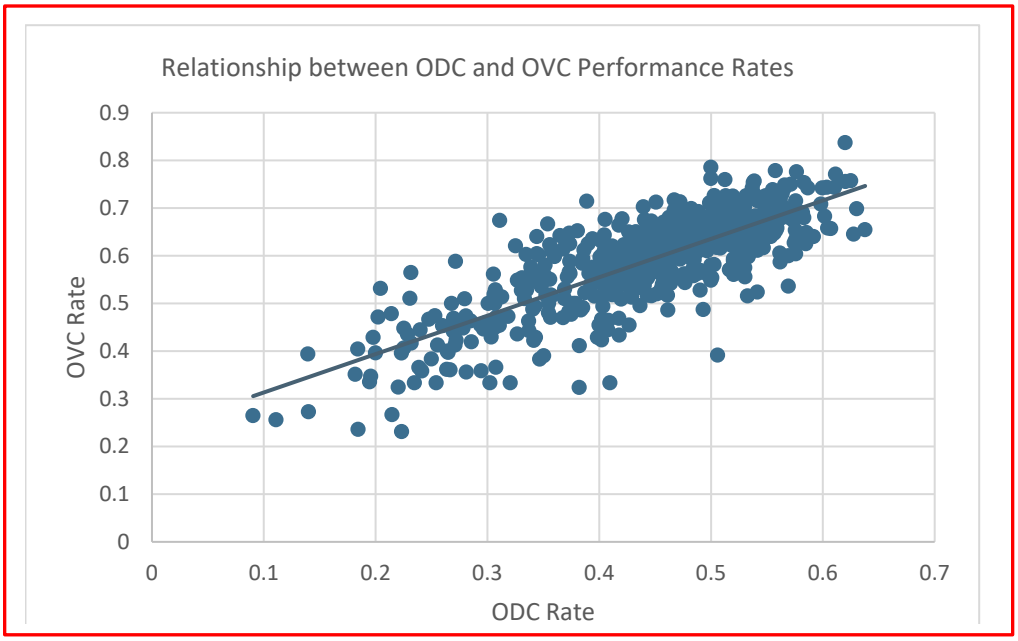
618 clinics with 306,509 patients

Correlation of performance with another similar chronic disease composite measure

ODC = Optimal Diabetes Care and OVC = Optimal Vascular Care

[2018]

r^2 value = 0.629



Validity Testing Component Measures

Please refer to question 2c. Empirical Analysis to Support Composite Construction Approach.

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?*)

[2014] High compliance with critical data element validity as demonstrated by annual validation audit processes.

[2018] Based on linear regression analysis, a medical group's performance on the Optimal Diabetes Care measure is associated with its performance on the Optimal Vascular Care measure, as demonstrated by an r^2 value of .629, representing a fairly strong correlation.

Validity testing of the individual components demonstrate strong correlation between the components and the composite measure. Pearson r coefficient values: HbA1c at 0.77714, Blood Pressure at 0.71245, Tobacco Free at 0.54201, Aspirin or Antiplatelet Use at 0.26253, and Statin Use at 0.68083

2b2. EXCLUSIONS ANALYSIS

Note: *Applies to the composite performance measure, as well all component measures unless they are already endorsed or are being submitted for individual endorsement.*

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

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

[2014] Medical groups submitting patient level data to MNCM have the option of either excluding patients appropriately via their EMR query process and provide an attestation through our process of denominator certification and providing their query code for MNCM staff to review and/or submitting a file of excluded patients. Exclusion testing was performed on a sample of groups who submitted files of patients they excluded from the measure. Sample included 11 medical groups representing metro, rural, endocrinology, federally qualified health care centers and teaching/ tertiary. Included were 232 clinics and over 109,000 diabetic patients with dates of service 1/1/2013 to 12/31/2013.

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*)

[2014]

Medical Group Code	Clinics	Type	Diabetic Patients	Nursing Home	Hospice	Deceased	Coded in error	Pregnancy	Total Exclusions
Medical Group A	51	Metro	32,110	0	0	266	0	118	384
Medical Group B	8	Rural	2,588	2	2	33	0	1	38
Medical Group C	12	Rural	3,757	24	1	52	0	8	85
Medical Group D	1	Metro; Endo	80	0	0	1	0	0	1
Medical Group E	39	Metro	17,289	0	0	124	0	30	154
Medical Group F	28	Metro	15,182	61	64	3	0	51	179
Medical Group G	13	Metro; FQHC	5,461	3	1	0	0	1	5
Medical Group H	12	Rural	2,490	0	0	20	0	1	21
Medical Group I	23	Metro	17,628	6	6	95	2	88	197
Medical Group J	44	Metro/ Rural	9,733	34	0	1	0	17	52
Medical Group K	1	Metro/ Teaching	3,346	0	0	1	7	8	16
	232		**109,664**	**130**	**74**	**596**	**9**	**323**	**1,132**
				0.1%	**0.1%**	**0.5%**	**0.01%**	**0.3%**	**1.0%**

2018 Update

During the CMS Measure Under Consideration process, the exclusion for "diagnosis coded in error" was removed from the measure specification. Note the above very low occurrence (0.01%). This exclusion had a purpose prior to 2015 because in the ICD-9 classification of disease there was no code to signal a pre-diabetes state to support billing for an A1c test and patients with pre-diabetes were given a 250.xx code for diabetes. With the advent of ICD-10, there is now a code for pre-diabetes [R73.03] and this exclusion is no longer necessary.

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*)

[2014] While the exclusions to this measure are have clinical importance related to patient safety in achieving targets or utilizing medications to reduce cardiovascular risk, the total number of exclusions is relatively small, 1.0% and therefore do not significantly impact measure performance.

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

***Note**: Applies to all outcome or resource use component measures, unless already endorsed or are being submitted for individual endorsement.*

***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? *(check all that apply)*

☐ **Endorsed (or submitted) as individual performance measures**

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with 4 risk factors**

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

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

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

Risk factors are

- patient age (continuous variable)
- insurance product (proxy for socioeconomic status)
- diabetes type (1, 2 or unknown)

- deprivation index (proxy for socioeconomic status based on 5-digit zip code)

  Comprised of percentage with SNAP benefits, percentage in poverty, percentage unemployment, percentage on public assistance and percentage single female with child

Since our outcome (dependent) variable is binary (yes/no optimal care was obtained), we use a logistic regression model with the following risk factors included:

- patient age as a continuous variable
- insurance product type as a categorical variable including commercial, Medicare, Medicaid, uninsured, and unknown insurance type as categories, commercial is the reference group in the model, this variable is a proxy for socioeconomic status
- diabetes type as a categorical variable including type 1, type 2 and unknown diabetes type, type 2 is the reference group
- deprivation index as a continuous variable, this variable is a proxy for socioeconomic status based on patient 5-digit zip code, it considers the percentage of people in that 5-digit zip code with SNAP benefits, in poverty, unemployed, on public assistance and single females with children using US Census Data.

Indirect standardization is used for risk adjustment. In this method, **the actual clinic result is not changed**, no matter the degree of patient risk. Instead, an expected value is calculated for each clinic using the logistic regression model run at the patient level and results are aggregated to the clinic level as described above. In this process, the clinics are not to be compared to the state or regional average but instead to their own expected rate. Comparisons between clinics are achieved with a calculation of actual result/expected result and significance testing is performed by using a chi square test.

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

NA

2b3.3a. Describe the conceptual/clinical <u>and</u> 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?

[2014] Normally, during our measure development process, the expert panel discusses potential variables for risk adjustment that are important to consider for the measured population. Variables are included in public comment and collected during pilot testing to assess feasibility. For this measure, which has been in place since ~ 2004, MNCM was not the developer but has taken on the stewardship. Variables for risk adjustment have been selected and tested over the years. In addition to our standard demographic variables (gender, age, zip, race/ethnicity, country of origin, primary language and insurance product), we have also assessed clinical variables including type of diabetes, major depression, and ischemic vascular disease. The potential risk adjustment variables are then evaluated for appropriate inclusion in the model based on a t value outside the range of -2.0 and +2.0.

The variables that have been selected are insurance product (commercial, Medicare, MN government programs, self-pay/uninsured), age bands (18-25, 26-50, 51-65 and 65 to 75) and diabetes type (1 or 2).

2018 Update

Risk model and variables have evolved since our last endorsement in 2014.

As of 2016, MNCM developed a framework for selection of risk variables. This framework is based both on NQF recommendations and consensuses from MNCM Risk Adjustment and Segmentation Committee.

Guiding Principles adopted by the MNCM Risk Adjustment & Segmentation Committee (June 2017)

Risk Adjustment: Framework and Guiding Principles for Selecting Risk Factors of Clinical Quality Measures

Risk adjustment refers to statistical methods to control or account for patient-related factors when computing performance measure scores. Risk adjusting outcome performance measures seeks to account for differences in patient health status, clinical factors and, when appropriate, socio-demographic status (SDS).

The selection of risk factors for adjustment is directed by a framework of criteria that must be considered for each factor. **The following criteria are to be applied during:**

1.) **Measure development when recommending variables for data collection and testing for potential risk adjustment, and**
2.) **Selection of tested variables for the application of risk adjustment.**
3.) **Reevaluation of currently applied risk adjustment factors.**

CRITERIA	RATIONALE
Clinical/conceptual relationship with the outcome of interest	A logical theory must explain the association between the factor and the outcome. Begin with conceptual model informed by research and experience; does not require a direct causal relationship
Empirical association with the outcome of interest	A statistical association to confirm the conceptual relationship
Variation in prevalence of the factor across the measured entities	If there is no variation in prevalence across providers being measured, it will not bias performance results
Not confounded with quality of care – risk factors should: • be present at the start of care and • not represent the quality of care provided (e.g., treatments, interventions, expertise of staff)	Trying to isolate effects of the provider – quality of care Ensures not a result of care provided Although these could explain variation in outcome, trying to isolate differences in performance due to differences in the care provided
Resistant to manipulation or gaming – generally, a diagnosis or assessment data (e.g., functional status score) is considered less susceptible to manipulation than a clinical procedure or treatment (e.g., physical therapy)	Ensures validity of performance score as representing quality of care (vs. for example, up coding)
Accurate data that can be reliably and feasibly captured at a reasonable cost	Data and resource limitations often represent a practical constraint to what factors are included in risk models.
Contribution of unique variation in the outcome (i.e., not redundant or highly correlated with another risk factor)	Prevent over-fitting and unstable estimates, or coefficients that appear to be in the wrong direction, reduce data collection burden

Risk Adjustment: Framework and Guiding Principles for Selecting Risk Factors of Clinical Quality Measures

CRITERIA	RATIONALE
Potentially, improvement of the risk model (e.g., risk model metrics of discrimination – i.e., sensitivity/specificity, calibration) and sustained with cross validation	Change in R-squared or C-statistic may not be significant, but calibration at different deciles of risk might improve May not appear to be a big change but could represent meaningful differences in terms of the outcome (e.g., lives, dollars) Order of entry into a model may influence this result
Potentially, face validity and acceptability	Some factors may not be indicated empirically, but could improve acceptability – need to weigh against negative impact on model, feasibility and burden of data collection

Segmentation

Segmentation is the process of dividing a population into meaningful categories and reporting them separately. For example, the reporting of performance measures by payer type (commercial, Medicare, Medicaid) is appropriate when the differences between the populations are the main objective of the report (to highlight disparities) or when the different groups are independently measured and evaluated.

The criteria for the selection of risk adjustment factors and meaningful categories for segmentation are not necessarily identical, as the application of these types of comparative reporting serve different purposes.

Application of Risk Adjustment and Segmentation

When appropriate, risk adjustment and segmentation will be applied to measures as was approved by MARC on March 13, 2013, and described below.

Patient age continues as a statistically significant variable for risk adjustment; however, we are now applying age as a continuous variable and not restricting to age band categories. Additionally, insurance product type and diabetes type are also still included in the model and continue to be statistically significant.

Insurance product affects diabetes outcomes as shown in ODC analyses stratified by product (MCHP is Medicaid and Other Purchasers include all other insurance types):

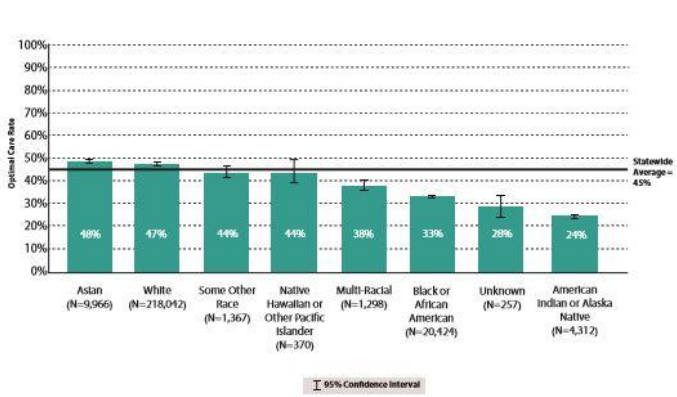
Year	MHCP Rate	MHCP CI (U/L)	MHCP Denominator (Patients Sampled)	Other Purchaser Rate	Difference (Other-MHCP)
2017	32.5%	32.1% - 33.0%	45,023	47.6%	15.1%**
2016	33.6%	33.2% - 34.1%	36,757	48.9%	15.3%**

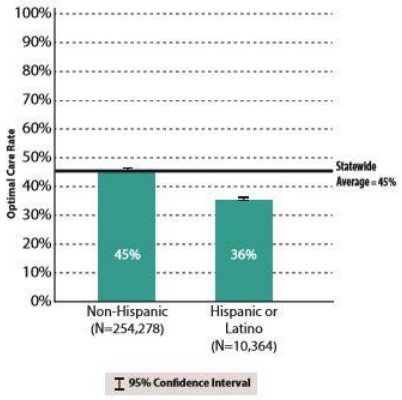
* A confidence interval gives an estimated range of values calculated from a given sample of data. A 95 percent confidence interval implies a 95 percent level of confidence that the interval includes the true mean or parameter.
** Denotes a statistically significant difference.

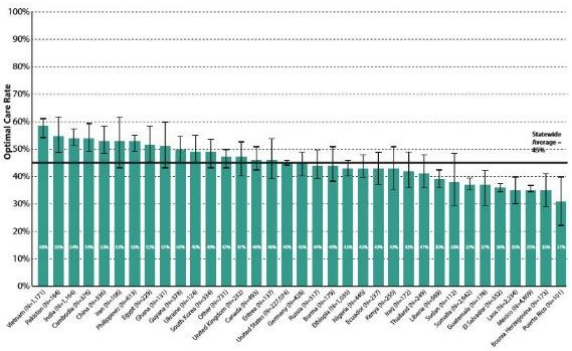
* Annual MNCM Health Care Disparities Report

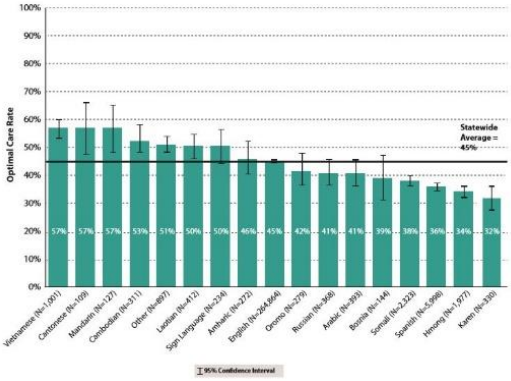
mncm.org/wp-content/uploads/2018/03/2017-Disparities-Report-FINAL-3.26.2018.pdf

Empirical analysis on RELO showed that there were differences in diabetes outcomes based on these social risk factors:









MNCM 2017 Health Equity Report

mncm.org/wp-content/uploads/2018/01/2017-Health-Equity-of-Care-Report_unencrypted-1.pdf

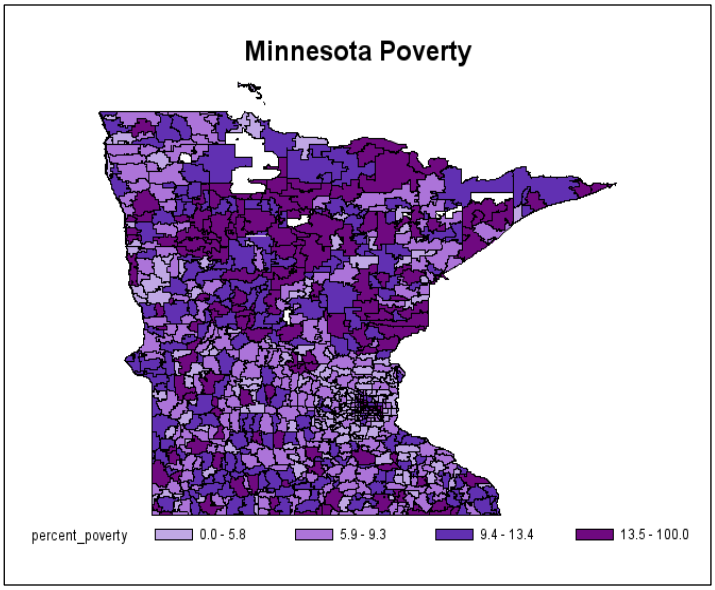
As mentioned in section 1.8; RELO variables were analyzed and MNCM decided not to use in a risk adjustment capacity. (Excerpt from 1.8)

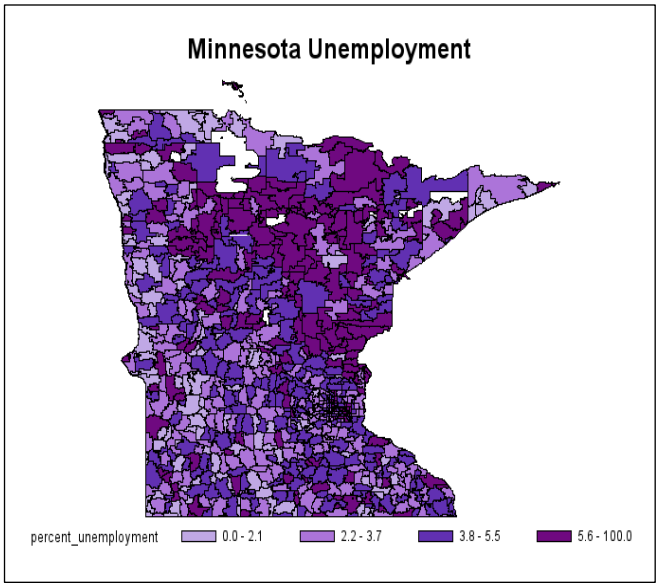
MNCM considered the inclusion of race, Hispanic ethnicity, preferred language and country of origin (RELO) as potential risk adjustment variables. While there are significant differences in outcome when segmenting the data by RELO variables, these variables are not used in risk adjustment because it is impossible to separate the

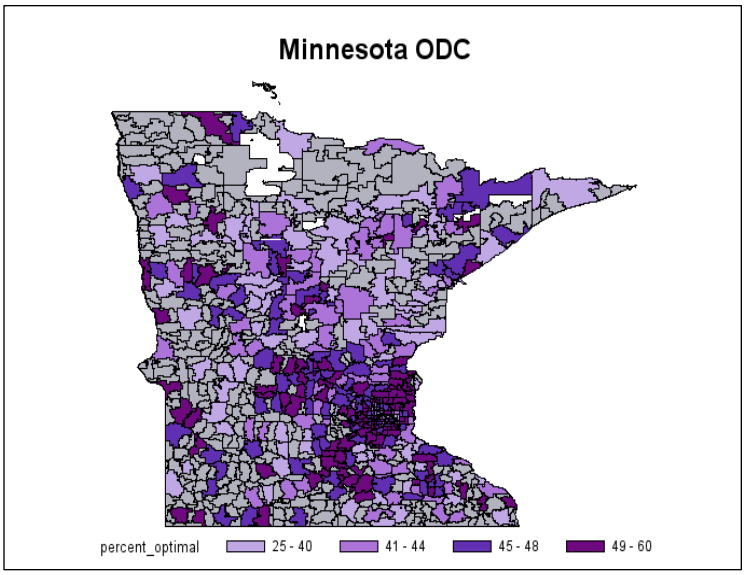
disparity in outcome between the patient's environment and the clinic's contribution to the disparity specific biases from healthcare providers that influence their interactions with patients.

- MNCM convened a panel of social science researchers from the University of Minnesota in August 2016 to understand if RELO variables were confounded by the clinic's contribution. For every measure and for every factor, as a group, the researchers were hard pressed to find a conceptual relationship that was not confounded by the clinic's contribution. This is confirmed in the social science literature available on the topic.
- The panel recommended use of geography instead, specifically neighborhood characteristics which lead to the development of the deprivation index mentioned above.

MNCM investigated optimal diabetes care and several social risk factors by patient zip code and observed that there is significant variation by location:







These empirical results led us to the development of the deprivation index which is described in detail in section 1.8 and in section 2b3.4b. The deprivation index calculated from U.S. Census information on socioeconomic factors based on the patient's 5-digit zip code.

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?

[2014] The effect of risk adjustment on clinic ranking was examined. First, the clinic's unadjusted and adjusted quality measures were compared using correlation analysis. Pearson's correlation examined the correlation when the measures are treated as continuous measures. A high correlation (close to 1) means that the two measures strongly co-vary, when one is high the other is high.
The second comparison ranked the clinics into performance rank deciles based on the unadjusted and adjusted scores and then examines how decile rankings based on unadjusted measures compared to decile rankings based on adjusted measures. The third comparison ranks clinics into Below Average, Average, and Above Average categories using statistical methods that consider the quality measure's confidence interval which is calculated based on the number of patients each clinic report. These two methods are compared directly in our accompanying report on the quality deviations ranking approach.

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.

[2014] Because ODC is a binary variable (0 or 1), the risk adjustment model was estimated using a logistic model implemented in the SAS Logistic Procedure that accounts for its non-continuous nature. The risk adjusters and an indicator for each clinic were included in the model. The estimated coefficient for the clinic indicator measures the clinic's ODC adjusting for the patient risk adjusters that were included in the model. The clinic level indicator was used to construct a risk adjusted ODC score at the clinic level that ranged from 0 to 1 (0% to 100%). The effect of risk adjustment on clinic rankings was calculated by comparing the risk adjusted ODC to the unadjusted ODC measure, the average ODC for all patients reported by the clinic.

The social risk factors included in our model are patient's insurance type and a deprivation index which have been described above. Because deprivation index is a new risk factor since our last submission, a description of the process used to determine inclusion of the deprivation index in our risk adjustment model is included.

Summary of MNCM Study: Impact of Adjusting Measures for Patients' Neighborhood Socio-Economic Status

Reminder of Goal and Methodology:

- Goal: to isolate the clinic/medical group's true impact on patients' health and allow them to be compared more easily.
- MNCM utilizes an indirect standardization methodology for risk adjustment.
- Each clinic/medical group's rate is compared to a unique benchmark rate for that clinic/medical group that is based on the mix of patient risk seen at that clinic/medical group

Study question:

How to measure the impact of where patients live on MNCM quality measures?

1. Literature Review

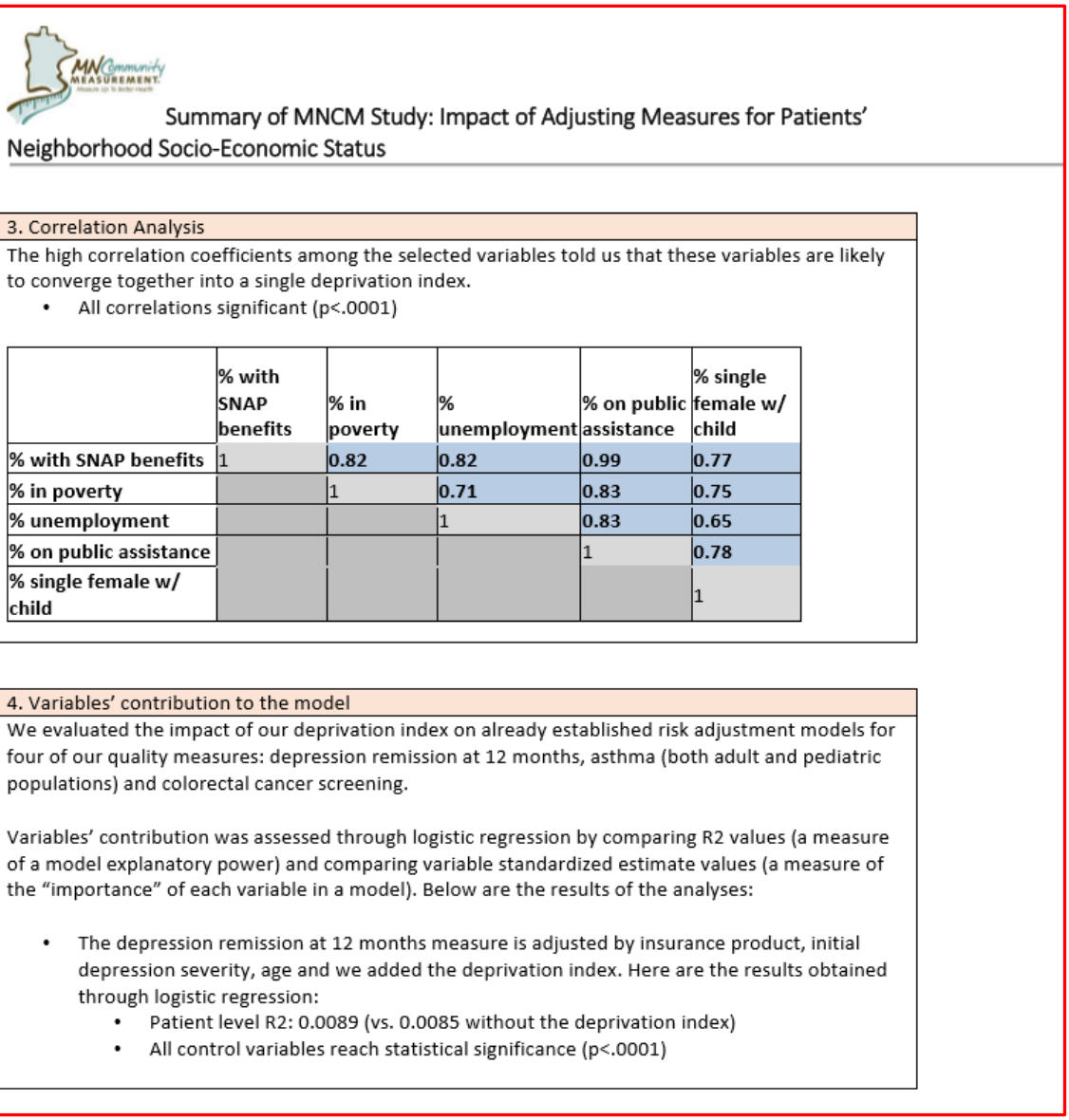
- We examined published, peer-reviewed articles on the creation of a measure for areas socio-economic status (SES).
- Key findings:
  - Census data at the ZIP code level is typically used
  - There was not consistent evidence that a more granular geographical unit (Census Track, Census Block) always produced more significant results. It was very measure and situation dependent.
  - Principal Components Analysis is used
  - A Deprivation Index is generated.

