

## QMC Measure Testing Protocol Draft – April 2019

### Background

Testing quality measures for implementation in quality improvement or accountability programs is a key component in assuring measures perform in the ways they are intended. Tested measures support the achievement of healthcare that is safe, effective, patient-centered, timely, efficient and equitable. Over the past 15-plus years, a wide range of organizations have contributed to assuring the scientific validity and clinical importance of quality measures. These include, for example, the National Quality Forum (NQF), the Physician Consortium for Performance Improvement (PCPI), medical associations, specialty societies, the Centers for Medicare & Medicaid Services (CMS) and private health plans.

### Purpose

Building upon the Quality Measures Committee's (QMCs) measure prioritization and development protocol, the QMC measure testing protocol aligns closely with NQF's measure evaluation criteria: (1) whether the measure focus is evidence-based, addresses a documented performance gap, and is of high-priority (Impact); (2) the extent to which the measure produces consistent (reliable) and credible (valid) results about the quality of care when implemented (Scientific Acceptability); (3) the extent to which the measure specification includes data elements that are readily available so as to reduce reporting burden and ease implementation (Feasibility); and (4) the extent to which the measure can be used for both accountability and performance improvement by potential audiences (e.g., consumers, providers, policymakers, purchasers) to achieve the goal of high-quality care (Usability and use).

After QMC workgroups/subject matter experts (SMEs), in collaboration with other relevant AGA Institute Committees, develop and specify a *de novo* measure, the QMC oversees initial feasibility and face validity testing of new measures through recruiting member volunteers. For measures identified as both feasible and meeting face validity requirements, a testing vendor is selected to perform more robust and complete measure testing work. Based on findings from measure testing, the QMC will use testing information to determine if a measure should be recommended for use in a national program, by a private payor, for inclusion in a quality improvement initiative or in other implementation mechanism such as a Qualified Clinical Data Registry (QCDR) or Qualified Registry (QR). Based on its review of measure testing data, the QMC may also recommend measure discontinuation, placement of a measure in inactive status, or retirement of a measure. This protocol describes the elements of the measure testing process, the QMC's strategy for recommending measures for testing, and the use of measure testing results following completion of the measure testing process.

### Measure Testing Process

The *2010 Measure Testing Protocol for Physician Consortium for Performance Improvement (PCPI™) Performance Measures* recommends the following testing requirements:

**Reliability Testing:** Can the measure be consistently calculated using a variety of sources? Does it measure this aspect of care well?

- Manual Review – Two reviewers audit the same sample of charts, calculate agreement rates, evaluate the reliability of the numerator, denominator, exclusions, and calculation/algorithm

- Automated Reporting vs. Manual Review – compare automated calculation of the measure to manual review; evaluate the reliability of the numerator, denominator, exclusions, CPT Category II codes, and calculation/algorithm.
- Administrative data (including CPT II): compare automated calculation of the measure to manual review; evaluate the reliability of the numerator, denominator, exclusions, and calculation/algorithm

**Validity Testing:** Does it measure what it is intended to measure? Does high performance lead to better outcomes? Are there any unintended consequences or adverse effects from measuring this aspect of care?

- Measure specifications developed and measures in use.
- Statistical analysis performed to determine whether higher measure performance correlates with better outcomes
- Prospective testing and retrospective monitoring of unintended consequences

### **Reliability Testing**

*Inter-rater reliability testing* can be conducted by either manual chart abstraction or EHR-based chart abstraction where a minimum of two different individuals review the same charts to determine if the same results are attained. The tool(s) that are used should be constructed in a way to make observations as uniform as possible.

### **Validity Testing**

*Face validity* attempts to determine if the measure definition and specifications measure the aspect of care that was intended.

*Construct validity* refers the extent to which the concept is designed to measure using either convergent or discriminant validity associations. Convergent validity is the degree to which multiple indicators of a single underlying concept are correlated. Essentially, is what we intend to measure actually being measured? Discriminant