2. Variable Selection

In line with published literature, we chose the following variables (from the 2015 census data) to evaluate for our deprivation index:

- % with SNAP benefits
- % in poverty
- % unemployment
- % on public assistance
- % single female w/ child

Staff note: Median Income was tested and not retained as a component of the index because this variable behaves differently from the other variables listed above.



Summary of MNCM Study: Impact of Adjusting Measures for Patients' Neighborhood Socio-Economic Status

3. Correlation Analysis

The high correlation coefficients among the selected variables told us that these variables are likely to converge together into a single deprivation index.

- All correlations significant (p<.0001)

	% with SNAP benefits	% in poverty	% unemployment	% on public assistance	% single female w/ child
% with SNAP benefits	1	0.82	0.82	0.99	0.77
% in poverty		1	0.71	0.83	0.75
% unemployment			1	0.83	0.65
% on public assistance				1	0.78
% single female w/ child					1

4. Variables' contribution to the model

We evaluated the impact of our deprivation index on already established risk adjustment models for four of our quality measures: depression remission at 12 months, asthma (both adult and pediatric populations) and colorectal cancer screening.

Variables' contribution was assessed through logistic regression by comparing R2 values (a measure of a model explanatory power) and comparing variable standardized estimate values (a measure of the "importance" of each variable in a model). Below are the results of the analyses:

- The depression remission at 12 months measure is adjusted by insurance product, initial depression severity, age and we added the deprivation index. Here are the results obtained through logistic regression:
  - Patient level R2: 0.0089 (vs. 0.0085 without the deprivation index)
  - All control variables reach statistical significance (p<.0001)

* 5-digit zip code was utilized for the analysis of the deprivation index

For the ODC patient population, the Deprivation Index ranges from -6.84 to 1.42. The index is centered at 0 and a higher number indicates a higher socioeconomic level.

The table below is a snapshot of a larger table that shows the impact of the deprivation index. The larger table included clinics with 750-1250 patients and ranked the clinics by impact of adding the zip code deprivation index to the indirect standardization risk adjustment model. The smaller snapshot table (next page) shows the 10 clinics where the expected value (indirect standardization) is increased the most by the addition of the deprivation index and the 10 clinics where the expected value is decreased the most with the addition of the deprivation index. It is important to note that the 10 clinics with raised expected values are all in wealthier suburbs and the 10 with lower expected values are either rural or inner city. This is exactly the result that was anticipated and shows that the deprivation index is working as expected.

Sample Impact of Adding Zip Code Level Economic Averages to 2017 Diabetes Measure
Clinics with 750-1250 patients

Clinic	Patients	Actual Rate	Expected Rate without SES	Expected rate with SES	Rate Difference	Impact of SES Variable	Change in Significance Test
1	970	48.1%	43.8%	45.6%	1.8%	raised expectation	Above to Average
2	834	45.2%	44.7%	46.4%	1.6%	raised expectation	
3	873	46.2%	44.5%	45.9%	1.5%	raised expectation	
4	1001	54.8%	46.7%	48.2%	1.4%	raised expectation	
5	1092	47.0%	45.9%	47.3%	1.4%	raised expectation	
6	1079	51.8%	45.6%	46.9%	1.3%	raised expectation	
7	930	51.8%	47.1%	48.5%	1.3%	raised expectation	
8	871	35.6%	48.6%	49.9%	1.3%	raised expectation	
9	827	51.8%	46.1%	47.3%	1.3%	raised expectation	
10	1034	51.1%	45.1%	46.3%	1.3%	raised expectation	
…	…	…	…	…	…	…	
78	829	41.0%	43.8%	42.2%	-1.6%	lowered expectation	
79	711	45.9%	43.5%	41.7%	-1.8%	lowered expectation	Average to Above
80	906	30.4%	45.9%	44.1%	-1.8%	lowered expectation	
81	1097	44.3%	44.4%	42.5%	-1.9%	lowered expectation	
82	1030	42.9%	45.8%	43.9%	-2.0%	lowered expectation	Below to Average
83	1093	46.2%	44.3%	42.1%	-2.2%	lowered expectation	Average to Above
84	1107	35.8%	41.4%	38.8%	-2.6%	lowered expectation	
85	914	44.5%	47.2%	44.6%	-2.7%	lowered expectation	
86	709	40.3%	40.7%	37.6%	-3.1%	lowered expectation	
87	1057	18.5%	36.2%	33.0%	-3.2%	lowered expectation	
88	703	30.7%	40.6%	37.2%	-3.4%	lowered expectation	
89	833	37.3%	40.9%	37.5%	-3.4%	lowered expectation	Below to Average

* 5-digit zip code was utilized for the analysis of the deprivation index

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*)

[2014] Because ODC is a binary variable (0 or 1), the risk adjustment model was estimated using a logistic model implemented in the SAS Procedure Logistic that accounts for its non-continuous nature. The risk adjusters and an indicator for each clinic were included in the model. The estimated coefficient for the clinic indicator measures the clinic's ODC adjusting for the patient risk adjusters that were included in the model. The clinic level indicator was used to construct a risk adjusted ODC score at the clinic level that ranged from 0 to 1 (0% to 100%). The effect of risk adjustment on clinic rankings was calculated by comparing the risk adjusted ODC to the unadjusted ODC measure, the average ODC for all patients reported by the clinic.

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*)**:**

[2014] At the clinic level, the average ODC measure was 35.1% (standard deviation = 12%). The average number of patients reported by a clinic was 348 (standard deviation = 405). At the patient level, the average ODC was 39.7%. The average age in the examined population was 58, 46% were female, 7.3% had Type I diabetes, 19.6% were depressed, 48.8% had commercial insurance, 33.1% had Medicare coverage, and 7.1% had Medicaid coverage.

Risk adjustment is necessary only when there is heterogeneity across clinics. There was significant heterogeneity across clinics in insurance product mix ($\chi 2 = 65617$, $p < .001$), patient age ($\chi 2 = 12522$, $p < .001$), gender ($\chi 2 = 5256$, $p < .001$), depression ($\chi 2 = 4290$, $p < .001$), Type 1 Diabetes ($\chi 2 = 67297$, $p < .001$), and distance to the clinic ($\chi 2 = 63638$, $p < .001$).

Update 2018

At the clinic level*, the average ODC measure was 43.2% (standard deviation = 10.2%). The average number of patients reported by a clinic was 495 (standard deviation = 519). At the patient level**, the average ODC was 44.9%. The average age in the examined population was 58, 54% were male, 8.0% had Type I diabetes, 40.5% had commercial insurance, 36.4% had Medicare coverage, 15.0% had Medicaid coverage and 2.7% had no insurance.

* When evaluating rates and comparison among clinics, a clinic is only included in the analysis if they have $\geq$ 30 eligible patients in the measurement period (calendar year).

** When evaluating the entire population (statewide), all eligible patients submitted for rate calculation are included even if their clinic's number of eligible patients was < 30.

Pearson Correlation Coefficient

Correction between specific risk variable and overall result

Value between -1 and 1 where 1 is a total positive linear correlation and -1 is a total negative correlation

Variable		Pearson
Gender		0.081
Patient Age		0.233
Diabetes Type	Type One	-0.058
Insurance Product		
	Commercial	0.424
	Medicare	0.151
	Medicaid	0.151
	Uninsured	-0.293
Deprivation Index (Socioeconomic ratio for patient zip code)		0.398

Independent groups t-tests comparing insurance product type groups with Commercial*

	N	Optimal Care Rate	T value	p-value
Commercial	127,849	44.95%	-	-
Medicare	111,797	51.37%	-31.44	<0.001
Medicaid	45,743	32.65%	70.54	<0.001
Uninsured	8,203	26.69%	48.33	<0.001
Unknown	13,566	43.91%	-16.43	<0.001

Independent groups t-tests comparing diabetes type groups with type 2*

	N	Optimal Care Rate	T value	p-value
Type 1	23,157	34.71%	-32.72	<0.001
Type 2	282,681	45.82%	-	-
Unknown	1,320	32.30%	10.56	<0.001

*Pooled t-tests were calculated when variances were equal and Satterthwaite were calculated when variances were unequal.

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

Logistic Regression Output results

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-1.5500	0.0232	4479.8773	<.0001
Patient_Age	1	0.0252	0.000405	3857.8729	<.0001
Medicare	1	-0.0521	0.00961	29.3463	<.0001
Medicaid	1	-0.4350	0.0118	1354.3824	<.0001
Uninsured	1	-0.7222	0.0259	776.4681	<.0001
Unknown	1	-0.1363	0.0185	54.3630	<.0001
Type1	1	-0.1673	0.0153	119.4319	<.0001
Unk_type	1	-0.4271	0.0603	50.2018	<.0001
Dep_index	1	0.1346	0.00459	858.3075	<.0001

Test of correlation between variables

Pearson Correlation Coefficients, N = 307158 Prob > |r| under H0: Rho=0										
	medicare	medicaid	commercial	uninsured	unknown	patient_age	type1	type2	unk_type	dep_index
medicare	1	-0.31644	-0.63877	-0.12531	-0.16261	0.52416	-0.11265	0.11081	-0.00408	0.01944
		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0237	<.0001
medicaid	-0.31644	1	-0.35322	-0.06929	-0.08992	-0.20847	0.00133	-0.00469	0.01404	-0.20136
	<.0001		<.0001	<.0001	<.0001	<.0001	0.4613	0.0094	<.0001	<.0001
commercial	-0.63877	-0.35322	1	-0.13987	-0.18151	-0.324	0.10372	-0.09962	-0.0062	0.13291
	<.0001	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	0.0006	<.0001
uninsured	-0.12531	-0.06929	-0.13987	1	-0.03561	-0.09471	-0.01518	0.01549	-0.00286	-0.0588
	<.0001	<.0001	<.0001		<.0001	<.0001	<.0001	<.0001	0.1134	<.0001
unknown	-0.16261	-0.08992	-0.18151	-0.03561	1	-0.0146	0.02456	-0.02451	0.00235	0.03067
	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001	<.0001	0.1928	<.0001
patient_age	0.52416	-0.20847	-0.324	-0.09471	-0.0146	1	-0.30913	0.30708	-0.0236	0.06906
	<.0001	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001	<.0001	<.0001
type1	-0.11265	0.00133	0.10372	-0.01518	0.02456	-0.30913	1	-0.9704	-0.01876	0.03128
	<.0001	0.4613	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001	<.0001
type2	0.11081	-0.00469	-0.09962	0.01549	-0.02451	0.30708	-0.9704	1	-0.22326	-0.02747
	<.0001	0.0094	<.0001	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001
unk_type	-0.00408	0.01404	-0.0062	-0.00286	0.00235	-0.0236	-0.01876	-0.22326	1	-0.01252
	0.0237	<.0001	0.0006	0.1134	0.1928	<.0001	<.0001	<.0001		<.0001
dep_index	0.01944	-0.20136	0.13291	-0.0588	0.03067	0.06906	0.03128	-0.02747	-0.01252	1
	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	

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

Comparison of Unadjusted and Adjusted Decile Ranks (N/Percent of Row)*											
	Risk Adjusted Decile Rank										
Unadjusted Decile Rank	0% to 10%	10% to 20%	20% to 30%	30% to 40%	40% to 50%	50% to 60%	60% to 70%	70% to 80%	80% to 90%	90% to 100%	Total
0% to 10%	0/0.00	0/0.00	0/0.00	1/100.00	0/0.00	0/0.00	0/0.00	0/0.00	0/0.00	0/0.00	1
10% to 20%	0/0.00	0/0.00	0/0.00	8/44.44	10/55.56	0/0.00	0/0.00	0/0.00	0/0.00	0/0.00	18
20% to 30%	0/0.00	0/0.00	0/0.00	21/41.18	30/58.82	0/0.00	0/0.00	0/0.00	0/0.00	0/0.00	51
30% to 40%	0/0.00	0/0.00	0/0.00	18/15.00	102/85.00	0/0.00	0/0.00	0/0.00	0/0.00	0/0.00	120
40% to 50%	0/0.00	0/0.00	0/0.00	9/3.42	248/94.30	6/2.29	0/0.00	0/0.00	0/0.00	0/0.00	263
50% to 60%	0/0.00	0/0.00	0/0.00	0/0.00	151/98.69	2/1.31	0/0.00	0/0.00	0/0.00	0/0.00	153
60% to 70%	0/0.00	0/0.00	0/0.00	0/0.00	11/91.67	1/8.33	0/0.00	0/0.00	0/0.00	0/0.00	12
70% to 80%	0/0.00	0/0.00	0/0.00	0/0.00	0/0.00	0/0.00	0/0.00	0/0.00	0/0.00	0/0.00	0
80% to 90%	0/0.00	0/0.00	0/0.00	0/0.00	0/0.00	0/0.00	0/0.00	0/0.00	0/0.00	0/0.00	0
90% to 100%	0/0.00	0/0.00	0/0.00	0/0.00	0/0.00	0/0.00	0/0.00	0/0.00	0/0.00	0/0.00	0
Total	0	0	0	57	552	9	0	0	0	0	618
*Grey cells indicate no change in rank. Blue cells indicate an increase in rank after risk adjustment and green cells indicate a decrease in rank after risk adjustment. N is the number of clinics in each cell and the percent of row is the percent of the total unadjusted decile ranked clinics in each cell.											

	Impact of risk adjustment on clinic level measurement (N/Overall percent)*			
	With Risk Adjustment			
Without Risk Adjustment	Below Expected	As Expected	Above Expected	Total Counts
Below	125/20.2%	38/6.2%	0/0.0%	163
Average	8/1.3%	282/45.6%	7/1.1%	297
Above	0/0.0%	63/10.2%	95/15.3%	158
Total Counts	133	383	102	618
*Grey cells indicate no change in rank. Blue cells indicate an increase in rank after risk adjustment and green cells indicate a decrease in rank after risk adjustment.				

Direction of Impact	N	Percent
Moved toward As Expected	101	16.3%
Moved away from Average	15	2.4%
Improved	45	7.3%
Worse	71	11.5%
Impacted	116	18.8%

2b3.9. Results of Risk Stratification Analysis:

[2014] NA, measure is not stratified.

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*)

Our analysis of risk adjustment factors for the Optimal Diabetes Care measure indicates that age, diabetes type, deprivation index and insurance product variables are related to ODC and warrant attention for risk adjustment.

For most clinics there is no change in clinic ranking due to risk adjustment, although some increase in ranking and others decrease in ranking. For clinics whose rankings are impacted by the risk adjustment, it is valid and based on disparate differences among these clinics.

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*)

NA

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

Note: *Applies to the composite performance measure.*

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)*

[2014] Annual Health Care Quality Report and Health Care Disparities reports available at http://mncm.org/reports-and-websites/reports-and-data/

Methodology:

Identifying High Performing Medical Groups/Clinics

For each measure, both individual medical group rates and a medical group average rate were calculated. Medical groups that achieved high performance were identified by comparing the individual medical group/clinic rate with the medical group average. Medical groups that had rates that were fully above the medical group average and 95 percent confidence intervals were noted as high performers.

Additionally, the Top 15 performers are identified.

Identifying Medical Groups and Clinics with Biggest Improvements

For each measure, individual medical group and clinic rates during report year 2017 were compared with their rates during report year 2016, calculating an absolute percentage point difference. Medical groups and clinics with the largest percentage point increases were identified.

Medical Group and Clinic Performance Over Time (Three Years)

This analysis was done to determine patterns of medical group and clinic performance over time per measure. Patterns were reviewed for the three reporting years (2015, 2016 and 2017).

The percent and number of medical groups were reported for each of the following patterns of rate changes over the past three years for each measure:

- Consistently improved: Medical groups with more than a two percentage point increase between each consecutive year.
- Relatively stable: Medical groups that had no more than a two percentage point increase or decrease between each consecutive year (-2 percent – +2 percent).
- Consistently decreased: Medical groups with more than a two percentage point decrease between each consecutive year.
- Variable performance (with an improvement or with a decline): Medical groups with an up/down pattern that was not consistent and did not fall into one of the other categories.

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*)

[2014]

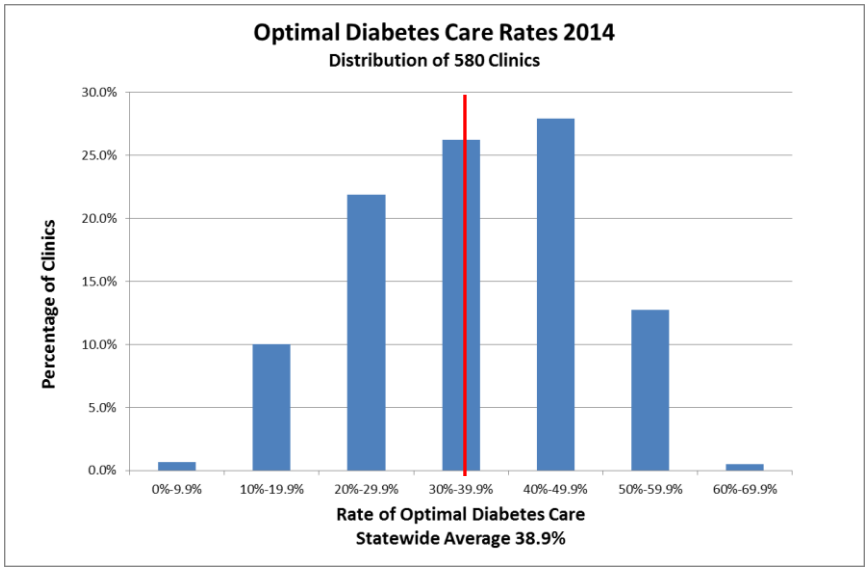
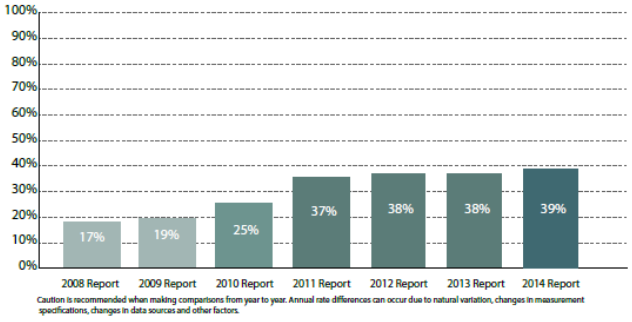


TABLE 10: STATEWIDE RATE FOR OPTIMAL DIABETES CARE

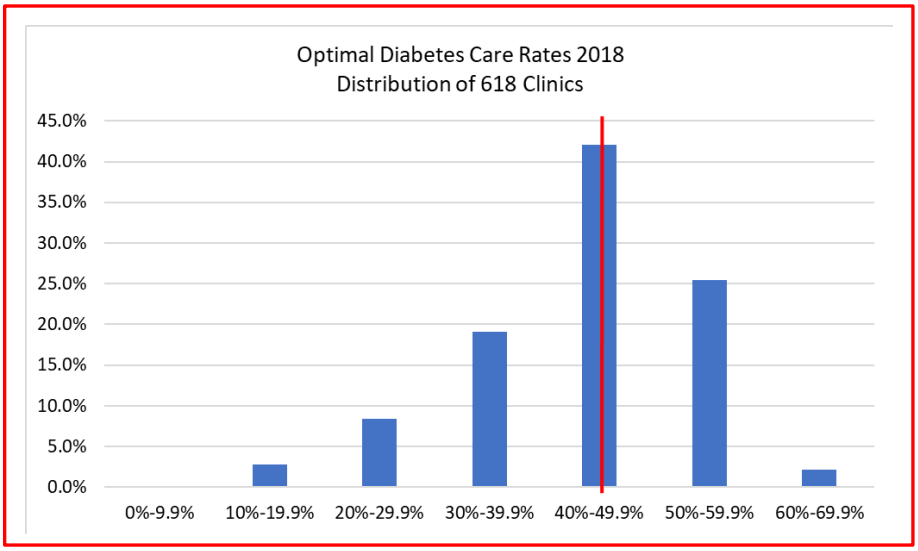
	Statewide Average (Weighted)	95% CI	Numerator (Patients who met treatment goals)	Denominator (Patients sampled)	Total Eligible
Optimal Diabetes Care	38.9%	38.7%-39.1%	90,499	230,818	237,354

FIGURE 7: STATEWIDE RATES FOR OPTIMAL DIABETES CARE OVER TIME



Caution is recommended when making comparisons from year to year. Annual rate differences can occur due to natural variation, changes in measurement specifications, changes in data sources and other factors.

Update for 2018



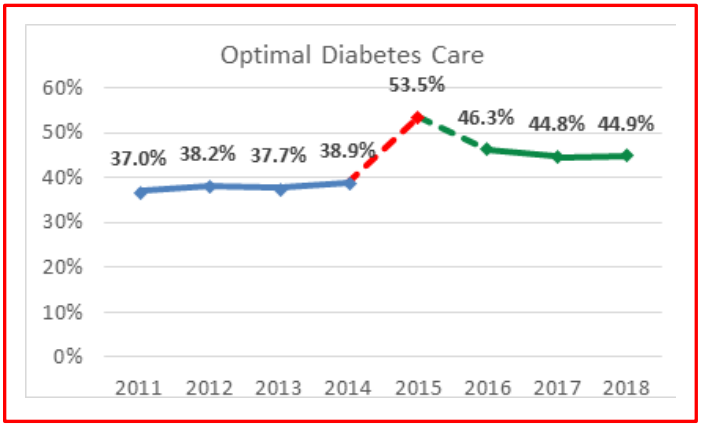
Performance Rates Over Time

Report Year	Statewide Average	Numerator	Denominator
2018	44.9%	137,985	307,158
2017^	44.8%	132,106	295,049
2016+	46.3%	119,194	257,078
2015*	53.5%	131,847	249,878
2014	38.9%	90,499	230,818
2013	37.7%	80,190	208,809
2012	38.2%	73,037	184,234
2011	37.0%	61,930	158,770

* No cholesterol component during redesign

+ New cholesterol component for statin use in place

^ Established patient criteria replaces visit counting



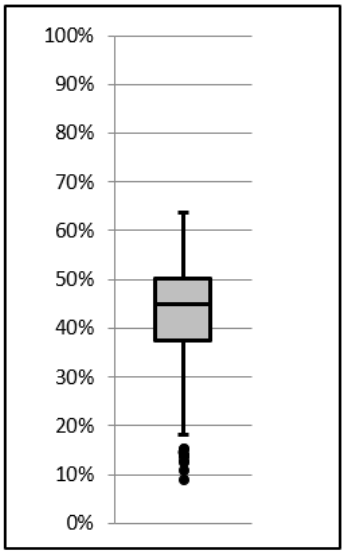
Performance Rates Most Recent Reporting Period

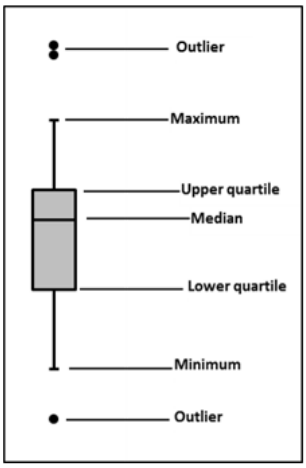
Performance variation 2018 (2017 DOS): (618 reportable clinics)

Percentiles:
10th: 28.1%
25th: 37.4%
50th: 45.0%
75th: 50.3%
90th: 55.2%

Range: 9.0% - 63.8%

Mean: 43.2%





* quartile box plot analysis comparing clinic level results includes only clinics with ≥ 30 patients; n = 618 clinics and 306,509 patients

Highest Performers in 2017 by Medical Group Type for Clinical Measures

The following **tables 44-47** list the high-performing medical groups by specialty type based on the applicable clinical measures on which they were reported. These groups had above-average rates on most clinical measures.

Eight primary care medical groups achieved rates that were above average more frequently than others. Each of these medical groups achieved above-average rates on more than half of the clinical measures for which they were reportable.

TABLE 44: HIGH PERFORMING MEDICAL GROUPS IN 2017 - **PRIMARY CARE**

Measure	Park Nicollet Health Services 22 out of 25	Allina Health 21 out of 25	Health Partners Clinics 21 out of 25	Entira Family Clinics 15 out of 25	CentraCare Health 14 out of 25	Mankato Clinic 14 out of 25	Fairview Health Services 14 out of 25	Stillwater Medical Group 9 out of 15
ADHD							●	NA
Adolescent Immunizations		●	●					NA
Breast Cancer Screening	●	●	●		●	●	●	NA
Bronchitis	●	●	●					NA
Cervical Cancer Screening	●		●				●	NA
Childhood Immunization Status (Combo 10)		●	●	●		●		NA
Chlamydia Screening	●	●	●	●	●		●	NA
Colorectal Cancer Screening^	●	●	●		●	●	●	
Controlling High Blood Pressure	●	●		●			●	NA
COPD	●		●	●		●	●	NA
Depression Remission at Six Months^	●	●	●	●	●			
Depression Remission at Twelve Months^	●	●	●	●	●	●		●
Depression response at six months^	●	●	●	●	●	●		
Depression response at twelve months^	●	●	●	●	●	●		●
*Maternity Care: Primary C-Section Rate^	●							
Optimal Asthma Control - Adults^	●	●	●	●		●	●	●
Optimal Asthma Control - Children^	●	●	●		●	●	●	●
Optimal Diabetes Care^	●	●	●	●	●		●	●
Optimal Vascular Care^	●	●	●	●			●	●

Excerpt 2017 Health Care Quality Report pg. 188

http://mncm.org/reports-and-websites/reports-and-data/health-care-quality-report/

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?*)

[2014] Measure continues to demonstrate opportunity for improvement as well as statistically significant and clinically meaningful differences between medical group practices and clinics.

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

Note: *Applies to all component measures, unless already endorsed or are being submitted for individual endorsement.*

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

NA

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

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

NA

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*)

NA

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?*)

NA

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

Note: *Applies to the overall composite measure.*

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*)

[2014] For this patient level all-or-none composite measure, elements missing from any component (e.g. visit but no blood pressure during the measurement year) are counted as a numerator component fail and therefore the patient would be accounted for and remain in the denominator.

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*)

[2014] The impact of missing data on measure calculations is minimal. For 2013 dates of service on over 230,800 diabetic patients submitted for rate calculation two variables were considered 1) with in the appropriate measurement timeframe and 2) valid values submitted:

Variable	Within measure period	Invalid values
A1c	96.8%	0.003%
Blood Pressure	99.8%	0.02%
LDL	89.8%	0.9%

Tobacco Status documented – 99.8%

Aspirin or anti-platelets if IVD- 97.2% had documented aspirin or anti-platelet in the measurement year or the date of a valid contraindication.

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*)

[2014] Patients with missing data are not excluded from the measure. Elements missing from any component are counted as a numerator component fail and remain in the denominator.

2c. EMPIRICAL ANALYSIS TO SUPPORT COMPOSITE CONSTRUCTION APPROACH

Note: *If empirical analyses do not provide adequate results—or are not conducted—justification must be provided and accepted in order to meet the must-pass criterion of Scientific Acceptability of Measure Properties. Each of the following questions has instructions if there is no empirical analysis.*

2d1. Empirical analysis demonstrating that the component measures fit the quality construct, add value to the overall composite, and achieve the object of parsimony to the extent possible.

2d1.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

[2014] This composite measure is a patient level all-or-none composite in which the desired goal is for the patient is to achieve multiple intermediate physiological clinical outcome and medication use targets to best reduce their overall risk of developing long term complications (acute MI, cardiovascular and peripheral vascular disease, kidney damage and failure, loss of vision, amputation, etc.) Reducing modifiable risks was the reason why this measure was developed. The components of this measure include blood sugar and blood pressure control, being tobacco-free, appropriate use of statins and daily aspirin or anti-platelet use if ischemic vascular disease.

Achieving the intermediate physiological outcome targets related to blood pressure and glycemic control in addition being tobacco free and use of daily aspirin and statins where appropriate are the diabetic patient's best mechanisms of avoiding or postponing long term complications associated with this chronic condition which affects millions of Americans. Measuring providers separately on individual targets is not as patient centric as a measure that seeks to reduce multiple risk factors for each patient. Diabetic patients are more likely to reduce their overall risk and maximize health outcomes by achieving several intermediate physiological targets.

A multifactorial approach to diabetes care that includes emphasis on blood pressure, lipids, glucose, aspirin use and non-use of tobacco will maximize health outcomes far more than a strategy that is limited to just one or two of these clinical domains (American Diabetes Association, 2014; Duckworth, 2009; Gaede, 2008; Holman, 2008a).

The components of this patient level all-or-none composite measure, though they can be analyzed as individual components especially for purposes of understanding opportunities within the composite measure, are treated as a whole. There is no weighting of the components; it is an all-or-none measure.

The methods used for analysis demonstrating soundness of the composite construct include distribution rates of performance for the individual components over time, rates of performance for all possible combinations of the components and Pearson product-moment correlation as a measure of the strength of linear regression of the relationships between the composite and the components.

The Pearson product-moment correlation coefficient (or Pearson correlation coefficient, for short) is a measure of the strength of a linear association between two variables and is denoted by *r*. Basically, a Pearson product-moment correlation attempts to draw a line of best fit through the data of two variables, and the Pearson correlation coefficient, *r*, indicates how far away all these data points are to this line of best fit (i.e., how well the data points fit this new model/line of best fit).

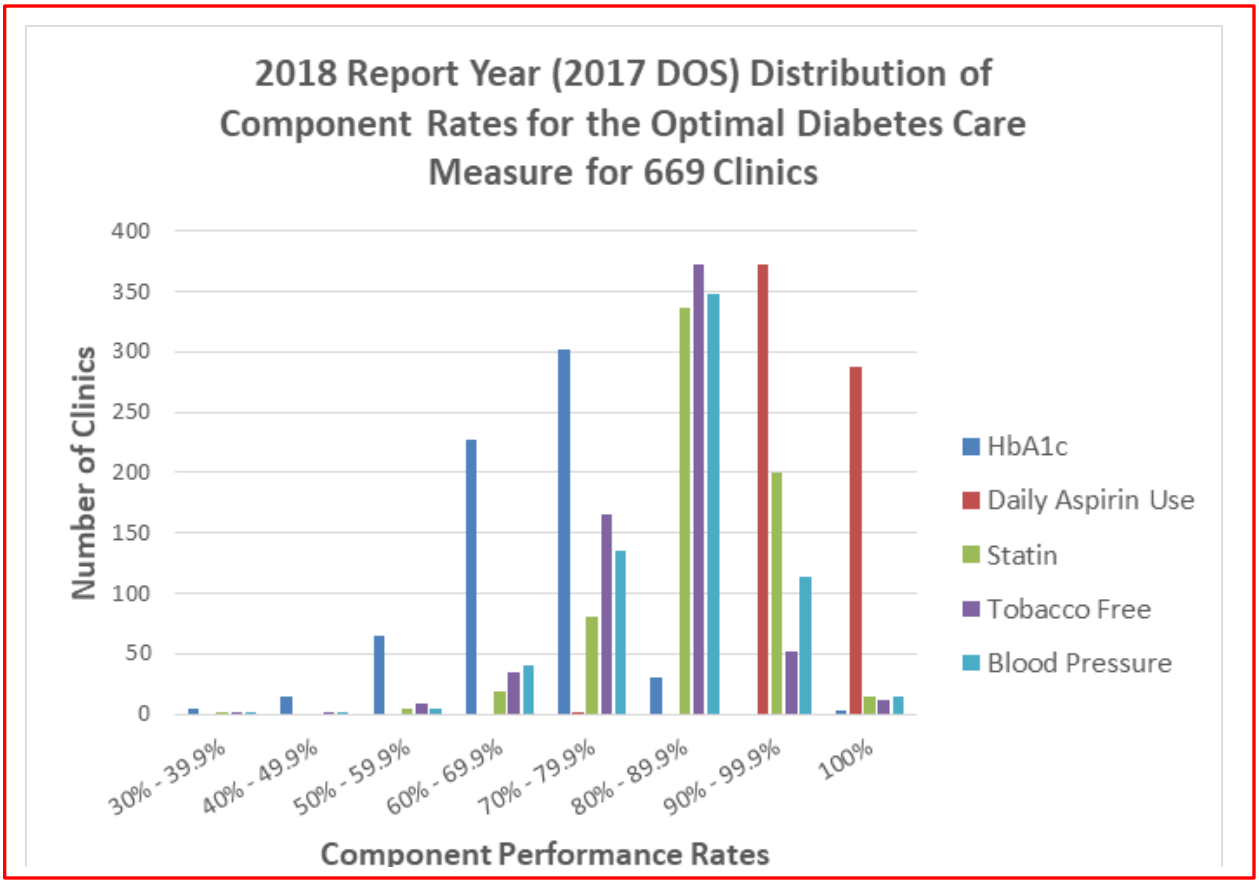
The Pearson correlation coefficient, r, can take a range of values from +1 to -1. A value of 0 indicates that there is no association between the two variables. A value greater than 0 indicates a positive association; that is, as the value of one variable increases, so does the value of the other variable. A value less than 0 indicates a negative association; that is, as the value of one variable increases, the value of the other variable decreases.

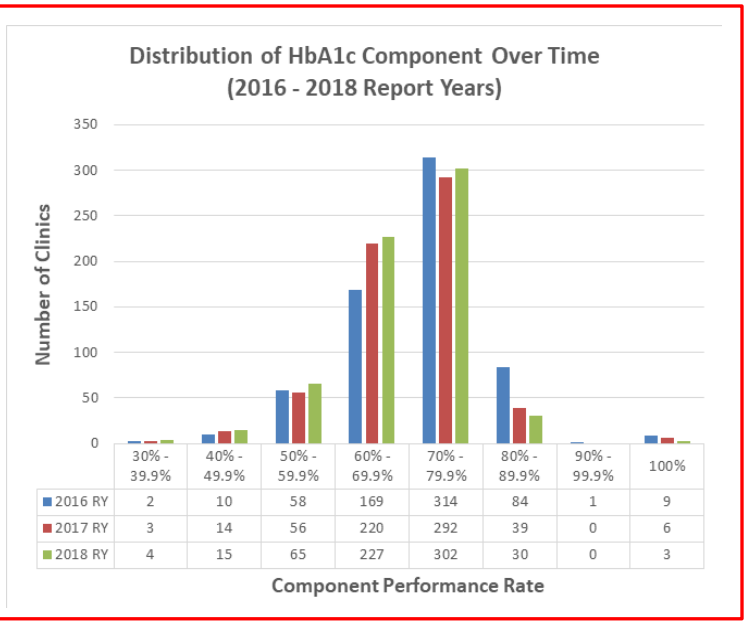
The stronger the association of the two variables, the closer the Pearson correlation coefficient, r, will be to either +1 or -1 depending on whether the relationship is positive or negative, respectively. Achieving a value of +1 or -1 means that all your data points are included on the line of best fit - there are no data points that show any variation away from this line. Values for r between +1 and -1 (for example, r = 0.8 or -0.4) indicate that there is variation around the line of best fit. The closer the value of r to 0 the greater the variation around the line of best fit.

2d1.2. What were the statistical results obtained from the analysis of the components? (e.g., *correlations, contribution of each component to the composite score, etc*.; *<u>if no empirical analysis</u>, identify the components that were considered and the pros and cons of each*)

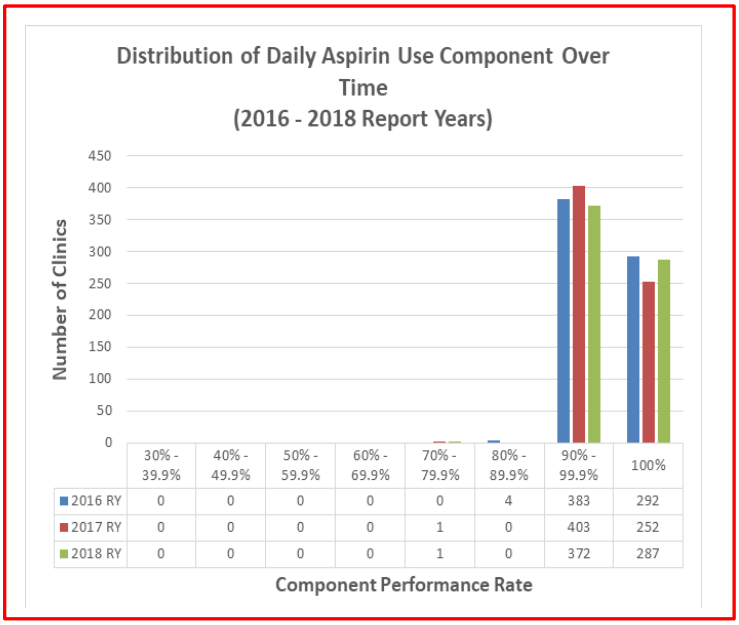
[2014] The components of this measure were selected as intermediate outcomes and medication use that can significantly reduce the diabetic patient's risk of developing or delaying microvascular and macrovascular or other long term complications associated with this chronic condition.

Updated Analysis for 2018

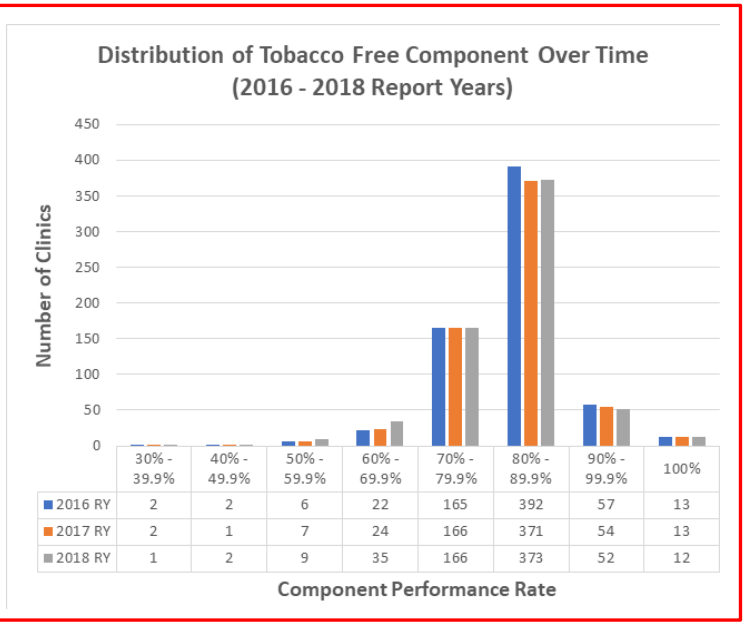




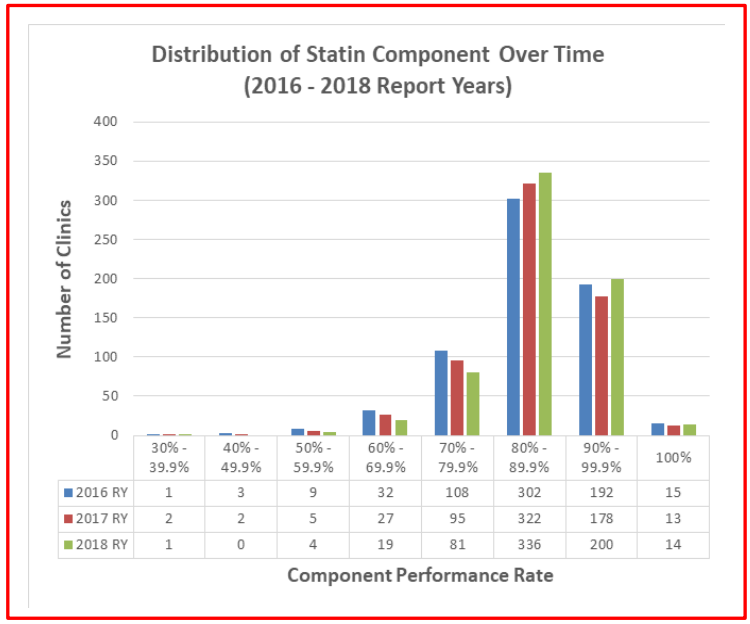
	30% - 39.9%	40% - 49.9%	50% - 59.9%	60% - 69.9%	70% - 79.9%	80% - 89.9%	90% - 99.9%	100%
2016 RY	2	10	58	169	314	84	1	9
2017 RY	3	14	56	220	292	39	0	6
2018 RY	4	15	65	227	302	30	0	3



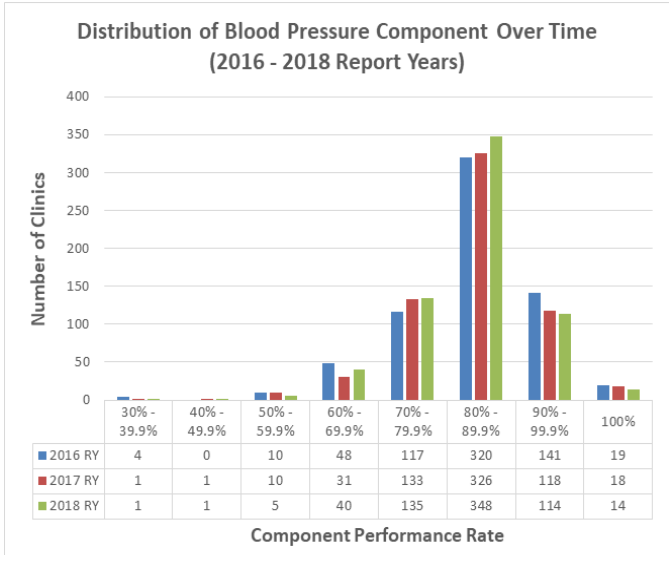
	30% - 39.9%	40% - 49.9%	50% - 59.9%	60% - 69.9%	70% - 79.9%	80% - 89.9%	90% - 99.9%	100%
2016 RY	0	0	0	0	0	4	383	292
2017 RY	0	0	0	0	1	0	403	252
2018 RY	0	0	0	0	1	0	372	287



	30% - 39.9%	40% - 49.9%	50% - 59.9%	60% - 69.9%	70% - 79.9%	80% - 89.9%	90% - 99.9%	100%
2016 RY	2	2	6	22	165	392	57	13
2017 RY	2	1	7	24	166	371	54	13
2018 RY	1	2	9	35	166	373	52	12



	30% - 39.9%	40% - 49.9%	50% - 59.9%	60% - 69.9%	70% - 79.9%	80% - 89.9%	90% - 99.9%	100%
2016 RY	1	3	9	32	108	302	192	15
2017 RY	2	2	5	27	95	322	178	13
2018 RY	1	0	4	19	81	336	200	14



	30% - 39.9%	40% - 49.9%	50% - 59.9%	60% - 69.9%	70% - 79.9%	80% - 89.9%	90% - 99.9%	100%
2016 RY	4	0	10	48	117	320	141	19
2017 RY	1	1	10	31	133	326	118	18
2018 RY	1	1	5	40	135	348	114	14

Proportion of how many patients are meeting component targets (2018)

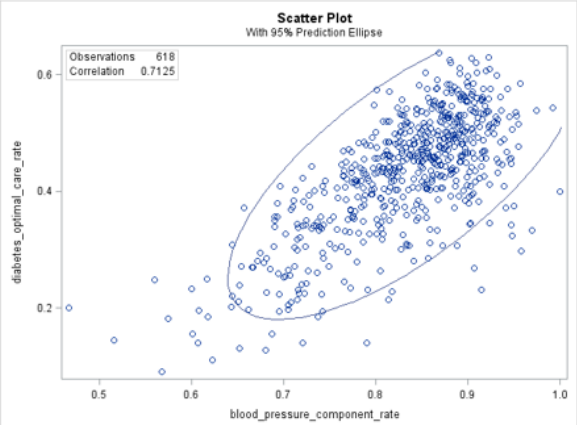
Category	# of Patients	Proportion
HbA1c alone	212,638	69%
Daily Aspirin Use alone	305,619	99%
Statin alone	269,592	88%
Tobacco Free alone	257,808	84%
Blood Pressure alone	256,323	83%
HbA1c + Daily Aspirin Use	211,658	69%
HbA1c + Statin	188,357	61%
HbA1c + Tobacco Free	181,729	59%
HbA1c + Blood Pressure	180,705	59%
Daily Aspirin Use + Statin	268,358	87%
Daily Aspirin Use + Tobacco Free	256,586	84%
Daily Aspirin Use + Blood Pressure	255,127	83%
Statin + Tobacco Free	226,784	74%
Statin + Blood Pressure	226,670	74%
Tobacco Free + Blood Pressure	216,209	70%
HbA1c + Daily Aspirin Use + Statin	187,546	61%
HbA1c + Daily Aspirin Use + Tobacco Free	180,928	59%
HbA1c + Daily Aspirin Use + Blood Pressure	179,933	59%
HbA1c + Statin + Tobacco Free	161,175	52%
HbA1c + Statin + Blood Pressure	161,088	52%
HbA1c + Tobacco Free + Blood Pressure	155,226	51%
Daily Aspirin Use + Statin + Tobacco Free	225,790	74%
Daily Aspirin Use + Statin + Blood Pressure	225,702	73%
Statin + Tobacco Free + Blood Pressure	191,582	62%
HbA1c + Daily Aspirin Use + Statin + Tobacco Free	160,506	52%
HbA1c + Statin + Tobacco Free + Blood Pressure	138,517	45%
HbA1c + Daily Aspirin Use + Statin + Blood Pressure	160,448	52%
Daily Aspirin Use + Statin + Tobacco Free + Blood Pressure	190,801	62%
All 5 components (statewide average)	**137,985**	**45%**

Pearson Correlation Coefficients for Each Component

Optimal Diabetes Care and Blood Pressure

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
blood_pressure_component_rate	618	0.83170	0.07832	513.98962	0.46667	1.00000
diabetes_optimal_care_rate	618	0.43182	0.10228	266.86641	0.09045	0.63781

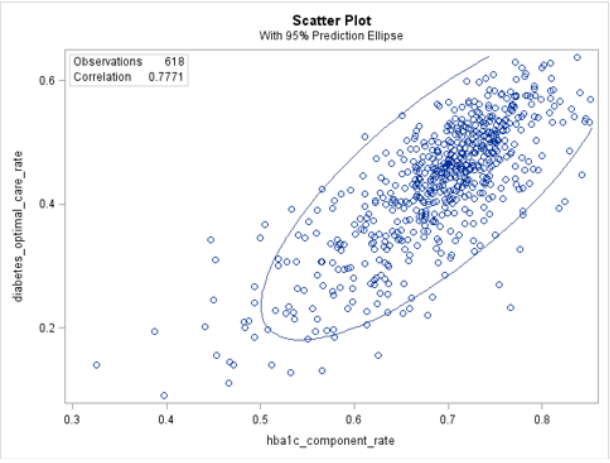
Pearson Correlation Coefficients, N = 618 Prob > |r| under H0: Rho=0	
	diabetes_optimal_care_rate
blood_pressure_component_rate	0.71245 <.0001



Optimal Diabetes Care and HbA1c

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
hba1c_component_rate	618	0.68937	0.07686	426.02911	0.32558	0.85128
diabetes_optimal_care_rate	618	0.43182	0.10228	266.86641	0.09045	0.63781

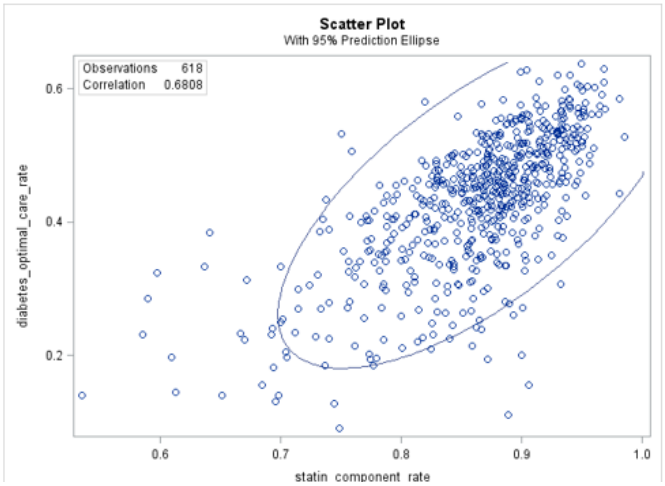
Pearson Correlation Coefficients, N = 618 Prob > |r| under H0: Rho=0	
	diabetes_optimal_care_rate
hba1c_component_rate	0.77714 <.0001



Optimal Diabetes Care and Statin Use

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
statin_component_rate	618	0.86328	0.06742	533.50798	0.53488	0.98571
diabetes_optimal_care_rate	618	0.43182	0.10228	266.86641	0.09045	0.63781

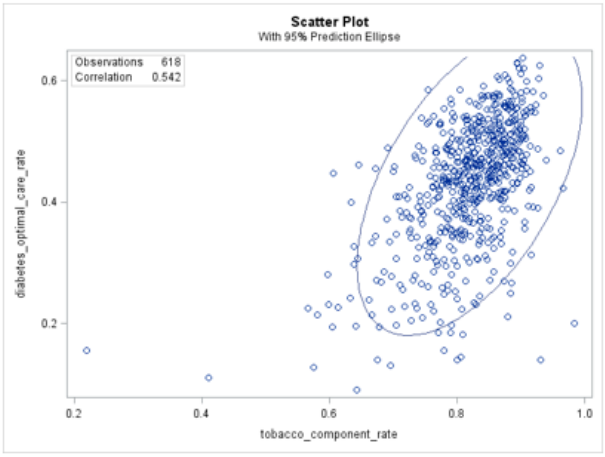
Pearson Correlation Coefficients, N = 618 Prob > |r| under H0: Rho=0	
	diabetes_optimal_care_rate
statin_component_rate	0.68083 <.0001



Optimal Diabetes Care and Tobacco Use

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
tobacco_component_rate	618	0.81948	0.07189	506.44022	0.21875	0.98333
diabetes_optimal_care_rate	618	0.43182	0.10228	266.86641	0.09045	0.63781

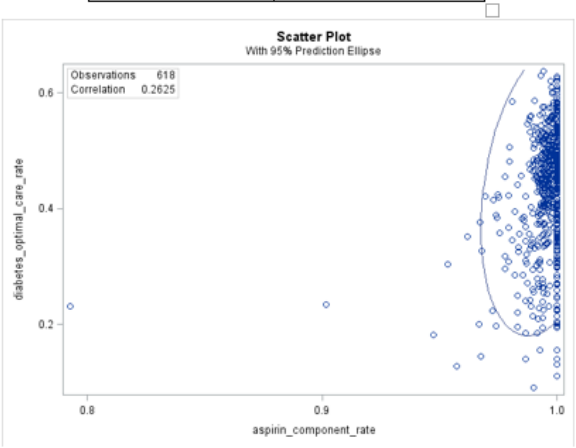
Pearson Correlation Coefficients, N = 618 Prob > |r| under H0: Rho=0	
	diabetes_optimal_care_rate
tobacco_component_rate	0.54201 <.0001



Optimal Diabetes Care and Aspirin

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
aspirin_component_rate	618	0.99495	0.01114	614.87973	0.79268	1.00000
diabetes_optimal_care_rate	618	0.43182	0.10228	266.86641	0.09045	0.63781

Pearson Correlation Coefficients, N = 618 Prob > |r| under H0: Rho=0	
	diabetes_optimal_care_rate
aspirin_component_rate	0.26253 <.0001



2d1.3. What is your interpretation of the results in terms of demonstrating that the components included in the composite are consistent with the described quality construct and add value to the overall composite? (i.*e., what do the results mean in terms of supporting inclusion of the components; if no empirical analysis, provide rationale for the components that were selected)*

A multifactorial approach to diabetes care that includes emphasis on blood pressure, lipids, glucose, aspirin use and non-use of tobacco will maximize health outcomes far more than a strategy that is limited to just one or two of these clinical domains (American Diabetes Association, 2014; Duckworth, 2009; Gaede, 2008; Holman, 2008a).

Four of the components have very high correction with the overall Optimal Diabetes Care result and Aspirin Use is effectively topped out so therefore does not have the same correlation as the other four components. In Minnesota, this measure and statewide efforts to increase the use of aspirin / antiplatelets have been in place for over 10 years and this component is essentially topped out, but this is not necessarily the case for other areas of the United States. Paerkh A.K. et al cite rates as low as 34.8% in primary care settings. [Aspirin in the Secondary Prevention of Cardiovascular Disease, NEJM Jan 2013)

Pearson Correlation Analysis

Components as compared to the composite Optimal Care Rate demonstrate a strong correlation with the following Pearson r coefficient values: Blood Pressure at 0.7124, Hemoglobin A1C at 0.7771, Statin Use at 0.6808, Aspirin Use at 0.2625 and Tobacco use at 0.5420

2d2. Empirical analysis demonstrating that the aggregations and weighting rules are consistent with the quality construct and achieve the objective of simplicity to the extent possible

NA

2d2.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

NA

2d2.2. What were the statistical results obtained from the analysis of the aggregation and weighting rules? (e.g., *results of sensitivity analysis of effect of different aggregations and/or weighting rules; if no empirical analysis, identify the aggregation and weighting rules that were considered and the pros and cons of each*)

NA

2d2.3. What is your interpretation of the results in terms of demonstrating the aggregation and weighting rules are consistent with the described quality construct? (i.*e., what do the results mean in terms of supporting the selected rules for aggregation and weighting; if no empirical analysis, provide rationale for the selected rules for aggregation and weighting*)

NA

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 by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition, Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)

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).

The majority of fields needed to calculate and risk adjust this measure are available in standardly available fields stored discretely in EHRs (e.g. date of birth, zip code, race/ ethnicity, provider NPI, provider specialty, insurance coverage, diagnosis codes- ICD, lab values, blood pressure values, smoking status, medications), however some data fields could be built into the EHR but may not be readily available unless a discrete field is built (e.g. some exceptions to statin such as breast feeding or women of child bearing years not actively taking birth control or aspirin/ antiplatelet use).

To date, MNCM has attempted to receive funding for technical assistance and development of an e-CQM. We are willing to move forward with e-CQM development upon appropriate funding and interest from key stakeholders.

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.

Over the last several years we have learned the following:

* Confidentiality- Patient confidentiality has been addressed by numerous mechanisms. MNCM only receives the patient level information needed to calculate the rates, determine eligibility for inclusion in the measure and support the administration of pay for performance programs. The PHI submitted is minimal and the data is protected by 1) password protection with password only available to the medical group submitting data, 2) file upload process is encrypted as data is transferred and 3) Data is stored on a separate secure server and meets all HIPAA protection rules.

* Specifications- Detailed specifications with instructions on how to handle most situations (e.g. detailed instructions on blood pressure values) has been valuable to medical groups, increased data accuracy and resulted in 98% of groups submitting data successfully.

* Pre-submission certification insures that medical group's queries are pulling the denominator population correctly prior to submission of patient level data to the HIPPA secure data portal for rate calculation.

* Audit- Audit methods have ensured the accuracy of our data and we are able to successfully compare providers because everyone is pulling their data the same way and subject to the same rules.

* Electronic Medical Record- It is easier for groups that have an electronic medical record to submit data and to submit their full population of patients, however groups with paper chart systems can successfully submit their sample.

* Acceptance of Data- Vast improvement in terms of sample sizes and timeliness of the data submitted by medical groups six weeks after the end of the measurement year as compared to prior method of health plan's samples and the results over a year old. Providers are more accepting of the results as compared to previous methods of pooling health plan samples.

* Health Plans: pay for performance and the inclusion of measures within contracts significantly impacts the number of groups participating in each measure (Diabetes, Ischemic Vascular, and Depression)

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)*.

There are no fees associated with participation and submitting data for this measure. Results are available to 1) all data submitters within the HIPAA secure MNCM data portal and 2) to the public on our consumer facing website MN Health Scores at www.mnhealthscores.org and 3) annual health care quality report on our corporate website at www.mncm.org. There are costs to the medical groups in terms of extract programs or abstraction to submit patient level clinical information for rate calculation.

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 MN Community Measurement- MN HealthScores Website http://www.mnhealthscores.org MN Community Measurement- MN HealthScores Website http://www.mnhealthscores.org Payment Program MN Bridges to Excellence http://mnhealthactiongroup.org/wp-content/uploads/2017/07/2017_MNBTE_winners.pdf MN Bridges to Excellence http://mnhealthactiongroup.org/wp-content/uploads/2017/07/2017_MNBTE_winners.pdf Regulatory and Accreditation Programs MN Department of Health- Statewide Quality Reporting and Measurement System http://www.health.state.mn.us/healthreform/measurement/index.html MN Department of Health- Statewide Quality Reporting and Measurement System http://www.health.state.mn.us/healthreform/measurement/index.html

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

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

MN Bridges to Excellence
Pay for Performance Program for top performers and those attaining improvement goals/ benchmarks. All clinics participating in MNCM's data portal who submit full population are eligible for inclusion in this program. Annual recognition event and rewards distributed.
MN Community Measurement- MN HealthScores Website
Public Reporting consumer-facing website
All primary care and endocrinology clinics in Minnesota (mandatory) and bordering communities (voluntary)
100 Medical groups representing 618 clinic sites; 2017 dates of service 307,158 patients with diabetes
MN Community Measurement- Health Care Quality Report

Public Reporting: Hard-copy report (pdf) highlighting top performers, most improved
100 Medical groups representing 618 clinic sites; 2017 dates of service 307,158 patients with diabetes
MN Department of Health- Statewide Quality Reporting and Measurement System
Based on 2008 health reform state legislation; this program requires mandatory submission of data from Minnesota physician clinics that have provider specialties that are applicable to the measured population. For the Optimal Diabetes Measure: family medicine, general practice, internal medicine, geriatric medicine and endocrinology.

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?*)

Not applicable, used in public reporting and accountability applications.

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.*)

Not applicable, used in public reporting and accountability applications.

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.

Performance results are provided to all medical groups who submit data for this state-wide measure via several options:

• Preliminary measure rates are provided immediately after file upload to HIPAA secure, password protected data portal

• A two-week review process is conducted to allow groups to review and potentially appeal prior to public reporting of rates

• Rates are reported by medical group and clinic level on public website MN Healthscores at www.mnhealthscores.org/

• Additionally, rates including all historical rates can be obtained from the MNCM data portal (pass-word protected)

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.

Currently, data is collected once per year and results are provided on an annual basis. See question 4a2.1.1. for the process and list of multiple mechanisms for receiving results and providing feedback.

MNCM provides recorded webinars for each measure or measure set that provides education for measure specification (denominator, numerator, exclusions) measure calculation and understanding results.

Education and explanation are also included in our hard copy reports. The annual Health Care Quality Report provides descriptive information along with the results for each measure plus appendices for guidelines for comparing measures over time, data sources and data collection, and methodology (attribution, weighting, rate calculation, risk adjustment). http://mncm.org/reports-and-websites/reports-and-data/health-care-quality-report/

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.

MNCM's Measure Review Committee (MRC) is tasked with the annual review of all publicly reported measures on MN Healthscores. As part of this process, which includes evaluation against NQF criteria (importance, scientific acceptability, feasibility and use), each measure is assessed for appropriateness to continue reporting. MRC recommendations are reviewed by MNCM's multi-stakeholder Measurement and Reporting Committee and the slate of publicly reported measures is approved by the MNCM Board of Directors http://mncm.org/news/slate-2018-measures/

For Optimal Diabetes Care review in 2018 Committee ratings were as follows:

On a four-point scale (1 = insufficient, 2 = low, 3 = moderate and 4 = high)

Evidence 3.71

Performance gap 3.57

Validity 3.43

Reliability 3.86

Feasibility and burden 3.86

Use and usability 3.57

Impact of the measure and level of effort (burden) was assessed on a 10-point scale

Does this measure help move the quality of care needle forward and improve health outcomes?

0 = no impact; 10 = extremely high impact

Committee impact rating was 8.50

How much time, effort and resources are needed for data collection, reporting and improving performance?

0 = no effort; 10 = extremely high effort

Committee effort rating was 6.33

The MRC voted unanimously to continue the measure without changes.

In May of 2018, MNCM's Measurement and Reporting Committee reviewed the BP Redesign Workgroup's rationale and recommendations for no change to the blood pressure target of < 140/90 and voted unanimously to accept the recommendation as presented.

The Optimal Diabetes Care measure was on the 2017 Measures Under Consideration list for CMS and in 2018 it was recommended for use by NQF Measures Application Partnership (MAP) Clinician and MAP Rural Health Workgroups.

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

Feedback from those being measured is obtained in several ways.

The MN Department of Health conducts several comment periods during its annual rule making process, those being measured are invited to comment. Comments are reviewed by measure development staff to identify areas of concern for potential redesign.

MNCM provides a year-round staffed support through a helpline 612-746-4522 or email support@mncm.org.

During the measure development process, formal public comment is sought from the clinics and medical groups who will be measured. All comments are reviewed by the measure development workgroup for identifying any redesign or tweaks to the measure specifications prior to pilot testing the measure. Pilot testing provides an additional source for feedback from users; pilot participants are surveyed with questions around feasibility and data element ease or difficulty.

MNCM conducts an annual medical group survey which all clinics in the state are invited to participate and provide feedback. There are structured questions asking the users about measure value and burden.

2018 Medical Group Survey

To what degree does your medical group find value in the measure? (n = 112)

High Value 45.5% (51)

Moderate Value 33.0% (37)

Minimal Value 14.3% (16)

No Value 7.1% (8)

How easy or difficult is it to obtain the data needed for DDS submission for this measure? (n = 104)

Very Easy 19.2% (20)

Easy 44.2% (46)

Difficult 29.8% (31)

Very Difficult 8.7% (9)

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

Please refer to feedback, rating and voting by the MNCM Measure Review Committee in question 4a2.2.1.

The Optimal Diabetes Care measure was on the 2017 Measures Under Consideration list for CMS and in 2018 it was recommended for use by NQF Measures Application Partnership (MAP) Clinician and MAP Rural Health Workgroups.

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

This measure, originally developed by HealthPartners with stewardship transferred to MNCM in ~ 2008, has undergone three component re-design activities based on changes in evidence and guidelines. Each re-design has involved a multi-stakeholder measure development workgroup who use a consensus-based decision making process. To recap briefly:

• The A1c component was changed from < 7.0 to less than 8.0 following ACCORD study results

• The cholesterol component was changed from LDL < 100 to appropriate statin use in 2015

• The blood pressure component has undergone some changes based on guidelines and alignment with other national measures until stabilizing at < 140/90 in 2010. Blood pressure component workgroup was convened in 2018 to review evidence and guideline change with the recommendation to remain at <140/90

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.

Since the start of public reporting of this measure in 2004, there has been steady improvement in composite rates for achieving all targets; statewide average from 9.5% to 44.9% with demonstrated variability and opportunity for improvement. What this translates to is 137,985 patients with diabetes in MN and surrounding communities are managing their modifiable risks and are less likely to develop long term complications.

HealthPartners, a large integrated health system in MN, has reduced the incidence of long-term complications of diabetes (heart attacks, amputations and blindness) by more than 30% since 1994 by focusing on optimal

diabetes care. Incidence of long-term complications per 1000 members between 2000 and 2016 fell from 17.8 to 11.3 for acute myocardial infarction, from 4.8 to 4.2 for amputations, and from 68.0 to 37.9 for retinopathy.

In 2018, a study of over 200,000 patients with type 2 diabetes found that excess risk of cardiovascular event outcomes decreased in a stepwise fashion for each risk factor variable that was within the target range. Patients who were successful in achieving targets for all five risk factors had little or no excess risk of death, myocardial infarction or stoke as compared to the general population. [Rawshani, A et al Risk Factors, Mortality, and Cardiovascular Outcomes in Patients with Type 2 Diabetes N Engl J Med 2018; 379:633-644 DOI: 10.1056/ NEJMoa1800256]

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.

Two unexpended findings over the years:

• Adults of Medicare age 65 and older with the potential for more co-morbidities, have better outcome rates than their younger counterparts with diabetes. This may be in part due to generational differences related to compliance with providers orders.

• We have noticed over time that statewide averages of A1c values are drifting upward and this is a trend also confirmed by the American Diabetes Association.

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

Statewide focus on the measure has led to a common way of measuring performance and communicating results. Many health systems create this measure internally and use for benchmarking and provider performance bonuses. Many clinics post their rates over time in the hallways for patients to view. Comparison provided through transparency provides competition and motivation to improve.

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)

0061 : Comprehensive Diabetes Care: Blood Pressure Control (<140/90 mm Hg)

0545 : Adherence to Statins for Individuals with Diabetes Mellitus

0575 : Comprehensive Diabetes Care: Hemoglobin A1c (HbA1c) Control (<8.0%)

2712 : Statin Use in Persons with Diabetes

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?

No

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

Denominator differences due to data source, different composite measure construct and philosophical beliefs of our measure development work group. Please see 5b.1.

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.)

2 measures are part of a composite measure that is stewarded by NCQA.

0061: Comprehensive Diabetes Care: Blood Pressure Control (<140/90 mm Hg)

0575: Comprehensive Diabetes Care: Hemoglobin A1c (HbA1c) Control (<8.0%)

NCQA's composite is a different measure construct; it is calculated at the physician panel level (what percentage of my patients have an A1c < 8.0, what percentage had BP < 140/90) but is not a patient level composite. MNCM believes that its patient level all-or-none composite is superior, patient-centric (not provider centric) and individual patients achieving as many health targets as possible only increases their likelihood of reducing long term microvascular and macrovascular complication of diabetes.

These two measure's numerators are harmonized.

We have philosophical differences in the denominator definitions and this is due in part to the data source. NCQA uses claims data to identify diabetic patients, MNCM used EMR based data. NCQA's methodology looks for diabetes diagnosis codes but additionally will include patients on oral medications and insulin who do not have the diagnosis. We also believe that is important to exclude diabetic women who are currently pregnant during the measurement year, related to cholesterol management. NCQA's denominator value sets intentionally include these patients.

This measure is related (but not exactly the same)

0545: Adherence to Statins for Individuals with Diabetes Mellitus (CMS)

Uses the same denominator definition as the NCQA composite. From information available in QPS, it does not appear that there are exceptions to this measure related to liver disease, rhabdomyolysis, pregnancy, etc. This is different from our planned cholesterol component for statin use. We believe our cholesterol component is superior in that it takes into account patient safety.

This measure is related (but not exactly the same)

2712: Statin Use in Persons with Diabetes (PQA)

This measure uses a different data source; pharmacy claims. Because the data source relies on filled prescriptions, the only way to identify the denominator is if the patient is on a diabetes drug, which does not encompass all diabetic patients that should be on a statin. Exclusions for this measure do not take into account the exceptions and contraindications for use of statins. We believe our cholesterol component is superior.

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): MN Community Measurement

Co.2 Point of Contact: Collette, Pitzen, pitzen@mncm.org, 612-454-4815-

Co.3 Measure Developer if different from Measure Steward: MN Community Measurement

Co.4 Point of Contact: Collette, Pitzen, pitzen@mncm.org, 612-454-4815-

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.

An expert panel was convened in December 2008 to determine the most appropriate A1c target for this composite. The group reviewed literature and incorporated current ICSI Diabetes Guideline discussions as this guideline was also undergoing revision. Members included:

Beth Averbeck, MD Assoc Medical Director, Health Partners, MNCM Board

Barry Bershow, MD Medical Director, Quality & Informatics, Fairview, Co-Chair MNCM Reporting Advisory Committee (RAC) and MNCM Board Member

Rich Bergenstal, MD Executive Director, International Diabetes Center, ICSI Diabetes Guideline Workgroup

John Fredrick, MD Exec Vice President & Chief Medical Officer PreferredOne, MARC Member

Gene Ollila, MD Allina Medical Clinic, ICSI Diabetes Guideline Workgroup

Expert panel was re-convened in March 2010 to address the aspirin component and again in July 2010 to address the blood pressure component of the composite measure. This technical advisory panel included:

Beth Averbeck, MD | HealthPartners

Barry Bershow, MD | Fairview Health Services

Rich Bergenstal, MD | International Diabetes Center, Park Nicollet

John Fredrick, MD | Preferred One

Gene Ollila, MD | Allina Medical Clinic

Linda Walling, MD | HealthEast

Mark Nyman, MD | Mayo Clinic

JoAnn Sperl-Hillen, MD | HealthPartners

Victor Montori, MD | Mayo Clinic

Kari Retzer | ICSI Facilitator for Diabetes Guideline

Measure development work group was convened March 2014 thru October 2014 to redesign the cholesterol component for both the Optimal Diabetes Care (#0729) and Optimal Vascular Care (#0076) measures whose previous target of LDL < 100 was no longer appropriate or supported by updated evidence and guidelines (American College of Cardiology/ American Heart Association Treatment of Cholesterol Guidelines Nov 2013). Members included:

Beth Averbeck, MD | Chair Internal Med MNCM Board | HealthPartners

Mark Nyman, MD | Internal Med MARC member | Mayo Clinic

Victor Montori, MD | Endocrinology | Mayo Clinic

JoAnn Sperl-Hillen, MD | Internal Med | HealthPartners

Courtney Baechler, MD | Cardiologist | Allina Penny George Institute

J. Ward Godsall, MD | Endocrinology | Allina Medical Group

Christopher Restad, DO | Family Medicine | Health East

Rebecca Moxness, MD | Endocrinology | Park Nicollet

Thomas Knickelbine, MD | Cardiologist | Minneapolis Heart Inst.

Woubeshet Ayenew, MD | Cardiologist | Hennepin County Med Cen

Terry Murray, RN | Data Analyst | Allina Medical Group

Jeanine Rosner, RN | QI or Clinic Admin | Park Nicollet

Monica Simmer | Health Plan | Metropolitan Health Plan

Pam York | State Agency | MDH/ Chronic Disease

Kris Soegaard | Consumer/ Empl/ MARC Member | MN Health Action Group

Collette Pitzen | Facilitator/ Measure Dev | MNCM

Measure development workgroup was convened in April of 2018 to evaluate and discuss recent changes in guidelines and evidence surrounding blood pressure targets for patients with diabetes and vascular disease.

Name | Member Type | Organization

Beth Averbeck, MD | Internal Medicine; Chair | HealthPartners

Joseph Bianco, MD | Family Medicine & MARC | Essentia Health- Ely

Andrew Greenland, MD | Internal Medicine | Mayo Clinic

Christopher Fallert, MD | Family Medicine | University of Minnesota

Christian Anderson, MD | Family Medicine | Entira Family Clinics

Steven Bradley, MD MPH | Cardiology | Minneapolis Heart

David Homans, MD | Cardiology | Park Nicollet

Jesse Wheeler, MD | Nephrology & MARC | Park Nicollet

Nicole Paterson, PharmD | Pharmacist | Fairview Health Services

Karen Margolis, MD MPH | Data Analyst | HealthPartners

Cindy Ferrara, RN | Quality Improvement | Essentia Health- Duluth

Patrick Schultz, ACNS-BC | Clinic Administrator | Sanford

James Peacock, PhD MPH | State Agency | MN Dept. of Health

Cynthia Toher, MD | Health Plan/ Cardiology | Blue Cross/Blue Shield MN

David Klocke, MD | Health Plan/Hosp Med | Blue Cross/Blue Shield MN

Christine Norton | Consumer and MARC | Retired

Deb Krause | Employer | MN Health Action Group

Collette Pitzen | Facilitator/ Measure Dev | MNCM

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2007

Ad.3 Month and Year of most recent revision: 10, 2015

Ad.4 What is your frequency for review/update of this measure? Annual, but can be more frequently as evidence emerges and guidelines change.

Ad.5 When is the next scheduled review/update for this measure? 06, 2019

Ad.6 Copyright statement:

Ad.7 Disclaimers:

Ad.8 Additional Information/Comments:

r

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 #: 3475e

Measure Title: Appropriate Use of DXA Scans in Women Under 65 Years Who Do Not Meet the Risk Factor Profile for Osteoporotic Fracture

Measure Steward: Centers for Medicare & Medicaid Services, Center for Clinical Standards and Quality, Quality Measurement and Value-Based Incentives Group (QMVIG), Division of Electronic and Clinician Quality, MS S3-02-01

Brief Description of Measure: Percentage of female patients 50 to 64 years of age without select risk factors for osteoporotic fracture who received an order for a dual-energy x-ray absorptiometry (DXA) scan during the measurement period.

Developer Rationale: This measure is expected to increase the recording of patient risk for fracture data and to decrease the number of inappropriate DXA scans. Current osteoporosis guidelines recommend using bone measurement testing to assess osteoporosis risk in women ages 65 and older. In postmenopausal women younger than 65, guidelines recommend using a formal clinical risk assessment tool to establish patients' risk for osteoporosis in order to determine whether to screen them for osteoporosis using bone measurement testing. Clinical information such as age, BMI, parental history of hip fracture, smoking, and alcohol use can be used to determine a woman's fracture risk (U.S. Preventive Services Task Force, 2018).

In addition, there are potentially avoidable harms associated with screening for osteoporosis in general, including exposure to radiation, false-positive exams, and the side effects of unnecessary osteoporosis medications, which add costs to an already burdened health care system (Lim et al., 2009).

Citations:

Lim LS, Hoeksema LJ, Sherin K. Screening for osteoporosis in the adult U.S. population: ACPM position statement on preventive practice. Am J Prev Med. 2009;36(4):366-75.

U.S. Preventive Services Task Force. Screening for osteoporosis to prevent fractures: U.S. Preventive Services Task Force recommendation statement." JAMA. 2018;319(24):2521-31.

Numerator Statement: Female patients who received an order for at least one DXA scan in the measurement period.

Denominator Statement: Female patients ages 50 to 64 years with an encounter during the measurement period.

Denominator Exclusions: The measure excludes patients who have a combination of risk factors (as determined by age) or one of the independent risk factors.

Measure Type: Process: Appropriate Use

Data Source: Electronic Health Records

Level of Analysis: Clinician: Individual

Preliminary Analysis: New Measure

Criteria 1: Importance to Measure and Report

1a. Evidence

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

The developer provides the following evidence for this measure:

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

Evidence Summary

- This is an overuse measure aiming to decrease inappropriate DXA screenings for osteoporosis and reduce avoidable harms associated with screening patients who have a low risk of osteoporotic fractures. The measure is based on the 2018 USPTF guideline, which is based on Grade B evidence: "The USPSTF recommends screening for osteoporosis with bone measurement testing to prevent osteoporotic fractures in postmenopausal women younger than 65 years who are at increased risk of osteoporosis, as determined by a formal clinical risk assessment tool."
- About 40% of women who have received a DXA scan do not meet risk factors for frailty, and may receive inappropriate medication and treatment for osteoporosis or osteopenia; the developers cite a study that showed that up to two-thirds of newly prescribed osteoporosis medications were given based on abnormalities identified using DXA scans that do not meet clinical guidelines for diagnosis.
- Potential harms caused by overuse of screening for osteoporosis include "false-positive test results, which can lead to unnecessary treatment, and false-negative test results" as well as "radiation exposure from DXA and opportunity costs (time and effort required by patients and the health care system)."
- An evidence review conducted in 2018 captured 168 published articles of good or fair quality on screening for and treatment of osteoporotic fractures, risk assessment tools, and the efficacy of screening.

Exception to evidence

N/A

Questions for the Committee:

Is there at least one thing that the provider can do to achieve a change in the measure results?

Guidance from the Evidence Algorithm

Process measure based on systematic review (Box 3) → QQC presented (Box 4) → Quantity: high; Quality: moderate; Consistency: high (Box 5) → Moderate (Box 5b) → Moderate

The highest possible rating is moderate.

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

RATIONALE:

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

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

- In data on 7.5 million women from one large health plan, 6.7 percent of the women ages 50 to 64 had potentially inappropriate DXA scans.
- About 40% of women who have had a DXA scan do not meet the risk factors for frailty.
- A retrospective cohort study of 13 practices assessed the three-, five-, and seven-year incidence of inappropriate and appropriate DXA scans. This study revealed a three-year incidence of DXA scans of 18.4 percent in women ages 50 to 59 without osteoporosis risk factors, and 24.9 percent in women ages 60 to 64 without risk factors.

Disparities

Overuse rates vary by race, with white women and Asian women having higher rates of overuse.

Rates of potentially inappropriate DXA scans by age and race from three test sites – Percents of Scans
(calculated using earlier version of measure for ages 18-64)

Site	White	Black	Asian	Other	Missing
Site 1	0.11	0.07	0.12	0.05	0.08
Site 2	2.36	1.23	2.43	4.87	1.83
Site 3	2.79	2.67	1.76	1.72	2.28

There are also disparities in general use of DXA scans and osteoporosis care:

- A gender matched study on women ages 60 and older in primary care practices, only 29.8 percent of black women were referred for a DXA scan, compared with 38.4 percent of white women. Of the referred women, 20.8 percent of the black women had the scan, compared with 27.0 percent of the white women.
- Among included women with a diagnosis of osteoporosis, black women were less likely to receive medication (79.6 percent) than were white women (89.2 percent) ($p < 0.05$), controlling for both age and BMI. But there was no difference in the pattern of follow-up visits between the two races.
- The prevalence of osteoporosis differs across races and ethnicities. In 2010, an estimated 15.8 percent of non-Hispanic white women, 7.7 of non-Hispanic black women, and 20.4 percent of Mexican American women had osteoporosis of femoral neck or lumbar spine.

Questions for the Committee:

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

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

RATIONALE:

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

1a. Evidence

Comments*:*

** Yes the evidence relates well, is applied directly and is current.

**variation in care, and overuse of testing is shown.

**overuse of imaging in women less than65 without risk factors, and over treatment of women less than 65 with abnormalities not meeting clinical guidelines for treatment. Recent USPTF guideline cited.

**Decreasing unnecessary tests is a desired outcome.

**Inappropriate screening relates indirectly to the adverse outcome of treatment when harm exceeds benefit. I am aware of no new evidence. Inappropriate screening relates indirectly to the adverse outcome of treatment when harm exceeds benefit. I am aware of no new evidence.

**There is moderate evidence to support this overuse measure

**Evidence from literature review and the USPSTF report apply directly to the process of appropriate or inappropriate selection of women under 65 y/o to undergo DXA scans. The desired outcome is to provide the benefits of osteoporosis treatment when appropriate and avoid unnecessary expense, stress, and radiation for women not at significant risk. EHRs are utilized to identify the two populations.

**The evidence relates directly to this process measure. The rationale is to reduce the number of unnecessary DXA scans by requiring use of a risk assessment tool before ordering bone measurement.

**good evidence base supporting measure

**The evidence directly raes to the specific measure process.

**There is a direct relationship.

**The evidence supporting this measure is moderate.

1b. Performance Gap

Comments:

**Yes performance data present and it demonstrated a gap of care with opportunity of improvement and variation. Race/Ethnicity + Age also showed opportunity for improvement and variation.

**They cited both variation in practice, as well as significant rates of overuse. There were significant racial disparities.

**Overuse (18.4% in women 55-59, ~25% in women 60-63); overuse greater in caucasian and asian women.

**There is a gap demonstrated. Disparities among groups are also identified.

**Performance gap with disparities has been demonstrated.

**Performance gap and disparities exist. Performance measure will be useful to attempt to try reducing performance and disparities gaps.

**Data submitted from the literature and from studies at three health care institutions indicate DXA scans are overused in the target population, but at low rates. Both the studies and the literature indicate differences in the steps of care between population subgroups.

**Current performance data was provided. There is overall less than optimal performance for use of clinical osteoporosis risk assessment tools for determine appropriate referral for DXA scans. A national performance measure would provide the focus needed to improve appropriate referrals for bone measurement testing. Data on the measure by population subgroups was provided and showed that race and ethnicity factored in referrals and treatment.

**Current performance data was provided: 6.7% potentially unnecessary DXA scans done in women 50 to 64 y in a large health care system (over 7 million women); another study potentially 40% of DXA scans done in women who do not frailty criteria; finally a retrospective study done at three sites showing potentially inappropriate DXA scans in women between ages 18 and 64. This also showed a gap in care higher rates of overuse in Asian and white women. Also disparities in care noted in that treatment of osteoporosis is lower in African Americans and Latinos. A national performance measure would focus attention on proper use of DXA and awareness of the contributing risk factors.

**gap identified: 18% - 24% overuse varying by age group; racial disparities

**There is a performance gap based on age and race demonstrating potential disparities in the care being delivered.

**Current performance data is provided. It demonstrates a substantial gap in care and among women of differences races / ethnic groups.

**There is evidence of overuse of DXA scans, providing justification for this measure

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

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

Reliability

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

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

Validity

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

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

eCQM Technical Advisor review:

Submitted measure is an HQMF compliant eCQM	The submitted eCQM specifications follow the industry accepted format for eCQM (HL7 Health Quality Measures Format (HQMF)). **HQMF specifications** ☒ **Yes** ☐ **No**

Documentation of HQMF,QDM, or CQL limitations	N/A – All components in the measure logic of the submitted eCQM are represented using the HQMF, QDM, or CQL standards
Value Sets	The submitted eCQM specifications uses existing value sets when possible and uses new value sets that have been vetted through the Value Set Authority Center (VSAC).
Measure logic is unambiguous	Submission includes test results from a simulated data set demonstrating the measure logic can be interpreted precisely and unambiguously. – this includes 100% coverage of measured patient population testing with pass/fail test cases for each population

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

Evaluators: Staff

Evaluation of Reliability and Validity:

- Score level reliability testing was conducted, and the results indicate that measure is reliable for clinicians with at least 20 patients in the denominator.

Questions for the Committee regarding reliability:

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

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?

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

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

Evaluation A: Scientific Acceptability

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 3475e

Measure Title: Appropriate Use of DXA Scans in Women Under 65 Years Who Do Not Meet the Risk Factor Profile for Osteoporotic Fracture

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.**

RELIABILITY: TESTING

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

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

   ☐ **Yes** ☐ **No**

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

   Submission document: Testing attachment, section 2a2.2

   - Random split-half correlation, supplemented with bootstrapping. Done at LOA; three sites, approximately 126,000 women; also tested with claims data for 7.5 million women from one large health plan. Developer notes that there are potential issues with the sample used for testing:

     "We pursued testing sites that captured data elements for the measure in their existing EHR workflows. As a result, we recruited sites that could be considered advanced EHR users, suggesting that they are unlikely to be representative of the broader field of clinicians who treat the population of interest. Our approach thus offers evidence that the measure concept is achievable but does not provide conclusive evidence regarding the ability of all EHR users to implement these measures."

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

   Submission document: Testing attachment, section 2a2.3

   - Average reliability coefficient, for providers with at least 20 patients in the appropriate age range: 0.82. Indicates that measure is reliable for clinicians with at least 20 patients in the denominator.

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

   Submission document: Testing attachment, section 2a2.2

☒ **Yes**

☐ **No**

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

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

   Submission document: Testing attachment, section 2a2.2

   ☐ **Yes**

   ☐ **No**

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

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

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

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

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

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

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

    Score level testing with large sample showed reliability.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

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

    Submission document: Testing attachment, section 2b4.

    None

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**

    Submission document: Testing attachment, section 2b5.

    N/A

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

    Submission document: Testing attachment, section 2b6.

Distribution of missing data not explicitly tested but developer describes how the level of missing data can be found. Developer states:

> In the data files submitted by test sites, there was no distinction between a negative (for example, confirmation that the patient was diagnosed with osteoporosis) and missing data. Where sites reported data for at least one patient, we assumed that blank records indicated no relevant data for those patients. For example, we assumed a patient with no data indicating osteoporosis did not have osteoporosis; we did not exclude that patient from the denominator based on lack of data regarding osteoporosis.

- Kappa could not be calculated for all data elements at all sites due to low prevalence of many exclusions; however, they contribute to the measure's face validity.

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**

VALIDITY: TESTING

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

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

☐ **Face validity**

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

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

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

Submission document: Testing attachment, section 2b2.2

Developer drew a random sample and extracted data elements, and then compared with data manually abstracted data. Developer then assessed validity using kappa agreement. Results were stratified by site.

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

Submission document: Testing attachment, section 2b2.3

Chance-adjusted agreement between the EHR and manually abstracted data for the numerator was high at two sites (0.91 and 0.93) and extremely low at one site (-0.01: chart prevalence of 48.5% and EHR prevalence of 0.5%). The developer states this is "attributable to a lack of EHR documentation for DXA scans in structured fields"; however the feasibility scorecard, updated more recently, does indicate most data are available in structured fields in EHRs.

The developer states that chance-adjusted agreement for the denominator exclusions was not reliable due to low prevalence.

Staff concern: A lack of EHR documentation at one of the three testing sites raises concerns with the measure's validity and feasibility.

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

    Submission document: Testing attachment, section 2b1.

    ☒ **Yes**

    ☐ **No**

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

22. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?**

    NOTE that data element validation from the literature is acceptable.

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

    ☐ **Yes**

    ☒ **No**

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

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

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

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

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

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

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

ADDITIONAL RECOMMENDATIONS

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

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

2a1. Reliability-Specification

Comments:

**data element well defined along with codes-descriptors are provided. SDOH may impact ability to reach minorities--they addressed their concern as well.

**Most practices would have 20 patients in the denominator, which is the threshold for reliability. This is reasonable for implementation.

**Missing data

**Data elements are clearly defined. This measure could be consistently implemented.

**No identified concerns.

**No major concerns. Specifications are appropriate

**Reliability specificiations are clear. However, the experience at the four testing sites in which all sites could not complete data analysis with data already included and extratable from their EHRs raises concerns about consistent implementation

**According to the NQF staff review the data elements are clearly defined. The other specifications are clear. If the data can be put into the EHR it is likely to be consistently implemented.

**I have no concerns

**good reliability however, reliability was tested at what was considered "high level" EHR users which may not be representative of typical EHR users;

**The specifications are clear and should be able to be consistently implemented for measurement purposes.

**No concerns

**No concerns

2a2. Reliability-Testing

Comments:

**Not at this time

**No

**No

**None

**No concerns.

**No major concerns. Reliability is high

**See comment on 6.2a1.

**the developer indicated that test sites were chosen where the clinicians were thought to be expert users of an EHR. This indicated to them that the measure could be reliably implemented at least at these sites with advanced EHR users, but that it does not necessarily mean that all EHR users could implement the measure

**NQF staff indicated that reliability testing met if clinic had at least 20 patients in the denominator, therefore I have no concerns about this

**none

**No concerns with the reliability testing.

**It is not clear that all EHR practices will have access the needed data elements.

**No concerns

2b1. Validity-Testing

Comments:

**No

**No

**Missing data? and its impact.

**None.

**No concerns.

**Assumptions regarding lack of data-meaning lack of exclusions-may be problematic

**No

**Chance adjusted agreement for the numerator was performed at three sites correlating EHR data with manually abstracted data. Two of the three clinics had high correlations, one was very poor. The developers indicated that the problem was due to inadequate EHR support for entry of DXA information. Validity testing for the denominator compromised by low prevalence of the exclusions.

**The developers looked at random samples and correlate data abstracted from paper records and EHR data. Two of the three sites had high correlation; the third site had low correlation. Developers attributed that to lack of DXA documentation facility on the existing EHR. If the developer is correct, this issue will fade as most clinicians/groups use robust EHR's

**Developer makes assumption that if nothing in the record regarding osteoporosis then patient is assumed to not have risk factors and is therefore not excluded from denominator; perhaps review some of these cases specifically (by manual review) to confirm that they should not be excluded from denominator;

**No concerns with the validity testing.

**No

**No concerns

2b4-7. Threats to Validity
2b.4. Meaningful Differences

Comments:

**Not at this time

**The concern over missing data due to EHR is a concern.While this is not common, it can not be assumed it could not happen at other sites.

**perhaps as interpreted by the developer (no data interpreted as no osteoporosis)

**No majors threats.

**No recognized threats to validity.

**yes, this can be assumed to be a threat unless there's information to suggest otherwise.

**Missing data: It is assumed that a patient with no data indicating osteoporosis did not have osteoporosis. This is consistent with the current state of medical records but is not necessarily true.

**Higher scores will indicate higher potentially inappropriate referrals. 2b5 There is only one set of specifications. 2b6 Missing data is treated as lack of an osteoporosis diagnosis, or lack of adequate score needed to reliably order a DXA

**Missing data will be treated as "no osteoporosis". There is wide variation among clinicians/clinics in the use of DXA in women who do not meet the requirement for use of a formal clinical risk assessment tool. There is a large number of potential exclusions and for any one clinic the number of individuals who represent those exclusions may be low.

**see above

**I did not identify any threats to the validity of the measure or to measure results.

**2b.6 There was no distinction between a negative and missing data which may constitute a threat to validity.A lack of EHR documentation at one of the three testing sites raises concerns with the measure’s validity and feasibility.

**No concerns

2b2-3. Other Threats to Validity
2b2. Exclusions
2b3. Risk Adjustment

Comments:

**Exclusions are consistent with evidence.

**n/a

**Deemed no applicable by developer

**Exclusions are appropriate.

**No recognized other threats to validity.

**No other issues or concerns

**Exclusions appear to be consistent with the current evidence and complete. No risk adjustment or stratification.

**2b2 No groups are excluded. 2b3 there is a conceptual relationship between potential social risk factor variables and the measure focus. Potentially inappropriate DXA scans are performed more frequently in Asian and white women than in African American and Latino women.

**Exclusions are consistent with evidence. There are some women who lose bone mineral rapidly at menopause who do not have any of the accepted risk factors. There is no way to capture this population at present without bone measurement. There is a clear conceptual relationship between potential social risk factor variables and the measure focus. The risk-adjustment variables were present at the start of care. Risk adjustment was properly developed and tested. Results are acceptable and there is an appropriate risk-adjustment strategy included in the measure.

**none

**There appeared to be some challenges with abstacting data correctly from the EHR and some conflicting information on whether that was due to the EHR field structure or how information was documented in the EHR.

**Certain exclusions (gastric bypass for example) could not be completely evaluated.

**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.

- The data for this measure are generated/collected by healthcare providers during the provision of care.
- All data elements are available in EHRs.
- The measure is an eMeasure and has been reviewed by the eMeasure team.
- The FRAX can be accessed online for free or a desktop edition can be purchased. The measure is available for public use and there are no other fees associated.

Feasibility Testing	Number of data elements included in measure calculation: 41 Number of data elements scoring less than 3 on scorecard: 31 **Questions for the Committee:** Consider the following questions for each data element that scored less than a 3 in any of the feasibility domains: *How is the data element used in computation of measure?* *How the data element is feasible within the context of the measure logic?* *What is the plan for readdressing the data element?* The following data elements scored less than 3 in the Workflow domain at one of the four sites: • Diagnosis: Type 1 Diabetes • Diagnosis: Rheumatoid Arthritis • Diagnosis: Psoriatic Arthritis • Diagnosis: Ankylosing Spondylitis • Diagnosis: Ehlers Danlos Syndrome • Diagnosis: Marfan's Syndrome • Diagnosis: Osteopenia • Diagnosis: Osteoporosis • Diagnosis: Osteogenesis Imperfecta • Diagnosis: Osteoporotic Fractures • Diagnosis: Cushings Syndrome • Diagnosis: Lupus • Diagnosis: Hyperthyroidism • Diagnosis: Hyperparathyroidism • Diagnosis: Chronic Liver Disease • Diagnosis: Chronic Malnutrition • Diagnosis: Malabsorption Syndromes • Diagnosis: End Stage Renal Disease Data Element 1 • List low scoring domains: Availability – Accuracy – Standards - Workflow How is the data element used in computation of measure? How the data element is feasible within the context of the measure logic? What is the plan for readdressing the data element? The following data elements scored less than a 3 in the Availability, Accuracy, Standards, and Workflow domains in at least of two of the sites: • Risk Category Assessment: History of hip fracture in parent Feedback from developer's Feasibility Assessment: *Clinicians at three sites did not collect the history of hip fracture in a parent in a structured field in the EHR. Two of these sites did not have a structured field for this element, and they did not consistently inquire about it as part of the clinical workflow. The other site could capture this element in a structured field, but clinicians did not always ask about it as part of the patient's medical history. Therefore, the record was not always accurate and would require a workflow*

change to ensure routine documentation of the data element by all providers.

- Risk Category Assessment: Ten-year probability of all major osteoporosis related fracture (FRAX Score)

  Feedback from developer's Feasibility Assessment:
  One of the four sites used FRAX to determine whether to order a DXA scan, but clinicians at this site did not document the FRAX score in a structured field in the EHR. Test-site staff noted that if clinicians were to start documenting this score in the EHR, it would most likely be entered as free text.

  Clinicians at two other sites used the FRAX tool—but not to determine when to order DXA scans. Instead, they typically ordered DXA scans and calculated the FRAX score afterward, using information from the scans (such as bone-mineral density) as an input. Clinicians cited two reasons for using the FRAX tool after receiving the scan results: (1) some believe that bone-mineral density was a required input to calculate the FRAX score, although it is actually optional, and (2) clinicians felt that the FRAX score would be more accurate if the DXA scan results were included. At one of these sites, clinicians typically entered the score as free text in the EHR, which was linked to a diagnosis (such as osteoporosis) and a date and time. Clinicians at the other site entered the score in a structured field.

  Staff at the fourth site, which was not using the FRAX tool, said that they hoped the tool would be incorporated into the clinical workflow and EHR in the next one or two years, but they noted that the scores would also most likely be entered post-DXA scan to determine the appropriate treatment for patients.

The following data element scored less than 3 in the Data Accuracy and Workflow domains in 2/4 sites:

- Risk Category Assessment: Average Number of Drinks per Drinking Day

  Feedback from developer's Feasibility Assessment:
  Clinicians at three sites administered the Alcohol Use Disorders Identification Test (AUDIT) as a screening questionnaire for alcohol abuse in patients, and this questionnaire includes a question about the average number of drinks per drinking day. Clinicians at one of these sites only recently started administering the AUDIT for new patients and estimated that the results were available in their EHR only for 25 percent of patients. Clinicians at the fourth site documented the average drinks per day, but not per drinking day, in a structured field in the EHR. As with all measures that require self-reported information on substance use, data accuracy is an issue at all of these sites because patients might not provide truthful answers about their use. However, where AUDIT is consistently used and the results are stored in structured fields, the data element is available and feasible to extract.

The following data element scored less than 3 in the Data Accuracy domain in 3/4 sites:

- Medication, Active: Aromatase Inhibitors
- Medication, Order: Aromatase Inhibitors
- Medication, Active: Glucocorticoids (oral only)
- Medication Dosage, Glucocorticoids (oral only)
- Medication Duration: Glucocorticoids (oral only)

Feedback from developer's Feasibility Assessment:

The measure excludes patients who have taken aromatase inhibitors at any point during their history and considers whether a patient has taken 5 mg per day or more of oral glucocorticoids

	over a period of at least 90 days at any point during their history. Although sites captured active medications and medication orders in structured fields in their EHRs, test-site staff said that medication reconciliation does not always occur. Therefore, the EHR might not accurately reflect when patients stop taking medications. Medication history for new patients or patients seen by external providers might also be incomplete in the EHR. *Clinicians at one site could find medications that had been e-prescribed or could manually enter new patients' medication history from transferred medical records. At another site, clinicians could request a list of medications for the patient from the pharmacy, but only for the past two years. A third site switched to a new EHR in January 2017, and site staff said that previous medical information was transferred inconsistently, resulting in an incomplete medication history for some patients.* *In addition, none of the sites captured the daily dosage of active or ordered medications in a structured field, but providers routinely documented prescription quantity, strength (for example, 5 mg per pill), and number of refills in structured fields. The frequency of medications (for example, two pills a day) was documented as free text at three sites. Because not all inputs necessary to calculate daily dosage are available in structured fields of the EHR, manual calculation would be required to determine the daily dosage for oral glucocorticoids, and these calculations would be subject to error. Furthermore, two sites did not have structured fields for the stop and start dates of medications; practices would therefore need to calculate the duration of active and ordered medications based on refill dates, which could reduce accuracy.* The following data element scored less than a 3 in the Availability, Accuracy, Standards, and Workflow domains at two sites: • Procedure, Performed: Gastric Bypass Surgery Feedback from developer's Feasibility Assessment: *Two of the four sites captured gastric bypass surgeries in structured fields of the EHR and indicated that it was feasible to use this data element to exclude patients from the measure. The other two sites, both using GE Centricity, did not consistently capture gastric bypass surgery in structured fields. Clinicians at one of the sites documented gastric bypass surgery as free text, and staff at the other site said that clinicians did not always ask about gastric bypass surgery, so documentation depended on whether the patient volunteered the information or if the clinician was involved in the patient's care at the time of the procedure.* The following data elements scored less than a 3 in the Data Standards domain at one site: • Encounter, Performed: Face-to-Face Interaction • Patient Characteristic Race: Race • Patient Characteristic Payer: Payer • Patient Characteristic Ethnicity: Ethnicity

Questions for the Committee:

- Does the Committee think the identified feasibility issues are fixable, as suggested by the developer, or do they raise larger concerns around the measure's overall feasibility?
- Is it reasonable to assume providers will be able to modify EHRs and/or clinical workflows to accurately report the measure?
- Do the developer's plan for the issues encountered in testing suffice?

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

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

RATIONALE:

- This measure scored low on feasibility due to a number of potential issues. NQF staff have identified potential issues the Committee should discuss.
- The measure received mixed results at the four testing sites. None of the sites were able to fully implement the measure. However, the developer notes this is a new measure and that minor changes to workflow or products should allow sites to capture all of the data elements needed.
- There are 41 data elements included in the measure, of which 31 scored less than three in at least one of the four testing sites.
- NQF's eMeasure Feasbility Report states that if any data element scores as 1, the data element has low feasibility, regardless of summary scores. Three of the data elements scored a 1 at all four test sites.

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. Feasibility

Comments:

**well documented and tested

**not all reports come back in form of structured data, so could require manual entry. frax tool is not regularly documentee in emr, so could create extra steps.

**Screening tools to assess risk are, in many practices not part of the EHR. Using risk calculators can provide estimate of patient risk but may not be found in a common data element in the EMR, thus leading to challenges with having the risk score that purportedly may have initiated the test.

**Barriers identified during feasibility assessment are FRAX calculation, Gastric bypass surgery, parental hip fx history. Largely due to not in a discrete field in the ehr.some variability due to type of ehr used.

**No identified concerns about feasibility.

**The identified feasiblity issues may be problematic. Additional discussion during committee call is needed. Providers are able to modify EHRs and clinical workflows to accurately report the measure; however, whether they WILL is another issue entirely. The developers plan to counter the issues seem sufficient, but further discussion during committee call is needed.

**Feasibility is worrisome as many of the required data elements (including diagnoses leading to exclusions & current medications lists were not available to be extracted by the EHR. This is proposed as a MIMS addition. I suppose if this proposed measure does not require much more "manual" extraction of data by reading the record for free text entries, etc., than current MIPS measures, then this problem might not be a "deal breaker".

**The required data elements are routinely generated during care delivery. If the clinic has a robust EHR, then the required data elements should be available. This measure can be put into operational use when the proper EHR software installed

**5 exclusion criteria (FRAX, hip fracture in parent, gastric bypass, medication reconciliation, # drinks/day) were difficult to capture and could influence results; the developer needs to reconcile how these 5 elements could be captured in spite of EHR challenges;

**There was significant inconsistency across testing sites indicating challenges with the feasibiity of collecting the data necessary to calculate the measure.

**I'm concerned about inconsistent collection of the history of hip fracture in a parent in a structured field in the EHR. Also, inconsistent use of FRAX tool, that medication reconciliation does not always occur, and thatnone of the sites captured the daily dosage of active or ordered medications in a structured field.

**Concerns were raised over feasibility, but these concerns seem to be addressable

Criterion 4: Usability and Use

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

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

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

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

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

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

OR

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

Accountability program details

This measure will be in MIPS starting in 2019.

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

Feedback on the measure by those being measured or others

The measure is not yet in use so hasn't been shared with anyone being measured. However, results were shared with test sites during testing as well as with the technical expert panel and a DXA Overuse expert work group. None of these groups had any significant concerns about their performance/clinician performance on the measure.

Additional Feedback:

Not available

Questions for the Committee:

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

Preliminary rating for Use: ☒ **Pass** ☐ **No Pass**

RATIONALE:

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

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

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

Improvement results

The measure is not yet in use so no improvement results were submitted.

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

This measure has not yet been implemented so no results are available, but the developers note that it should encourage providers to follow the USPSTF guideline because it will encourage the use of clinical risk assessment tools and because it may "increase clinicians' consistency in determining which patients are at high risk for osteoporotic fracture—and therefore eligible for a DXA scan."

Potential harms

Potentially, the measure could cause women who do not have the risk factors identified, or not enough of them, to miss needed DXA scans and therefore not receive or be delayed in receiving needed treatment. Also, the screening tool (FRAX) has not yet been widely studied in nonwhite groups, so women of color could not receive appropriate treatment, or have delays in receiving treatment.

Additional Feedback:

Not available

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

4a.1 Use-Accountability and Transparency

Comments:

**not public; feedback from the technical expert panel and a DXA Overuse expert work group--one of these groups had any significant concerns

**unsure

**Does not appear to be publicly reported nor does there appear to be opportunities for feedback

**Credible plan. Feedback not yet given as not in use but centers in feasibility given feedback and concerns not voiced .

**I'm not aware of public reporting or feedback to those being measured.

**It is an ambitious measure, with feasibility problems. The issues related to "USE" can be addressed only if the solutions to address feasibility are implemented, and if they work.

**Credible plan for implementation. "The two test sites did not share any significant concerns about their performance on their measure." Were providers and staff asked to comment on the process of the measuring?

**There are 41 data elements required for this measure. There were 31 that scored less than 3 in the workflow domain of 1 of 4 clinics. About 19 data elements in particular scored less than 3 at one clinic. Workflow change will have to occur at 2 of the 4 clinics to accurately indicate hip fracture in a parent. All four of the clinics had some issue, actually different issues employing the FRAX tool at present. Alcohol use (drinks per day) and glucocorticoid and aromatase inhibitor use could not always be reliably assessed from review of the EHR. Gastric bypass surgery data capture was inadequate. NQF staff commented LOW feasibility. Developer thinks all of this can be fixed with EHR upgrade.

**not publicly reported yet;

**The measure is being used in an accountability program and will be included in MIPS in 2019.

**No concerns.

**No concerns.

4b1. Usability-Improvement

Comments:

**I agree with their statement: “increase clinicians’ consistency in determining which patients are at high risk for osteoporotic fracture—and therefore eligible for a DXA scan.”

**would require working with emr vendor to have better way to document use of frax tool

**The measure has importance for over utilization of testing, over utilization of treatment, with potential for long term medication use that can be expensive with little overall patient benefit. Mitigation of individual harm and improving healthcare value are important potential benefits.

**Reducing unnecessary testing would contribute to high-quality case. Only potential harm is reduction in DXA use among in patient appropriate for the test.

**I am not aware of actual unintended consequences.

**Uninted harms are appropriately listed. No additional comments from me.

**The rationale of how results would further healthcare improvement seems solid. As described, benefits outweigh harms.

**If this measure can be successfully used DXA overuse will be reduced. Hopefully more individuals who would benefit from getting the test and subsequent treatment will be served. The unintended consequence is that women who experience rapid and severe bone mineral loss after menopause who do not meet the requirement for score/age or special disease consideration will be missed. Overall, considering that many DXA are currently incorrectly performed and reported and that inappropriate treatment may result, the benefits outweigh the harms.

**not publicly reported yet; there is need (not necessarily responsibility of the developer) to more carefully study use of FRAX across different populations (race disparity)

**The results of the measure could reduce unnecessary testing, inappropriate or unneeded treatment. May also create cost-efficiencies in treating osteoporosis.

**My only concern is the measure could cause women who do not have the risk factors identified, or not enough of them, to miss needed DXA scans and therefore not receive or be delayed in receiving needed treatment.

**No concerns.

Criterion 5: Related and Competing Measures

Related or competing measures

This measure is related to 0046 Screening or Therapy for Osteoporosis for Women Aged 65 Years and Older: Percentage of female patients aged 65-85 years of age who ever had a central dual-energy X-ray absorptiometry (DXA) to check for osteoporosis.

The developer states that the two measures complement each other.

Harmonization

The measures are harmonized to the extent possible, but have significant differences:

- The measures are different levels of analysis: 0046 is for claims and registry LOA; this measure is clinician LOA
- The measures have different intents: 0046 assesses documentation of DXA results, and is limited to DXA scans of the hip or spine (central DXA scans); 3475e assesses DXA orders for both central and peripheral scans
- The measures cover different populations: 0046 is for women ages 65 and older; 3475e is for women under 65

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

5. Related and Competing

Comments:

**I agree with their statement: This measure is related to 0046 Screening or Therapy for Osteoporosis for Women Aged 65 Years and Older: Percentage of female patients aged 65-85 years of age who ever had a central dual-energy X-ray absorptiometry (DXA) to check for osteoporosis. The developer states that the two measures complement each other.

**No

**no need for harmonization

**The other measure regarding DXA screening in patients 65-85 is in a different age group so doesn't compete or harmonize.

**I am not aware of any competing measures.

**Complements 0046 well

**NQF 0046 is different but related and not competing.

**#3475e is related to #0046 screening or rx for Osteoporosis in women >= 65: % females 65-85 who ever had a central DXA to check for Osteoporosis. The developer states these measures have been harmonized to the extent possible but they differ in level of analysis, intents, and population age. No additional harmonization is possible without major changes in the measures

**no concerns

**I am not aware of competing or related measures.

**Measures appear complementary.

**No concerns

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: January/25/2019

- **No NQF members who have submitted a support/non-support choice**

Brief Measure Information

NQF #: 3475e

Corresponding Measures:

De.2. Measure Title: Appropriate Use of DXA Scans in Women Under 65 Years Who Do Not Meet the Risk Factor Profile for Osteoporotic Fracture

Co.1.1. Measure Steward: Centers for Medicare & Medicaid Services, Center for Clinical Standards and Quality, Quality Measurement and Value-Based Incentives Group (QMVIG), Division of Electronic and Clinician Quality, MS S3-02-01

De.3. Brief Description of Measure: Percentage of female patients 50 to 64 years of age without select risk factors for osteoporotic fracture who received an order for a dual-energy x-ray absorptiometry (DXA) scan during the measurement period.

1b.1. Developer Rationale: This measure is expected to increase the recording of patient risk for fracture data and to decrease the number of inappropriate DXA scans. Current osteoporosis guidelines recommend using bone measurement testing to assess osteoporosis risk in women ages 65 and older. In postmenopausal women younger than 65, guidelines recommend using a formal clinical risk assessment tool to establish patients' risk for osteoporosis in order to determine whether to screen them for osteoporosis using bone measurement testing. Clinical information such as age, BMI, parental history of hip fracture, smoking, and alcohol use can be used to determine a woman's fracture risk (U.S. Preventive Services Task Force, 2018).

In addition, there are potentially avoidable harms associated with screening for osteoporosis in general, including exposure to radiation, false-positive exams, and the side effects of unnecessary osteoporosis medications, which add costs to an already burdened health care system (Lim et al., 2009).

Citations:

Lim LS, Hoeksema LJ, Sherin K. Screening for osteoporosis in the adult U.S. population: ACPM position statement on preventive practice. Am J Prev Med. 2009;36(4):366-75.

U.S. Preventive Services Task Force. Screening for osteoporosis to prevent fractures: U.S. Preventive Services Task Force recommendation statement." JAMA. 2018;319(24):2521-31.

S.4. Numerator Statement: Female patients who received an order for at least one DXA scan in the measurement period.

S.6. Denominator Statement: Female patients ages 50 to 64 years with an encounter during the measurement period.

S.8. Denominator Exclusions: The measure excludes patients who have a combination of risk factors (as determined by age) or one of the independent risk factors.

De.1. Measure Type: Process: Appropriate Use

S.17. Data Source: Electronic Health Records

S.20. Level of Analysis: Clinician : Individual

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

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

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

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? Not applicable. This measure is not paired or grouped.

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

DXA_Evidence_Attachment_Final-636772656013050280.docx

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

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

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number

Measure Title: **Appropriate Use of DXA Scans in Women Under 65 Years Who Do Not Meet the Risk Factor Profile for Osteoporotic Fracture**

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

Date of Submission: 11/8/2018

Instructions

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

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

1a. Evidence to Support the Measure Focus

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

- Outcome: 3 Empirical data demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service. If not available, wide variation in performance can be used as evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.

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

Notes

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

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

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

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

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

Outcome

☐ Outcome: Click here to name the health outcome

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

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

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

☒ Process:

☒ Appropriate use measure: Overuse of Dual-Energy X-Ray Absorptiometry (DXA) Scans in Women Under 65 Who Do Not Have Select Risk Factors for Osteoporotic Fracture

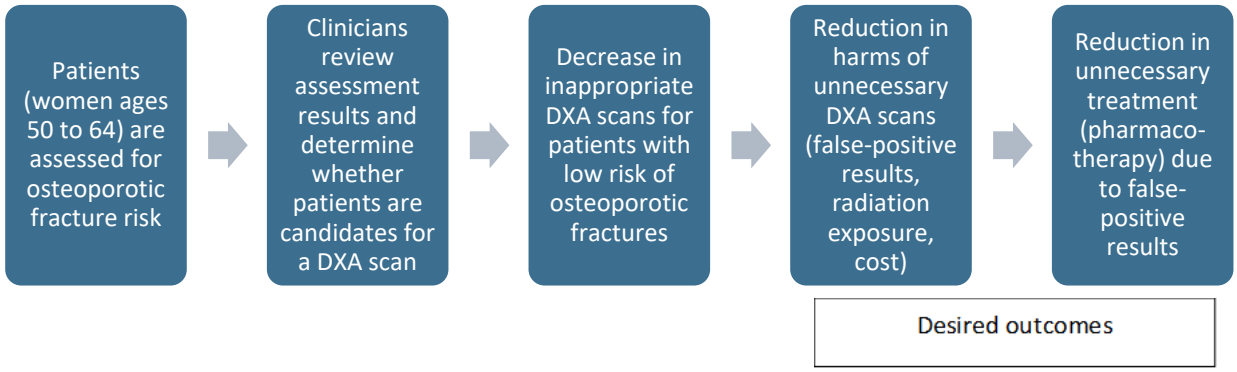
☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

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

The goals of this overuse measure are to (1) decrease inappropriate DXA screenings for osteoporosis and (2) reduce avoidable harms associated with screening patients who have a low risk of osteoporotic fractures.

The presumed pathway from process to outcomes is as follows:



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

Not applicable. This measure does not rely on patient-reported data.

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

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

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

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

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	**U.S. Preventive Services Task Force (USPSTF) recommendation:** • **Osteoporosis to Prevent Fractures: Screening** • **USPSTF** • **June 2018** • **USPSTF. Screening for osteoporosis to prevent fractures: U.S. Preventive Services Task Force recommendation statement. JAMA. 2018;319(24):2521-31.** • **https://jamanetwork.com/journals/jama/fullarticle/2685995** **Evidence review supporting USPSTF recommendation:** • **Screening to Prevent Osteoporotic Fractures: Updated Evidence Report and Systematic Review for the U.S. Preventive Services Task Force** • **Viswanathan, M., Reddy, S., Berkman, N., Cullen, K., Middleton, J., Nicholson, W., and Kahwati, L.** • **June 2018** • **Viswanathan M, Reddy S, Berkman N, Cullen K, Middleton J, Nicholson W, et al. Screening to prevent osteoporotic fractures: updated evidence report and systematic review for the U.S. Preventive Services Task Force. JAMA. 2018;319(24):2532-51.** • **https://jamanetwork.com/journals/jama/fullarticle/2685994**
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	"The USPSTF recommends screening for osteoporosis with bone measurement testing to prevent osteoporotic fractures in postmenopausal women younger than 65 years who are at increased risk of osteoporosis, as determined by a formal clinical risk assessment tool." (USPSTF, 2018)
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	The USPSTF does not grade the evidence. They review the evidence identified through the evidence review and determine if the benefits outweigh the harms. For the recommendation grade, see "Grade assigned to the recommendation with definition of the grade" below.
Provide all other grades and definitions from the evidence grading system	The USPSTF does not grade the evidence. They review the evidence identified through the evidence review and determine if the benefits outweigh the harms. For the grading system used by the USPSTF, see "Provide all other grades and definitions from the recommendation grading system" below.
Grade assigned to the **recommendation** with definition of the grade	"The USPSTF concludes with moderate certainty that the net benefit of screening for osteoporosis in postmenopausal women younger than 65 years who are at increased risk of osteoporosis is at least moderate." The USPSTF recommendation is a grade B recommendation. Grade B—There is high certainty that the net benefit is moderate, or there is moderate certainty that the net benefit is moderate to substantial.

Provide all other grades and definitions from the recommendation grading system	The USPSTF used the following system for grading the body of evidence: • Grade A—The USPSTF recommends the service. There is high certainty that the net benefit is substantial. • Grade B—Grade B is described above. • Grade C—The USPSTF recommends selectively offering or providing this service to individual patients based on professional judgment and patient preferences. There is at least moderate certainty that the net benefit is small. • Grade D—The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits. • I statement—The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.

Body of evidence:
• Quantity – how many studies?
• Quality – what type of studies?
	In 2018, Viswanathan et al. (2018) conducted a systematic review to support the USPSTF as it considered an update to its 2011 recommendation for osteoporosis screening. Viswanathan and colleagues reviewed the evidence published from November 2009 to October 2016 to identify evidence published since the 2011 review on screening for and treatment of osteoporotic fractures, risk assessment tools, and the efficacy of screening. Unless otherwise noted, we obtained information on the quality and quantity of the studies from this evidence review.
	
	Overall, the evidence review captured 168 published articles of good or fair quality.
	
	The USPSTF uses the following criteria to rate the quality of the evidence:
	
	"Randomized controlled trials and cohort studies
	• Initial assembly of comparable groups:
	  ○ For randomized controlled trials: adequate randomization, including first concealment and whether potential confounders were distributed equally among groups
	  ○ For cohort studies: consideration of potential confounders, with either restriction or measurement for adjustment in the analysis; consideration of inception cohorts
	• Maintenance of comparable groups (includes attrition, cross-overs, adherence, contamination)
	• Important differential loss to follow-up or overall high loss to follow-up
	• Measurements: equal, reliable, and valid (includes masking of outcome assessment)
	• Clear definition of interventions
	• All important outcomes considered
	• Analysis: adjustment for potential confounders for cohort studies or intention-to-treat analysis for randomized controlled trials
	
	<u>"Definitions of ratings based on [the] above criteria:</u>
	
	<u>Good:</u> Meets all criteria: Comparable groups are assembled initially and maintained throughout the study (follow-up ≥80%); reliable and valid measurement instruments are used and applied equally to all groups; interventions are spelled out clearly; all important outcomes are considered; and appropriate attention [is paid] to confounders in [the] analysis. In addition, intention-to-treat analysis is used for randomized controlled trials.
	
	<u>"Fair:</u> Studies are graded 'fair' if any or all of the following problems occur, without the fatal flaws noted in the 'poor' category below: Generally comparable groups are assembled initially, but some question remains [about] whether some (although not major) differences occurred with follow-up; measurement instruments are acceptable (although not the best) and generally applied equally; some but not all

	important outcomes are considered; and some but not all potential confounders are accounted for. Intention-to-treat analysis is used for randomized controlled trials. "Poor: Studies are graded 'poor' if any of the following fatal flaws exists: Groups assembled initially are not close to being comparable or maintained throughout the study; unreliable or invalid measurement instruments are used or not applied equally among groups (including not masking outcome assessment); and key confounders are given little or no attention. Intention-to-treat analysis is lacking for randomized controlled trials." (Viswanathan et al., 2018). The information on evidence quantity and quality is organized by key questions assessed in the evidence review. *Key question 1. Does screening (clinical risk assessment, bone density measurement, or both) for osteoporotic fracture risk reduce fractures and fracture-related morbidity and mortality in adults?* The authors identified one fair-quality controlled study that evaluated the effect of screening for hip-fracture risk and treatment. *Key question 2a. What is the accuracy and reliability of screening approaches to identify adults who are at increased risk for osteoporotic fracture?* Clinical risk assessment tools, like the FRAX, can be used to identify osteoporosis or to assess a person's risk for osteoporotic fracture. The authors reviewed the evidence for both uses of clinical risk assessment tools. However, this measure focuses on the use of the FRAX to assess fracture risk, and thus we present the review of the evidence specific to this use of risk assessment tools. The authors included one good-quality systematic review that assessed the accuracy of clinical risk assessment tools in predicting fracture in adults. This systematic review included 45 articles that assessed 13 risk-prediction tools. Twenty-six studies assessed the FRAX, six assessed the Garvan Fracture Risk Calculator, and four assessed the QFracture prediction tool. Other tools were assessed by only one or two studies each. The authors also identified and included in the evidence review 13 observational studies with low risk of bias or unclear bias that were not included in the systematic review, either because they were published after the systematic review search dates or because they were not identified or included in the systematic review. *Key question 3: What are the harms of screening for osteoporotic fracture risk?* The single fair-quality controlled study identified to answer this question is the same study referenced above in key question 1. *Key question 5: What are the harms associated with pharmacotherapy?* One of the potential harms from the overuse of a screening test is the downstream effects for patients who have a positive screening result. For DXA screening, a positive test (an osteoporosis diagnosis) could lead

	to the use of pharmacotherapy. The authors identified 16 fair- and good-quality studies reporting on the harms of alendronate, 4 fair- and good-quality studies on zoledronic acid, 6 fair-quality studies on risedronate, 2 fair-quality studies on etidronate, 7 fair-quality studies on ibandronate, 6 good-quality studies on raloxifene, 4 fair-quality studies on denosumab, and 1 fair-quality study on parathyroid hormone.

Estimates of benefit and consistency across studies	The USPSTF concluded "with moderate certainty that the net benefit of screening for osteoporosis in postmenopausal women younger than 65 years who are at increased risk of osteoporosis is at least moderate." This was partly based on the evidence for key questions 1, 2a, and 5. We provide the evidence reviews for these key questions below. The USPSTF does not have a specific recommendation on the overuse of DXA (that is, it does not explicitly state when not to screen women for osteoporosis). In the evidence review supporting the USPSTF recommendation, the authors assessed the evidence for the harms associated with osteoporosis screening (key question 3). They found one fair-quality controlled study that evaluated how screening for hip-fracture risk and treating those at high risk affects fracture rates in postmenopausal women ages 70 to 85. Participants in the intervention group were initially assessed for 10-year hip-fracture risk using the FRAX, and if the FRAX identified them as high risk, they were offered a DXA screening. Women were then offered treatment, as appropriate, based on the results of the DXA test and a revised FRAX (which incorporated the DXA results). This study showed no differences in anxiety or quality of life between participants in the intervention group versus the control group. However, the USPSTF notes that the potential harms of screening for osteoporosis include "false-positive test results, which can lead to unnecessary treatment, and false-negative test results" as well as "radiation exposure from DXA and opportunity costs (time and effort required by patients and the health care system)." Although the USPSTF did not specifically recommend against the use of DXA screening for osteoporosis in women at low risk for osteoporotic fracture, it did recommend osteoporosis screening (using bone measurement testing) only in postmenopausal women younger than 65 who are at increased risk of osteoporosis, "as determined by a formal clinical risk assessment tool." This measure attempts to identify women who are not at increased risk for osteoporotic fracture and assesses whether they were potentially inappropriately screened for osteoporosis using a DXA scan. One exclusion for the measure is a FRAX score indicating a high risk of osteoporotic fracture. *Key question 1. Does screening (clinical risk assessment, bone density measurement, or both) for osteoporotic fracture risk reduce fractures and fracture-related morbidity and mortality in adults?* The one study identified for this question is the same as the study identified for key question 3 (described above). According to the evidence review, "this study reported no significant difference in the primary outcome of any osteoporotic fracture in women screened with FRAX compared to women receiving usual care." In addition, the study did not show a statistically significant difference for all clinical fractures or mortality. However, the study did reveal a statistically significant lower incidence of hip fracture in the intervention group. *Key question 2a. What is the accuracy and reliability of screening approaches to identify adults who are at increased risk for osteoporotic*

	fracture? Across 12 studies that included 190,795 women, the accuracy of the FRAX (without the use of bone measurement density in the calculation) in predicting hip fractures for women was 0.76. This was similar to or higher than the accuracy rates for other clinical risk-prediction tools, which ranged from 0.52 to 0.71 (however, no other tools assess the risk of hip fracture specifically). *Key question 3: What are the harms associated with osteoporosis screening?* The study showed no differences in anxiety or quality of life between participants in the intervention group versus the control group. The USPSTF notes that the potential harms of screening for osteoporosis include "false-positive test results, which can lead to unnecessary treatment, and false-negative test results" as well as "radiation exposure from DXA and opportunity costs (time and effort required by patients and the health care system)." *Key question 5: What are the harms associated with pharmacotherapy?* • **Bisphosphonates**: "The USPSTF identified 16 studies on alendronate, 4 studies on zoledronic acid, 6 studies on risedronate, 2 studies on etidronate, and 7 studies on ibandronate that reported on harms. Overall, based on pooled analyses, studies on bisphosphonates showed no increased risk of discontinuation, serious adverse events, or upper gastrointestinal events." • **Raloxifene**: "Six trials of raloxifene therapy in women reported on various harms. Pooled analyses showed no increased risk of discontinuation due to adverse events or increased risk of leg cramps. However, analyses found a nonsignificant trend for increased risk of deep vein thrombosis, as well as an increased risk of hot flashes." • **Denosumab**: "Four studies reported on harms of denosumab therapy in postmenopausal women. Pooled analyses showed no significant increase in discontinuation or serious adverse events but found a nonsignificant increase in serious infections." • **Parathyroid hormone**: "A single study of parathyroid hormone therapy in women reported an increased risk of discontinuation and other adverse events, such as nausea and headache" (Viswanathan et al. 2018). In **Section 1a.4. Other Source of Evidence**, we provide information from other sources that demonstrates DXA overuse by clinicians and the unintended consequences of these scans.

What harms were identified?	The USPSTF does not identify significant harms of FRAX assessments or DXA scans. But the task force notes that "potential harms of screening for osteoporosis include false-positive test results, which can lead to unnecessary treatment, and false-negative test results" as well as "radiation exposure from DXA and opportunity costs (time and effort required by patients and the health care system)." For more information about the harms of unnecessary DXA scans, see Section **1a.4. Other Source of Evidence**.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	No additional studies were identified since the publication of the guideline.

1a.4 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

This measure was developed based on the American Academy of Family Physicians' Choosing Wisely recommendation statement on DXA for osteoporosis, which states, "Don't use dual-energy X-ray absorptiometry (DEXA) screening for osteoporosis in women under age 65 or men under 70 with no risk factors" (AAFP, n.d.). This recommendation also encourages clinicians to use a risk assessment tool, such as the FRAX, to determine the need for a DXA scan. Although the recommendation and additional studies described below address overuse of DXA scans in women below age 65, there are no quality measures that assess the overuse of DXA scans.

Evidence in Support of Appropriate DXA Use

Studies suggest that among women who have had a DXA scan, about 40 percent do not meet risk factors for frailty (Schnatz et al., 2011). As a result of the DXA scan, these women may receive inappropriate medication and treatment for osteoporosis or osteopenia. One study showed that up to two-thirds of newly prescribed osteoporosis medications were given based on abnormalities identified using DXA scans that do not meet clinical guidelines for diagnosis (Fenton et al., 2016). (For more information about harms associated with osteoporosis medication, see Section **1a.3. Systematic Review of the Evidence**.) Furthermore, in a study of 451 reports from DXA scans, 80 percent contained an error related to image data analysis (Messina et al., 2015). As patients typically consider bone-scan results to be definitive, this poses problems for overdiagnosis and overtreatment for osteoporosis and osteopenia because patients may not question the findings (Moynihan et al., 2017).

Despite the problems associated with DXA scans, overdiagnosis of osteoporosis and osteopenia, and subsequent inappropriate medication and treatment, clinicians continue to overuse these scans. A retrospective longitudinal analysis conducted across 34 practices showed no difference in the rates of DXA scan usage before and after the publication of the Choosing Wisely recommendation about DXA overuse (Lasser et al., 2016). The authors of this study suggest that "targeted initiatives addressing providers with high ordering rates will be needed to change behavior."

In addition, a retrospective cohort study of 13 practices assessed the three-, five-, and seven-year incidence of inappropriate and appropriate DXA scans. The study team found a three-year incidence of DXA scans of 18.4 percent in women ages 50 to 59 without osteoporosis risk factors, and 24.9 percent in

women ages 60 to 64 without risk factors (Amaranth et al., 2015). These studies suggest that a measure targeting appropriate use of DXA scans, as informed by a risk assessment tool, could improve care delivery.

Evidence for Exclusions

This measure includes three types of exclusions: (1) high risk of hip fracture as determined by a FRAX score, (2) conditions or patient characteristics that are used to determine a FRAX score (called "combination" risk factors), and (3) conditions or patient characteristics that are associated with a high rate of osteoporotic fracture. The table below shows risk factors that fall into the third group and that we identified in the literature as having high-risk ratios. Patients with ankylosing spondylitis, for example, have a relative risk of 7.1, which means that this condition is associated with a 700 percent higher chance of an osteoporotic fracture compared with a healthy person's chances.

Exclusion	Risk of fracture
Ankylosing spondylitis	7.1 odds ratio (OR; 95 percent confidence interval [CI]: 6.0–8.4) for vertebral fractures in patients with ankylosing spondylitis (Weiss et al., 2010).
Aromatase inhibitors	In studies comparing the use of aromatase inhibitors versus no aromatase inhibitors, the medication increased fracture risk by 17 percent in women under age 65 (95 percent CI: 1.07–1.28) (Tseng et al., 2018).
Cushing's syndrome	Patients with Cushing's syndrome were significantly more likely to report a low-energy fracture (a fracture occurring after minimal or no trauma) compared with controls (9.5 percent compared with 1.8 percent; $p = 0.004$) (Vestergaard et al., 2002).
Ehlers-Danlos syndrome	Previous fracture was 10 times more common in patients with Ehlers-Danlos syndrome ($p < 0.001$) than in other patients; 86.9 percent of patients with Ehlers-Danlos syndrome reported low-impact fractures (fractures of a peripheral bone) compared with 8.7 percent of controls (Dolan et al., 1998).
End-stage renal disease	4.11 standardized incidence ratio (95 percent CI: 2.96–5.73) for hip fracture in female Caucasian patients with end-stage renal disease; 3.35 standardized incidence ratio (95 percent CI: 2.59–4.40) for all female patients with end-stage renal disease in the study population (Stehman-Breen et al., 2000). For female patients ages 45 to 54 on dialysis in the study population, the observed/expected ratio was 20.0 (95 percent CI: 13.5–30.8) for hip fracture. For female patients ages 55 to 64 on dialysis in the study population, the observed/expected ratio was 10.2 (95 percent CI: 8.2–12.8) for hip fracture (Alem et al., 2000).
Gastric bypass	In patients with diabetes, gastric bypass had a hazard ratio of 1.26 (95 percent CI: 1.05–1.53) for risk of any type of fracture. In patients without diabetes, the hazard ratio was 1.32 (95 percent CI: 1.28–1.47) (Axelsson et al., 2018).
Hyperparathyroidism	Patients with primary hyperparathyroidism had a standardized incidence ratio of: • 3.2 (95 percent CI: 2.5–4.0) for vertebral fracture. • 2.2 (95 percent CI: 1.6–2.9) for distal forearm fracture. • 2.7 (95 percent CI: 2.1–3.5) for rib fracture. • 2.1 (95 percent CI: 1.2–3.5) for pelvic fracture (Khosla et al., 1999).
Lupus	Compared with similar-age women from a U.S. population sample, women ages 45 to 64 with lupus had a 7.6 standardized morbidity ratio for any self-reported fracture (95 percent CI: 5.1–10.7) (Ramsey-Goldman et al., 1999).
Marfan syndrome	No studies were identified assessing fracture risk in patients with Marfan syndrome. However, a large case-control study showed that patients with Marfan syndrome had lower bone mineral density compared with controls, independent of body mass index (Moura et al., 2006).
Osteogenesis imperfecta	Compared with a reference population, women with osteogenesis imperfecta had an incidence rate ratio of: • 5.9 (95 percent CI: 4.7–7.4) for any type of fracture in women 20 to 54 years old. • 8.0 (95 percent CI: 5.6–11.4) for any type of fracture in women 55 years old and older. • 1.6 (95 percent CI: 0.5–2.6) for spine fracture in women 55 years old and older. • 4.52 (95 percent CI: 2.79–6.26) for hip fracture in women 55 years old and older (Folkestad et al., 2017).

Exclusion	Risk of fracture
Psoriatic arthritis	Compared with controls, patients with psoriatic arthritis had a hazard ratio of 1.16 (95 percent CI: 1.06–1.27) for any type of fracture. For hip fracture, the hazard ratio was 1.17 (95 percent CI: 0.86–1.59). For vertebral fracture, the hazard ratio was 1.07 (95 percent CI: 0.66–1.72) (Ogdie et al., 2017).
Type 1 diabetes	Compared with controls, patients with type 1 diabetes are more likely to have a hip-fracture hospitalization (incidence rate ratio of 6.39; 95 percent CI: 1.94–22.35) and hip fracture (cause-specific hazard ratio of 7.11; 95 percent CI: 2.45–20.64) (Hamilton et al., 2017).

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

Initially, we constructed the measure to line up with the 2011 USPSTF recommendation and its supporting evidence. In April and May 2018, we developed a search string to capture literature focused on the overuse of DXA scans and searched PubMed for articles published since the release of the 2011 USPSTF guideline (January 2011 to January 2018). We searched for literature that addressed overuse of DXA scans in women under age 65 and also completed a clinical guideline scan for guidelines about DXA scans published in the United States, United Kingdom, and Canada.

To identify evidence for exclusions, we conducted a literature search for supplementary work to accompany the guidelines. The goal of the search was to identify independent factors that put a person at higher risk for fractures.

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

Alem AM, Sherrard DJ, Gillen DL, Weiss NS, Beresford SA, Heckbert SR, et al. Increased risk of hip fracture among patients with end-stage renal disease. Kidney Int. 2000 Jul;58(1):396-9.

Amarnath ALD, Franks P, Robbins JA, Xing G, Fenton JJ. Underuse and overuse of osteoporosis screening in a regional health system: a retrospective cohort study. J Gen Intern Med. 2015;30(12):1733-40. https://doi.org/10.1007/s11606-015-3349-8

American Academy of Family Physicians. Choosing Wisely: DEXA for osteoporosis recommendation. https://www.aafp.org/patient-care/clinical-recommendations/all/cw-osteoporosis.html. Accessed October 2, 2018.

Axelsson KF, Werling M, Eliasson B, Szabo E, Näslund I, Wedel H, et al. Fracture risk after gastric bypass surgery: a retrospective cohort study. J Bone Miner Res. 2018 Jul (published online ahead of print). doi: 10.1002/jbmr.3553

Dolan AL, Arden NK, Grahame R, Spector TD. Assessment of bone in Ehlers Danlos syndrome by ultrasound and densitometry. Ann Rheum Dis. 1998 Oct;57(10):630-3.

Fenton JJ, Robbins JA, Amarnath ALD, Franks P. Osteoporosis overtreatment in a regional health care system. JAMA Intern Med. 2016;176(3):391-3. https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2478896

Folkestad L, Hald JD, Ersbøll AK, Gram J, Hermann AP, Langdahl B, et al. Fracture rates and fracture sites in patients with osteogenesis imperfecta: a nationwide register-based cohort study. J Bone Miner Res. 2017 Jan;32(1):125-34.

Hamilton EJ, Davis WA, Bruce DG, Davis TME. Risk and associates of incident hip fracture in type 1 diabetes: the Fremantle Diabetes Study. Diabetes Res Clin Pract. 2017 Dec;134:153-60.

Kanis JA, Johansson H, Oden A, Johnell O, de Laet C, Melton III L, et al. A meta-analysis of prior corticosteroid use and fracture risk. J Bone Miner Res. 2004 Jun;19(6):893-9.

Khosla S, Melton LJ, Wermers RA, Crowson CS, O'Fallon W, Riggs Bl. Primary hyperparathyroidism and the risk of fracture: a population-based study. J Bone Miner Res. 1999 Oct;14(10):1700-7.

Lasser EC, Pfoh ER, Chang HY, Chan KS, Bailey JC, Kharrazi H, et al. Has Choosing Wisely® affected rates of dual-energy X-ray absorptiometry use? Osteoporos Int. 2016;27(7):2311-6. https://doi.org/10.1007/s00198-016-3511-0

Messina C, Bandirali M, Sconfienza LM, D'Alonzo NK, Di Leo G, Papini GDE, et al. Prevalence and type of errors in dual-energy X-ray absorptiometry. Eur Radiol. 2015;25(5):1504-11. https://doi.org/10.1007/s00330-014-3509-y

Moura B, Tubach F, Sulpice M, Boileau C, Jondeau G, Muti C, et al. Bone mineral density in Marfan syndrome: a large case-control study. Joint Bone Spine. 2006 Dec;73(6):733-5.

Moynihan R, Sims R, Hersch J, Thomas R, Glasziou P, McCaffery K. Communicating about overdiagnosis: Learning from community focus groups on osteoporosis. PLoS ONE. 2017;12(2),1-16. https://doi.org/10.1371/journal.pone.0170142

Ogdie A, Harter L, Shin D, Baker J, Takeshita J, Choi HK, et al. The risk of fracture among patients with psoriatic arthritis and psoriasis: a population-based study. Ann Rheum Dis. 2017 May;76(5):882-5.

Ramsey-Goldman R, Dunn JE, Huang CF, Dunlop D, Rairie JE, Fitzgerald S, et al. Frequency of fractures in women with systemic lupus erythematosus: comparison with United States population data. Arthritis Rheum. 1999 May; 42(5):882-90.

Schnatz PF, Marakovits KA, Dubois M, O'Sullivan DM. Osteoporosis screening and treatment guidelines: are they being followed? Menopause. 2011;18:1072-8.

Silverman SL, Calderon AD. The utility and limitations of FRAX: a US perspective. Curr Osteoporos Rep. 2010;8(4):192-7. https://doi.org/10.1007/s11914-010-0032-1

Stehman-Breen CO, Sherrard DJ, Alem AM, Gillen DL, Heckbert SR, Wong CS, et al. Risk factors for hip fracture among patients with end-stage renal disease. Kidney Int. 2000 Nov;58(5):2200-5.

Tseng OL, Spinelli JJ, Gotay CC, Ho WY, McBride ML, Dawes MG. Aromatase inhibitors are associated with a higher fracture risk than tamoxifen: a systematic review and meta-analysis. Ther Adv Musculoskelet Dis. 2018 Apr;10(4):71-90.

Vestergaard P, Lindholm J, Jørgensen JO, Hagen C, Hoeck HC, Laurberg P, et al. Increased risk of osteoporotic fractures in patients with Cushing's syndrome. Eur J Endocrinol. 2002 Jan;146(1):51-6.

Weiss RJ, Wick MC, Ackermann PW, Montgomery SM. Increased fracture risk in patients with rheumatic disorders and other inflammatory diseases—a case-control study with 53,108 patients with fracture. J Rheumatol. 2010 Nov;37(11):2247-50.

1b. Performance Gap

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

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

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

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

This measure is expected to increase the recording of patient risk for fracture data and to decrease the number of inappropriate DXA scans. Current osteoporosis guidelines recommend using bone measurement testing to assess osteoporosis risk in women ages 65 and older. In postmenopausal women younger than 65, guidelines recommend using a formal clinical risk assessment tool to establish patients' risk for osteoporosis in order to determine whether to screen them for osteoporosis using bone measurement testing. Clinical information such as age, BMI, parental history of hip fracture, smoking, and alcohol use can be used to determine a woman's fracture risk (U.S. Preventive Services Task Force, 2018).

In addition, there are potentially avoidable harms associated with screening for osteoporosis in general, including exposure to radiation, false-positive exams, and the side effects of unnecessary osteoporosis medications, which add costs to an already burdened health care system (Lim et al., 2009).

Citations:

Lim LS, Hoeksema LJ, Sherin K. Screening for osteoporosis in the adult U.S. population: ACPM position statement on preventive practice. Am J Prev Med. 2009;36(4):366-75.

U.S. Preventive Services Task Force. Screening for osteoporosis to prevent fractures: U.S. Preventive Services Task Force recommendation statement." JAMA. 2018;319(24):2521-31.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis*. (This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

This measure has not yet been implemented and does not have performance data. However, from testing, we have an indication of performance scores based on 2013 encounters across 269 primary care providers (PCPs) at two sites: a primary care practice in suburban Michigan and a large multispecialty group in New York. (We also contracted with a third site, a large multispecialty group in Maryland. However, this site independently conducted analyses based on 2012 encounters and sent its results to measure developers. The site did not provide clinician-level performance scores.) In addition, we have data from 2,508,693 female patients ages 50 to 64 who were covered by one large multistate health plan and had a DXA scan in 2012.

In data on 7.5 million women from one large health plan, 6.7 percent of the women ages 50 to 64 had potentially inappropriate DXA scans. Although these data could not be analyzed at the clinician level, we present them because they indicate how the measure might perform if implemented nationally. Please note that the claims analysis is based on DXA scans performed rather than on DXA scans ordered (as specified in the measure), so the numbers might be lower than they would be if the measure were implemented.

The clinician-level data presented below are from only two sites, and thus they may not be representative of national performance.

In EHR data from 269 PCPs at two sites, the rates of potentially inappropriate DXA scans varied from 0.0 to 100 percent. Performance was skewed left, with the top decile of performers (that is, the worst performers) ordering inappropriate DXA scans for at least 10 percent of patients in the denominator. These results suggest that about 10 percent of clinicians have room for improvement.

Among the 269 PCPs at the two sites, the performance rate statistics were as follows:

Mean: 3 percent

Standard deviation: 9 percent

Minimum: 0 percent

Maximum: 100 percent

Interquartile range: 0 to 0.5 percent

10th percentile: 0 percent

50th percentile: 0 percent

90th percentile: 10 percent

95th percentile: 19 percent

99th percentile: 33 percent

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.

Studies suggest that among women who have had a DXA scan, about 40 percent do not meet the risk factors for frailty (Schnatz et al., 2011). Studies also indicate that DXA scans are overused, albeit at low rates. A retrospective longitudinal analysis conducted across 34 practices showed no difference in the rates of DXA scan usage before and after the publication of the Choosing Wisely recommendation about DXA overuse; rates were 2.6 percent before and 2.0 percent after (Lasser et al., 2016).

In addition, a retrospective cohort study of 13 practices assessed the three-, five-, and seven-year incidence of inappropriate and appropriate DXA scans. This study revealed a three-year incidence of DXA scans of 18.4 percent in women ages 50 to 59 without osteoporosis risk factors, and 24.9 percent in women ages 60 to 64 without risk factors (Amaranth et al., 2015).

Citations:

Amarnath ALD, Franks P, Robbins JA, Xing G, Fenton JJ. Underuse and overuse of osteoporosis screening in a regional health system: a retrospective cohort study. J Gen Intern Med. 2015; 30(12):1733-40. https://doi.org/10.1007/s11606-015-3349-8

Lasser EC, Pfoh ER, Chang HY, Chan KS, Bailey JC, Kharrazi H, et al. Has Choosing Wisely® affected rates of dual-energy X-ray absorptiometry use? Osteoporos Int. 2016; 27(7):2311-6. https://doi.org/10.1007/s00198-016-3511-0

Schnatz PF, Marakovits KA, Dubois M, O'Sullivan DM. Osteoporosis screening and treatment guidelines: are they being followed? Menopause. 2011; 18:1072-8.

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

This measure has not yet been implemented and does not have performance data. To understand how performance on this measure varies by patient characteristics, we compared patient-level measure results by age and race in the three test sites for which we had EHR data. Two sites provided data from 2013 encounters; the third conducted its own analyses based on 2012 encounters and sent the results to the measure developers.

The results below summarize the rates of potentially inappropriate DXA scans by age and race from these sites. The rate was highest among women ages 60 and older across two sites (the third site merged results for women ages 50 to 64). At two sites, black women had significantly lower rates of potentially inappropriate DXA scans than white women. Please note that the results stratified by race were calculated using an earlier version of the measure that included women ages 18 to 64.

RATES ON POTENTIAL DXA-OVERUSE MEASURE, BY AGE AND SITE

Note: Rates were calculated using EHR extracts from three sites.

Site 1

Ages 50–59: 0.25 percent

Ages 60–64: 0.29 percent

Site 2 (Site 2 combined the data for patients ages 50 to 64 in a single age bracket.)

Ages 50–64: 5.70 percent

Site 3

Ages 50–59: 6.20 percent

Ages 60–64: 8.19 percent

Rates on potential DXA-overuse measure, by race and site

Note: Rates were calculated using EHR extracts from three sites for women ages 18 to 64.

Site 1

White—0.11 percent

Black—0.07 percent

Asian—0.12 percent

Other—0.05 percent

Missing—0.08 percent

Site 2

White—2.36 percent

Black—1.23 percent

Asian—2.43 percent

Other—4.87 percent

Missing—1.83 percent

Site 3

White—2.79 percent

Black—2.67 percent

Asian—1.76 percent

Other—1.72 percent

Missing—2.28 percent

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4

The literature also suggests disparities between black and white women with regard to DXA scans. In a gender matched study on women ages 60 and older in primary care practices, only 29.8 percent of black women were referred for a DXA scan, compared with 38.4 percent of white women ($p < 0.05$) (Hamrick et al., 2012). Of the referred women, 20.8 percent of the black women had the scan, compared with 27.0 percent of the white women ($p < 0.05$) (Hamrick et al., 2012). Also, among included women with a diagnosis of osteoporosis, black women were less likely to receive medication (79.6 percent) than were white women (89.2 percent) ($p < 0.05$), controlling for both age and BMI. But there was no difference in the pattern of follow-up visits between the two races (Hamrick et al., 2012).

Although the literature shows that all ethnicities are at risk for osteoporosis, the prevalence of osteoporosis differs across races and ethnicities. In 2010, an estimated 15.8 percent of non-Hispanic white women, 7.7 of non-Hispanic black women, and 20.4 percent of Mexican American women had osteoporosis of femoral neck or lumbar spine (Wright et al., 2014). Understanding these differences among women of different ethnicities is helpful as we continue to look at DXA scans in the population.

Citations:

Hamrick I, Cao Q, Aqbafe-Mosley D, Cummings DM. Osteoporosis health care disparities in postmenopausal women. J Womens Health. 2012 Dec;21(12):1232-6.

Wright NC, Looker AC, Saag KG, Curtis JR, Delzell ES, Randall S, et al. The recent prevalence of osteoporosis and low bone mass in the United States based on bone mineral density at the femoral neck or lumbar spine. J Bone Miner Res. 2014 Nov;29(11):2520-6.

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)

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

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

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

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

No link to the current specifications exist; the specifications are attached in accordance with Question S.2a.

S.2a. If this is an eMeasure, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)

This is an eMeasure **Attachment:** AppropriateDXAScan_v5_5_Artifacts-636687330076328450.zip, CMS249v1_Bonnie_test_cases-636687330189610329.xlsx, cms249bonnie_-002-.docx

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:** CMS249_ValueSets.xlsx

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

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

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

Not an instrument-based measure

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

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

Not applicable. This is a new measure.

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

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

Female patients who received an order for at least one DXA scan in the measurement period.

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).

Female patients who received an order for at least one DXA scan in the measurement period

Please refer to the attached Measure Authoring Tool (MAT) output and value sets.

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

Female patients ages 50 to 64 years with an encounter during the measurement period.

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.)*

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

Female patients ages 50 to 64 years with an encounter during the measurement period

Please refer to the attached MAT output and value sets.

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

The measure excludes patients who have a combination of risk factors (as determined by age) or one of the independent risk factors.

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.)*

Exclude patients with a combination of risk factors (as determined by age) or one of the independent risk factors

Ages: 50-54 (>=4 combination risk factors) or 1 independent risk factor

Ages: 55-59 (>=3 combination risk factors) or 1 independent risk factor

Ages: 60-64 (>=2 combination risk factors) or 1 independent risk factor

COMBINATION RISK FACTORS [The following risk factors are all combination risk factors; they are grouped by when they occur in relation to the measurement period]:

The following risk factors may occur any time in the patient's history but must be active during the measurement period:

White (race)

BMI <= 20 kg/m2 (must be the first BMI of the measurement period)

Smoker (current during the measurement period)

Alcohol consumption (> two units per day (one unit is 12 oz. of beer, 4 oz. of wine, or 1 oz. of liquor))

The following risk factor may occur any time in the patient's history and must not start during the measurement period:

Osteopenia

The following risk factors may occur at any time in the patient's history or during the measurement period:

Rheumatoid arthritis

Hyperthyroidism

Malabsorption Syndromes: celiac disease, inflammatory bowel disease, ulcerative colitis, Crohn´s disease, cystic fibrosis, malabsorption

Chronic liver disease

Chronic malnutrition

Documentation of history of hip fracture in parent

Osteoporotic fracture

Glucocorticoids (>= 5 mg/per day) [cumulative medication duration >= 90 days]

INDEPENDENT RISK FACTORS (The following risk factors are all independent risk factors; they are grouped by when they occur in relation to the measurement period):

The following risk factors may occur at any time in the patient's history and must not start during the measurement period:

Osteoporosis

The following risk factors may occur at any time in the patient's history:

Gastric bypass

FRAX[R] ten-year probability of all major osteoporosis related fracture >= 8.4 percent

Aromatase inhibitors

Type I Diabetes

End stage renal disease

Osteogenesis imperfecta

Ankylosing spondylitis

Psoriatic arthritis

Ehlers-Danlos syndrome

Cushing's syndrome

Hyperparathyroidism

Marfan syndrome

Lupus

Please refer to the attached MAT output and value sets.

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.)*

Not applicable. This measure does not use 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.)*

Refer to items S.4 to S.9 for details, S2.a for the eCQM specification, and S2.b for value sets.

1. Determine the denominator. Identify female patients ages 50 to 64 who had an encounter during the measurement period.

2. Remove exclusions. Identify patients who meet the exclusion criteria and remove them from the denominator (female patients who have a combination of risk factors, as determined by age, or one of the independent risk factors).

3. Determine the numerator. Identify patients in the denominator (after removing patients who meet the exclusion criteria) who received at least one DXA scan order during the measurement period.

4. Calculate measure performance. Compute performance as a proportion: numerator cases divided by (denominator minus exclusions).

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.

This measure is not based on a sample. It is based on a clinician's entire patient population.

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

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

Not applicable. This measure is not based on survey or patient-reported data.

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

If other, please describe in S.18.

Electronic Health Records

S.18. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data are collected.)*

IF instrument-based, identify the specific instrument(s) and standard methods, modes, and languages of administration.

Not applicable. This measure is not instrument-based. Data are collected from structured fields of eligible clinicians' electronic health records (EHRs).

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)*

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

Clinician : Individual

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

Outpatient Services

If other:

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

Not applicable. This measure is not a composite measure.

2. Validity – See attached Measure Testing Submission Form

CMS249_Testing_Attachment.docx

2.1 For maintenance of endorsement

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

2.2 For maintenance of endorsement

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

2.3 For maintenance of endorsement

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

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

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

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

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

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

n.a.

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

Three sites provided electronic health records (EHR) data for women between the ages of 18 and 64 who had encounters with eligible clinicians (ECs) during the measurement period (measurement period was 2012 for Site 1; 2013 for Sites 2 and 3). We also used claims data from one large multistate health plan. We used claims data for female patients with dual-energy X-ray absorptiometry (DXA) orders in 2013. We used the claims data as an initial way to estimate the percentage of women receiving potentially inappropriate DXA scans before contracting with sites to do in-depth validity and reliability testing, as well as to initially estimate the prevalence of exclusions.

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

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.20***)	**Measure Tested at Level of:**

☒ individual clinician	☒ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

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

Starting in August 2013, we recruited and selected three testing sites (see Table 1 for site details). Using specifications defined by the measure developer, Site 1 conducted analyses and provided output to measure developers. Sites 2 and 3 provided raw data and developers completed analyses. Consequently, some results from Site 1 are presented differently—for example, results were provided for women ages 50–64 but not further stratified results for women ages 50–59 and 60–64. Furthermore, Site 1 only provided data to support a subset of EHR data analyses (see 1.7).

We pursued testing sites that captured data elements for the measure in their existing EHR workflows. As a result, we recruited sites that could be considered advanced EHR users, suggesting that they are unlikely to be representative of the broader field of clinicians who treat the population of interest. Our approach thus offers evidence that the measure concept is achievable but does not provide conclusive evidence regarding the ability of all EHR users to implement these measures.

Table 1. Testing site characteristics

Characteristics	Testing site		
	Site 1	Site 2	Site 3
State	Maryland	New York	Michigan
Encounter dates in EHR data	2012	2013	2013
EHR system	Centricity	Epic	NextGen
Overall EHR experience	7 years	12 years	6 years
Practice type and specialty mix	Large multispecialty group	Large multispecialty group	Family practice and internal medicine
Number of sites	35	75	12
Participation in quality programs	PQRS	PQRS, PCMH, local initiatives including those related to Choosing Wisely and appropriate ordering of radiology procedures	PQRS, eRX, PCMH

PCMH = patient centered medical home; PQRS = physician quality reporting system; eRX = Electronic Prescribing Incentive Program; EHR = electronic health record

We also used claims data from one large multistate health plan to calculate the frequency of denominator exclusions and the percentage of potentially inappropriate DXA scans among women ages 50–64. The data included 7.5 million covered lives, 7.1 million of which were insured in commercial plans. The majority of the remaining lives were insured by Medicaid; the data included about 50,000 Medicare beneficiaries.

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*)

As described in 1.5, we recruited and subcontracted with three sites to collect patient-level EHR data for the measure. During initial testing, the measure included women ages 18–64. Based on results of testing and feedback from experts, we later restricted the measure to women ages 50–64. In 2014, sites provided data for women between the ages of 18 and 64 who had encounters with ECs during the calendar year measurement period (see Table 1). When possible, we report results for women ages 50–64, since that age range aligns with the current measure specification. Data from 87,242 patients were collected from Site 1, 102,593 patients from Site 2, and 25,899 from Site 3. Tables 2 and 3 summarize patients' distribution by age and race, respectively.

Table 2. Patients' age distribution

Age	All sites		Site 1		Site 2		Site 3	
	N	%	N	%	N	%	N	%
18–29	44,642	20.7	20,952	24.0	18,524	18.1	5,166	19.9
30–39	55,187	25.5	20,099	23.0	30,506	29.7	4,582	17.7
40–49	49,127	22.8	19,532	22.4	23,667	23.1	5,928	22.9
50–59	54,403	25.2	26,659[a]	30.6	20,473	20.0	7,271	28.1
60–64	12,375	5.7	–	–	9,423	9.2	2,952	11.4
Total	215,734	100.0	87,242	100.0	102,593	100.0	25,899	100.0

Source: Testing site EHR extracts sent to Mathematica.

Note: Due to rounding, some percentages on the total row do not sum to exactly 100 percent.

[a] Includes women ages 50–64.

Table 3. Patients' race distribution

Race	All sites		Site 1		Site 2		Site 3	
	N	%	N	%	N	%	N	%
White	114,329	53.0	49,346	56.6	45,583	44.4	19,400	74.9
Black	38,798	18.0	26,880	30.8	9,597	9.4	2,321	9.0
Asian	13,380	6.2	3,985	4.6	8,243	8	1,152	4.4
Other	20,781	9.6	4,763	5.5	15,870	15.5	148	0.6
Missing	28,446	13.2	2,268	2.6	23,300	22.7	2,878	11.1
Total	208,792	100.0	87,242	100.0	102,593	100.0	25,899	100.0

Source: Testing site EHR extracts sent to Mathematica.

Note: Due to rounding, some percentages on the total row do not sum to exactly 100 percent.

We also used claims data for 7.5 million women between the ages of 18 and 64 from one large health plan. Of these, 2,508,693 were between the ages of 50 and 64.

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.

We used EHR data from three sites to test the validity of the data elements and the frequency of denominator exclusions. We also tested ECs' performance score distribution and the measure's reliability at Sites 2 and 3. We used claims data from one large health plan to test the frequency of denominator exclusions, and the frequency of inappropriate DXA scans among women ages 50–64.

1.8 What were the social risk factors that were available and analyzed? For example, patient-reported data (such as income, education, and language), proxy variables when social risk data are not collected from each patient (for example, census tract), or patient community characteristics (percent vacant housing, crime rate), which do not have to be a proxy for patient-level data.

We did not test social risk factors because none were available in the EHR or claims data.

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*)

To estimate the measure's reliability, we used a random split-half correlation approach supplemented with bootstrapping. This approach included the following steps. First, we split patients randomly into two groups. For each group, we calculated the average performance rate per EC. With the EC as the unit of analysis, we then estimated the Pearson correlation coefficient to measure the strength of association between the two rates, using a resampling (bootstrapping) technique to increase the precision of the estimate. The resampling repeated these steps 2,500 times, with the average correlation calculated across iterations. We considered reliability coefficients of 0.70 and higher satisfactory (Nunnally and Bernstein 1994). For each tested measure, we also assessed the proportion of variance in EC performance scores attributable to EC performance, which we calculated by squaring the reliability estimate.

We limited our sample in the reliability analysis to primary care physicians (PCPs) who ordered DXA scans during the measurement period for female patients ages 50–64. Sites included a provider type code in the EHR data reports, which we used to identify PCPs.

We tested reliability across different denominator thresholds because prior work had shown reliability is dependent on the number of denominator cases (Scholle et al. 2008).

References

Nunnally, J.C., and I.H. Bernstein. *Psychometric Theory*. New York: McGraw-Hill, 1994.

Scholle, S.H., J. Roski, J.L. Adams, D.L. Dunn, E.A. Kerr, D.P. Dugan, and R.E. Jensen. "Benchmarking Physician Performance: Reliability of Individual and Composite Measures," *Am J Manag Care*, vol. 14, no. 12, Dec. 2008, pp. 833–838.

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 4 displays the results of the reliability analysis. The average reliability coefficient among PCPs with at least 20 patients ages 50–64 was 0.82.

Table 4. Reliability results

Denominator threshold	Number of PCPs	Number of patients	Average	Min	25th Percentile	Median	75th Percentile	Max
1 or more patients	269	19,162	0.25	0.02	0.13	0.24	0.34	0.60
10 or more patient	170	18,791	0.68	0.35	0.61	0.69	0.76	0.91
20 or more patients	138	18,370	0.82	0.64	0.80	0.83	0.85	0.91

Source: Rates were calculated using EHR data from Sites 2 and 3. Site 1 did not conduct a reliability analysis.

Note: Reliability analysis restricted to primary care physicians and patients between the ages of 50 and 64.

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 results indicate that for ECs with 20 or more patients in the denominator, the measure is reliable, with a median reliability of 0.83 and an interquartile range of 0.80–0.85. Measures with reliability coefficients of 0.70 are generally considered adequately reliable (Nunnally and Bernstein 1994). The lowest reliability estimate among the total group of 138 PCPs with at least 20 patients ages 50–64 in the denominator was 0.64 and the reliability estimate for the first percentile among this group was 0.72. These results suggest that the vast majority of PCPs were close to or above the reliability threshold of 0.70. The measure was reliable for about half of the PCPs in our sample, with 10 or more patients eligible for the denominator.

Reference

Nunnally, J. C., and I.H. Bernstein. *Psychometric Theory*. New York: McGraw-Hill 1994.

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).*

To ensure data used in calculating a measure score accurately reflected the care a patient received, such as whether she received a DXA scan or had risk factors for osteoporotic fracture, we assessed validity at the data element level. We drew a random sample of patients and extracted data elements from their EHR records (EHR extract), which we compared with data manually abstracted through a detailed, visual review of the patients' EHR (manual abstract). Using an a priori power analysis, we determined that each site needed to abstract a minimum of 200 charts per measure to achieve 80 percent power to detect statistically significant differences between manually abstracted and EHR extracted data. We manually abstracted data for 200 patients for each measure at Sites 2 and 3. Clinical staff at Site 1 were responsible for abstracting data for 216 patients. We assessed validity using kappa agreement statistics to estimate the chance-adjusted agreement between the two data sources for the sampled patients at each site. This approach allowed us to assess the validity of the EHR extract against a definitive record of the patients' care and to report overall agreement, sensitivity, and specificity. We then stratified validity results by site to obtain an understanding of how site characteristics (for example, documentation patterns) affected data element validity.

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

Chance-adjusted agreement between sites' EHR extracts and manually abstracted data for the numerator condition (DXA order) was 0.91, -0.01, and 0.93 at the three respective sites (Table 5). We also calculated chance-adjusted agreement for denominator exclusions; however, given the low prevalence of these data elements, these results are not reliable.

Table 5. Agreement between chart abstracted data and EHR extract

Site	Measure element	Data element	Chart prevalence	EHR prevalence	Kappa	Overall agreement	Sensitivity	Specificity
1	Numerator	DXA order	30.6%	33.3%	0.91	96.3%	98.5%	95.3%
1	Exclusion	Chronic malnutrition	0.0%	0.5%	N/A	99.5%	N/A	99.5%
1	Exclusion	Marfan syndrome	0.5%	0.0%	N/A	99.5%	0.0%	100.0%
1	Exclusion	Ehlers-Danlos syndrome	0.0%	0.0%	N/A	100.0%	N/A	100.0%
1	Exclusion	Osteopenia	13.9%	7.9%	0.60	92.1%	50.0%	98.9%
1	Exclusion	Osteoporosis	5.6%	0.0%	N/A	94.4%	0.0%	100.0%
1	Exclusion	Prior osteoporotic fracture	0.0%	0.0%	N/A	100.0%	.	100.0%
1	Exclusion	Ankylosing spondylitis	0.5%	0.0%	N/A	99.5%	0.0%	100.0%
1	Exclusion	Lupus	1.4%	0.5%	0.50	99.1%	33.3%	100.0%
1	Exclusion	Rheumatoid arthritis	0.5%	0.5%	1	100.0%	100.0%	100.0%
1	Exclusion	Type 1 diabetes	0.5%	0.5%	1	100.0%	100.0%	100.0%
1	Exclusion	Hyperthyroidism	0.5%	0.5%	1	100.0%	100.0%	100.0%
1	Exclusion	Hyperparathyroidism	–	–	–	–	–	–

Site	Measure element	Data element	Chart prevalence	EHR prevalence	Kappa	Overall agreement	Sensitivity	Specificity
1	Exclusion	Cushing's syndrome	0.0%	0.0%	N/A	100.0%	N/A	100.0%
1	Exclusion	Malabsorption syndrome	0.9%	0.5%	-0.01	98.6%	0.0%	99.5%
1	Exclusion	Chronic liver disease	1.4%	1.4%	0.32	98.1%	33.3%	99.1%
1	Exclusion	End-stage renal disease	0.0%	0.0%		100.0%	.	100.0%
1	Exclusion	Psoriatic arthritis	0.5%	0.5%	1	100.0%	100.0%	100.0%
1	Exclusion	Gastric bypass	–	–	–	–	–	–
1	Exclusion	Glucocorticoids	–	–	–	–	–	–
1	Exclusion	Risk of osteoporotic fracture[1]	–	–	–	–	–	–
1	Exclusion	Smoker	86.1%	83.3%	0.82	95.4%	95.7%	93.3%
2	Numerator	DXA order	48.5%	0.5%	-0.01	51.0%	0.0%	99.0%
2	Exclusion	Chronic malnutrition	0.0%	0.5%	N/A	99.5%	N/A	99.5%
2	Exclusion	Marfan syndrome	0.0%	0.0%	N/A	100.0%	N/A	100.0%
2	Exclusion	Ehlers-Danlos syndrome	0.0%	0.0%	N/A	100.0%	N/A	100.0%
2	Exclusion	Osteopenia	17.5%	1.5%	-0.03	81.0%	0.0%	98.2%
2	Exclusion	Osteoporosis	10.0%	1.5%	-0.03	88.5%	0.0%	98.3%
2	Exclusion	Prior osteoporotic fracture	3.5%	0.0%	N/A	96.5%	0.0%	100.0%
2	Exclusion	Ankylosing spondylitis	0.0%	0.0%	N/A	100.0%	.	100.0%
2	Exclusion	Lupus	1.0%	0.0%	N/A	99.0%	0.0%	100.0%
2	Exclusion	Rheumatoid arthritis	2.0%	0.0%	N/A	98.0%	0.0%	100.0%
2	Exclusion	Type 1 diabetes	0.0%	0.5%	N/A	99.5%	.	99.5%
2	Exclusion	Hyperthyroidism	2.0%	1.0%	-0.01	97.0%	0.0%	99.0%
2	Exclusion	Hyperparathyroidism	1.5%	0.0%	N/A	98.5%	0.0%	100.0%
2	Exclusion	Cushing's syndrome	0.0%	0.0%	N/A	100.0%	N/A	100.0%

Site	Measure element	Data element	Chart prevalence	EHR prevalence	Kappa	Overall agreement	Sensitivity	Specificity
2	Exclusion	Malabsorption syndrome	3.0%	2.0%	-0.02	95.0%	0.0%	97.9%
2	Exclusion	Chronic liver disease	5.0%	1.5%	-0.02	93.5%	0.0%	98.4%
2	Exclusion	End-stage renal disease	0.5%	1.5%	-0.01	98.0%	0.0%	98.5%
2	Exclusion	Psoriatic arthritis	0.5%	0.0%	N/A	99.5%	0.0%	100.0%
2	Exclusion	Gastric bypass	0.0%	0.0%	N/A	100.0%	N/A	100.0%
2	Exclusion	Glucocorticoids	2.5%	0.0%	N/A	97.5%	0.0%	100.0%
2	Exclusion	Risk of osteoporotic fracture[1]	0.5%	0.0%	N/A	99.5%	0.0%	100.0%
2	Exclusion	Smoker	9.0%	7.0%	0.12	87.0%	16.7%	94.0%
3	Numerator	DXA order	11.0%	12.5%	0.93	99%	100%	98%
3	Exclusion	Chronic malnutrition	0.5%	0.5%	1.00	100%	100%	100%
3	Exclusion	Marfan syndrome	0.0%	0.0%	N/A	100%	N/A	100%
3	Exclusion	Ehlers-Danlos syndrome	0.0%	0.0%	N/A	100%	N/A	100%
3	Exclusion	Osteopenia	12.5%	7.5%	0.5(	92%	48%	98%
3	Exclusion	Osteoporosis	3.5%	0.0%	N/A	97%	0%	100%
3	Exclusion	Prior osteoporotic fracture	0.0%	0.0%	N/A	100%	N/A	100%
3	Exclusion	Ankylosing spondylitis	0.0%	0.0%	N/A	100%	N/A	100%
3	Exclusion	Lupus	0.0%	0.0%	N/A	100%	N/A	100%
3	Exclusion	Rheumatoid arthritis	1.0%	1.5%	0.8(	100%	100%	99%
3	Exclusion	Type 1 diabetes	0.5%	0.5%	-0.0	99%	0%	99%
3	Exclusion	Hyperthyroidism	1.0%	0.0%	N/A	99%	0%	100%
3	Exclusion	Hyperparathyroidism	1.5%	0.5%	0.5(	99%	33%	100%
3	Exclusion	Cushing's syndrome	0.0%	0.0%	N/A	100%	N/A	100%
3	Exclusion	Malabsorption syndrome	3.0%	1.5%	0.6(	99%	50%	100%

Site	Measure element	Data element	Chart prevalence	EHR prevalence	Kappa	Overall agreement	Sensitivity	Specificity
3	Exclusion	Chronic liver disease	1.5%	2.0%	0.8	100%	100%	99%
3	Exclusion	End-stage renal disease	0.0%	0.0%	N/A	100%	N/A	100%
3	Exclusion	Psoriatic arthritis	0.0%	0.0%	N/A	100%	.	100%
3	Exclusion	Gastric bypass	0.5%	0.5%	1.0	100%	100%	100%
3	Exclusion	Glucocorticoids	0.0%	0.0%	N/A	100%	N/A	100%
3	Exclusion	Risk of osteoporotic fracture[1]	0.0%	0.0%	N/A	100%	N/A	100%
3	Exclusion	Smoker	9.0%	9.5%	0.9	100%	100%	99%

Source: Results comparing EHR extracted data and manually abstracted EHR chart data from three sites in 2013–2014. Results are based on 200 records from Sites 2 and 3 and 216 records from Site 1.

Dashes (–) in this table indicate elements that Site 1 was unable to calculate results based on their provision of incomplete information.

Note: The following data elements in the current specification are not included in the results: race, body mass index (BMI), alcohol consumption, osteogenesis imperfecta, aromatase inhibitors, documentation of hip fracture in parent.

[1] Measure specification used in testing did not include a data element focused on the FRAX® 10-year probability of osteoporotic fracture. Instead, it excluded women with a ten-year probability of osteoporotic fracture >=20 percent without specifying a risk assessment tool. After testing, we added a data element to the measure's specification which defined risk of osteoporotic fracture using a FRAX® 10-year probability of osteoporotic fracture >=9.3 percent which aligned with the USPSTF recommendation on osteoporosis screening. The measure specification being submitted to NQF uses a FRAX® 10-year probability of >=8.4 percent, in alignment with the June 2018 USPSTF recommendation on osteoporosis screening.

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?*)

In general, we considered kappa scores below 0.40 to be indicative of poor agreement, scores of 0.40 to 0.75 to be indicative of intermediate to good agreement, and scores above 0.75 to be indicative of excellent agreement (Fleiss 1981).

Chance-adjusted agreement between sites' EHR extracts and manually abstracted data for the numerator condition (DXA order) was excellent at two of the three sites that participated in our testing (Sites 1 and 3). Staff at these sites mentioned during site visits that their physicians routinely used structured fields to capture orders for DXA scans. Site 2 had agreement equal to chance for DXA orders, which was attributable to a lack of EHR documentation for DXA scans in structured fields (0.5 percent in the EHR extract versus 48.5 percent in the manual abstract). We also calculated agreement between denominator exclusion data elements. However, due to low prevalence of these data elements, kappa results are not reliable. The most prevalent denominator exclusion (current smoker status) had very good kappa agreement at Sites 2 and 3 (0.82 and 0.97, respectively).

We did not test the data element validity of the 10-year probability of osteoporotic fracture because it was unavailable at all test sites. However, the data element is derived from the FRAX, a validated tool.

Reference

Fleiss, J.L. *Statistical Methods for Rates and Proportions*. New York: John Wiley & Sons, Inc., 1981.

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 explored the EHR records and claims data for the prevalence of exclusions among all women in the denominator. Women were excluded from the measure if they had at least one of the following conditions and/or met any of these criteria: osteopenia, osteoporosis, chronic liver disease, malabsorption syndrome, hyperthyroidism, rheumatoid arthritis, type I diabetes, lupus, chronic malnutrition, prior osteoporotic fracture, use of glucocorticoids, hyperparathyroidism, psoriatic arthritis, end-stage renal disease, ankylosing spondylitis, recent gastric bypass surgery, Cushing's syndrome, Ehlers-Danlos syndrome, Marfan syndrome, osteogenesis imperfecta, low BMI (≤20 kg/m^2), current smoker, high alcohol consumption (more than 2 units per drinking day, where one unit is 12 oz. of beer, 4 oz. of wine, or 1 oz. of liquor), or 10-year risk of osteoporotic fracture ≥9.3 percent.[1] We did not measure the prevalence of risk of osteoporotic fracture because sites did not include the variable in structured fields of their EHRs, nor were the data available in claims. Overall, prevalence rates for most exclusions were typically under 5 percent, with the exceptions of osteoporosis and osteopenia. Claims data show the prevalence of exclusions in all women ages 18 to 64, because that was the population included in the measure at the time of the analysis. However, we limited EHR extracts to women who were ages 50 to 64, to conform with the measure specification. Several exclusions were not available in claims data, such as BMI, smoking status, and alcohol consumption.

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 6 displays the prevalence of excluded data elements in claims and EHR data.

Table 6. Prevalence of measure exclusions

Exclusion	Claims (women ages 18–64 with DXA order)	Site 1 (women ages 50–64)	Site 2 (women ages 50–64)	Site 3 (women ages 50–64)
Osteopenia	46.9%	57.2%	4.3%	11.0%
Osteoporosis	26.2%	#	2.7%	0.5%
Chronic liver disease	4.9%	15.0%	1.9%	1.6%
Malabsorption syndrome	3.9%	8.1%	2.6%	1.7%
Rheumatoid arthritis	3.3%	6.9%	1.0%	1.1%

[1] The measure specification we submitted to NQF uses a FRAX® 10-year probability of >=8.4 percent, in alignment with the June 2018 USPSTF recommendation on osteoporosis screening.

Exclusion	Claims (women ages 18–64 with DXA order)	Site 1 (women ages 50–64)	Site 2 (women ages 50–64)	Site 3 (women ages 50–64)
Hyperthyroidism	2.5%	3.3%	1.1%	#
Type I diabetes	1.5%	7.0%	0.6%	0.5%
Lupus	1.3%	3.4%	0.7%	#
Prior osteoporotic fracture	1.3%	–	#	#
Chronic malnutrition	0.5%	1.5%	#	#
Hyperparathyroidism	1.3%	–	0.7%	#
Glucocorticoids (oral only)	0.8%	–	–	–
Psoriatic arthritis	#	1.1%	#	#
End-stage renal disease	#	1.0%	1.1%	1.2%
Ankylosing spondylitis	#	#	#	#
Gastric bypass surgery	#	–	#	#
Cushing's Syndrome	#	#	#	#
Ehlers-Danlos syndrome	#	#	#	#
Marfan syndrome	#	#	#	#
Osteogenesis imperfecta	#	#	#	#
BMI <=20 kg/m^2	–	–	16.4%	9.8%
Current smoker	–	–	7.5%	12.0%
>2 units of alcohol per drinking day	–	–	5.5%	–

Source: Claims from one large health plan and testing site EHR extracts.

#: Prevalence was <0.5%

Dashes (–) in this table indicate data is not available. BMI, smoking status, and alcohol consumption are not generally available in claims data. Site 1 did not provide all data elements.

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*)

Using claims data, we examined the rate of denominator exclusions among women with DXA scans during measure testing. Many denominator exclusion conditions had extremely low prevalence (less than 2 percent). The three most prevalent conditions were osteopenia (47 percent), osteoporosis (26 percent), and chronic liver disease (4.9 percent). The prevalence of exclusions in EHR data varied across sites. Site 1 had a higher prevalence of risk factors than Sites 2 and 3, including osteopenia, chronic liver disease, malabsorption syndrome, rheumatoid arthritis, and lupus. Sites 2 and 3 had a fairly high prevalence of patients with low BMI

or current smokers, both risk factors that can exclude patients from the measure if they co-occur with other risk factors.

We did not test the exclusion of FRAX® risk of osteoporotic fracture >=9.3 percent[2] because this data element was added to the specification after our 2013–2014 testing. Discussions with testing sites during 2018 feasibility testing suggest that practices rarely capture the FRAX® score in structured fields of their EHR, so we would expect prevalence of this data element to be low as well. However, the measure specification excludes patients based on combination and independent risk factors that serve as inputs to the FRAX® tool. Therefore, although use of the FRAX® score can facilitate ECs' identification of patients to exclude from the measure, the specifications provide an alternative way to identify these patients.

Although many exclusions have low prevalence, they are based on evidence and add to the face validity of the measure. Therefore, we retained them in the measure. Some risk factors, such as osteogenesis imperfecta, are rare and likely to be infrequent in PCP data but could be much more prevalent for specialists who chose to report this measure. Furthermore, variation across sites' EHR data demonstrate that risk factors might be more prevalent in some settings than others. Therefore, the exclusions are important for ensuring that practices' performance scores are based on patients lacking risk factors for whom DXA scans might be unnecessary.

_____

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

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

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

☒ No risk adjustment or stratification

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

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

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

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

n.a. This measure does not use risk adjustment or stratification.

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

n.a. This is a process 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**

[2] The measure specification we submitted to NQF uses a FRAX® 10-year probability of >=8.4 percent, in alignment with the June 2018 USPSTF recommendation on osteoporosis screening.

☐ **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*)

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

If stratified, skip to 2b3.9

n.a.

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

n.a.

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

n.a.

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

n.a.

2b3.9. Results of Risk Stratification Analysis:

n.a.

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*)

n.a.

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*)

n.a.

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)*

Using EHR data from Sites 2 and 3, we computed and examined the distribution of performance scores for 269 PCPs (Table 7), which represented one type of EC that might choose to report this measure using EHR data. (Site 1 did not provide EC performance score distributions). Using claims data, we also calculated the

percentage of potentially inappropriate DXA scans among 2,508,693 women ages 50–64 who were insured by one large health plan.

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*)

Among 2,508,693 women ages 50–64 insured by one large health plan, 6.7 percent had potentially inappropriate DXA scans as defined by the measure.

Table 7 displays the performance score distribution calculated from EHR data from two sites.

Table 7. Performance distribution

EP type	Number of EPs	50th percentile	75th percentile	90th percentile	95th percentile	99th percentile
PCPs (patients ages 50–64)	269	0.0%	0.5%	10.0%	19.2%	33.3%
PCPs (10 or more patients ages 50–64)	170	0.0%	1.6%	6.1%	14.3%	22.2%
PCPs (20 or more patients age 50–64)	140	0.0%	2.2%	6.4%	11.3%	21.0%

Note: Rates were calculated using EHR extracts from Sites 2 and 3. Lower scores indicate higher quality.

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 claims results demonstrate that there is an opportunity to reduce the number of women ages 50–64 receiving inappropriate DXA scans, with nearly 7 percent of women receiving a potentially inappropriate DXA scan.

The distribution from EHR data was skewed to the left (median performance was 0.0 percent), suggesting many PCPs were not ordering potentially inappropriate DXA scans for their patients. Among PCPs with 20 or more patients ages 50–64, PCPs in the highest decile of the distribution (that is, the poorest decile of physician performance) had performance scores between about 6 and 21 percent, similar to results for PCPs with 10 or more patients in the denominator.

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

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

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

more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.

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

n.a. This measure uses one set of specifications.

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*)

We did not explicitly test the distribution of missing data for this measure; however, our data file construction and data element validity results inform our understanding of the potential for systematic bias due to missing data. When comparing the data extracted from the EHR with a manual review of the full medical record, data element validity testing can be used to inform the level of missingness for individual data elements. Missing data will result in low overall agreement and chance-adjusted agreement (kappa).

The testing was limited to patients ages 50 to 64 eligible for the measure's denominator. In the data files submitted by test sites, there was no distinction between a negative (for example, confirmation that the patient was diagnosed with osteoporosis) and missing data. Where sites reported data for at least one patient, we assumed that blank records indicated no relevant data for those patients. For example, we assumed a patient with no data indicating osteoporosis did not have osteoporosis; we did not exclude that patient from the denominator based on lack of data regarding osteoporosis.

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)

As described above, data element validity testing, when comparing the data extracted from the EHR with a complete chart review for a sample of patients, can be used to inform the level of missingness for individual data elements. Missing data will result in low overall and chance-adjusted agreement (kappa). Lack of missing data will result in high overall and chance-adjusted agreement.

As shown in Table 5, agreement between the EHR extract and an abstract of the full chart showed high agreement and a lack of systematic missing information. Of the 22 data elements tested, all had overall agreement rates greater than 90 percent at two or more sites. Only four data elements scored less than 90 percent at Site 2, and three of these data elements scored above 80 percent. Due to low prevalence of many excluded data elements, we could not calculate kappa at all sites for all data elements. However, for DXA orders, the data element necessary to calculate the numerator, kappa was >.95 at two of the three sites. Smoking status, one of the most prevalent exclusions, had kappa >0.8 at two of the three sites. Osteoporosis

and osteopenia, two additional exclusion data elements that were fairly prevalent, had kappa >0.55 at two of the three sites. Variation across testing sites indicates that missingness for these data elements was not systematic.

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*)

As noted above, we interpreted data missing from numerator fields as indicating that the patient did not receive the service, and data missing from denominator exclusion fields as indicating that the patient should be included in the measure.

Based on our analysis of data element validity showing site-level variation in kappa and overall agreement rates for the four data elements with lowest EHR/chart agreement (overall agreement rates < 90 percent and kappa <55 percent for prevalent data elements), we conclude that missing data for these data elements is not systematic.

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

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

If other:

3b. Electronic Sources

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

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

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

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

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

Attachment: DXA_Feasibility_Scorecard_-1-.xlsx,DXA_Feasibility_Narrative_Final.docx

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.

This is a new measure that has not yet been implemented. Attached to this submission are two documents—a feasibility summary and scorecard—that describe the difficulties regarding data collection.

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

The FRAX may be accessed online for free. Clinicians can also purchase a desktop version if desired. To our knowledge, there are no fees, licensing, or other requirements associated with using any other aspect of the measure as specified, such as the value or code sets, programming code, or algorithm. The measure is available for public use.

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
Public Reporting	Payment Program CMS Merit-based Incentive Payment System (MIPS) https://qpp.cms.gov/mips/quality-measures

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

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

In the final CY2019 Medicare Physician Fee Schedule rule posted on November 1, 2018, CMS added this measure to MIPS beginning with performance period 2019. MIPS streamlines three historical Medicare

programs – the Physician Quality Reporting System, the Value-based Payment Modifier Program, and the Medicare Electronic Health Record Incentive Program – into a single payment program as part of CMS efforts to move clinicians to a performance-based payment system. MIPS is a national program where eligible clinicians can choose to report quality measures most meaningful to their practice. Clinicians will have the option to report this measure in 2019.

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

There are no reasons, such as policies or accessibility, which prohibit the use of this measure. CMS has adopted this measure for use in its MIPS program for performance period 2019 and future years. More information can be found in Section 4a1.3.

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

CMS submitted the measure to the Measures Under Consideration list for MIPS in June 2017. The Measure Applications Partnership reviewed the measure in December 2017 and recommended the measure for inclusion in the program with conditional support (pending NQF endorsement). CMS adopted this measure for use in its MIPS program for performance period 2019 and future years.

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

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

This is a new measure and has not yet been implemented, so we have not shared performance results with the entities being measured. However, as part of measure development and testing, we computed measure performance for two physician practices. We shared their performance data with various organizations or individuals, as described below.

With the two physician practices (test sites), we shared their individual performance rates but did not share the performance rates of the other test sites. The clinicians at these sites are the types of eligible clinicians who may, in the future, report on this measure as part of the MIPS program.

We shared performance data from both test sites with a technical expert panel (TEP). The TEP consisted of health system representatives, EHR vendors, patients, consumer representatives, and clinicians. It included clinicians who may, in the future, report on this measure as part of the MIPS program, along with other experts who would not report on this measure (for example, EHR vendors who do not work in a clinician practice).

We also shared performance data from both test sites with a DXA Overuse expert work group (EWG). The EWG consisted of experts in osteoporosis, skeletal health, and overuse measurement. It included clinicians who may, in the future, report on this measure as part of the MIPS program, along with other experts who would not report on this measure (for example, measure development experts who do not work in a clinician practice).

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

This is a new measure that CMS has not yet implemented, so we do not have national performance results to share. However, during measure development and testing, we computed measure performance for two clinician practices and shared the results with the test sites, the TEP, and the EWG. With each test site, we shared only the overall measure performance for that practice. With the TEP and EWG, we shared de-identified overall measure performance across the two test sites. We shared these data once with each group. During the meetings in which we shared the data, we also reviewed the measure specifications.

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.

During measure testing, we gave the test sites an opportunity to discuss any questions or concerns they had about their measure performance.

During our meetings with the TEP and EWG, we gave the members an opportunity to discuss any questions or concerns they had about the shared performance information.

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

The two test sites did not share any significant concerns about their performance on the measure.

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

The TEP and EWG did not share any significant concerns about clinician performance on the measure.

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.

The feedback described above did not result in changes to the measure specifications.

Improvement

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

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

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

The intent of this measure is to decrease the use of DXA scans among people who are at low risk for osteoporotic fracture, thereby reducing DXA-related harms. Although the measure is not yet in use, we expect that its implementation will improve quality of care by helping clinicians track their performance and by motivating them to reduce the number of inappropriate DXA scans they order.

4b2. Unintended Consequences

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

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

This is a new measure that CMS has not yet implemented in a program. When CMS implements the measure, it could cause women ages 50 to 64 with osteoporosis who do not have the risk factors identified in the measure—or who have the risk factors but not the number specified—to miss needed DXA screenings. Also, the applicability of the FRAX to nonwhite subgroups has not yet been widely studied (Viswanathan et al., 2018). Nonwhite women and women with risk factors other than those identified by the measure could fail to begin or experience unnecessary delays in appropriate treatment for osteoporosis.

Citation:

Viswanathan M, Reddy S, Berkman N, Cullen K, Middleton J, Nicholson W, et al. Screening to prevent osteoporotic fractures: updated evidence report and systematic review for the U.S. Preventive Services Task Force." JAMA. 2018;319(24):2532-51.

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

This measure does not explicitly assess clinician use of clinical risk assessment tools to determine patient risk for osteoporotic fracture (as recommended by the U.S. Preventive Services Task Force). However, it will encourage the use of those tools—particularly the FRAX—because clinicians will notice its inclusion in the measure as a method for identifying patients at high risk for fracture; clinicians may decide that this tool is an efficient way to screen patients before ordering a DXA scan. The measure could also increase clinicians' consistency in determining which patients are at high risk for osteoporotic fracture—and therefore eligible for a DXA scan.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

Yes

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

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

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.

(NQF 0046) Screening or Therapy for Osteoporosis for Women Aged 65 Years and Older: Percentage of female patients aged 65-85 years of age who ever had a central dual-energy X-ray absorptiometry (DXA) to check for osteoporosis. NQF 0046 is in MIPS and is specified for claims and registry reporting. It complements the proposed measure because it assesses the percentage of women who receive an appropriate osteoporosis screening after age 65. There are some differences between the measures, but these are appropriate based on the measures' intents. NQF 0046 assesses for documentation of DXA results, whereas the proposed measure assesses for DXA orders. Assessing for DXA orders makes sense because the proposed measure focuses on overuse of DXA screening. Also, NQF 0046 is limited to DXA scans of the hip or spine (that is, central DXA scans), whereas the proposed measure assesses for central and peripheral DXA scans. In its 2011 recommendation, the U.S. Preventive Services Task Force recommended using central DXA scans to assess for osteoporosis—and NQF 0046 complies with this recommendation. But the proposed measure, as an overuse measure, assesses for any type of DXA scan because any type could be inappropriate. Together, these two

measures assess the appropriate use of DXA scans in women 65 and older, along with inappropriate use of DXA scans in women under age 65.

5b. Competing Measures

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

OR

Multiple measures are justified.

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

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

Not applicable. We did not identify any competing measures.

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): Centers for Medicare & Medicaid Services, Center for Clinical Standards and Quality, Quality Measurement and Value-Based Incentives Group (QMVIG), Division of Electronic and Clinician Quality, MS S3-02-01

Co.2 Point of Contact: Susan, Arday, B.S.P.H., M.H.S., C.H.E.S., Susan.Arday@cms.hhs.gov, 410-786-3141-

Co.3 Measure Developer if different from Measure Steward: NCQA

Co.4 Point of Contact: Jenna, Williams-Bader, bader@ncqa.org, 202-955-5103-

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.

The following individuals participated in the DXA Overuse EWG. We selected EWG members based on their expertise in osteoporosis, skeletal health, and overuse measurement. They provided feedback throughout the measure's development, from 2013 to 2014—commenting on the clinical components of the measure, including the denominator, numerator, and exclusions, and on the measure's importance, feasibility, validity, and usability.

Itara Barnes, Medical University of South Carolina

Meryl S. LeBoff, M.D., Brigham and Women's Hospital

Michael LeFevre, M.D., M.S.P.H., University of Missouri

Mark Robbins, M.D., Harvard Vanguard

Kenneth Saag, M.D., M.Sc., University of Alabama at Birmingham

The following individuals participated in the TEP. This multistakeholder group had representatives from health systems, clinician practices, EHR vendors, and consumer advocacy organizations. The TEP provided feedback throughout the measure's development, from 2013 to 2014, on the importance, feasibility, validity, and usability of the measure.

Ayodola Anise, M.H.S., senior research associate, Engelberg Center for Health Care Reform, The Brookings Institute

Jessica Bartell, M.D., M.S., clinical informatics physician, Epic

Nate Bennett, M.D., physician, Preferred Primary Care Physicians

Jason Colquitt, executive director, research services, Greenway Medical Technologies, Inc.

William F. Groneman, M.H.A., executive vice president, system development, TriHealth, Inc.

Erin A. Mackay, M.P.H., associate director, health information technology systems, National Partnership for Women & Families

Jon D. Morrow, M.D., M.B.A., M.A., F.A.C.O.G, executive vice president, system development, General Electric Healthcare

Daniel Todd Rosenthal, M.D., M.Sc., M.P.H., director of health care intelligence, Inova Health Systems

Shannon Sims, M.D., Ph.D., director of clinical informatics and medical director of information services, Rush University Medical Center

Samuel S. Spicer, M.D., M.M.M., vice president of medical affairs, New Hanover Regional Medical Center

Rachelle "Shelly" Spiro, R.Ph., F.A.S.C.P., director, Pharmacy e-Health Information Technology Collaborative

Andy Steele, M.D., M.P.H, M.Sc., director of medical informatics, Denver Health

Jonathan P. Weiner, Dr.P.H., M.S., professor and program director, Johns Hopkins Bloomberg School of Public Health

Thomas R. Williams, M.P.H., M.B.A., Dr.P.H., executive director, Integrated Healthcare Association

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2018

Ad.3 Month and Year of most recent revision:

Ad.4 What is your frequency for review/update of this measure? CMS conducts an annual review to determine potential updates to the measure.

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

Ad.7 Disclaimers: The 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 MEASURE AND SPECIFICATIONS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND.

Due to technical limitations, registered trademarks are indicated by (R) or [R] and unregistered trademarks are indicated by (TM) or [TM].

Ad.8 Additional Information/Comments: Not applicable.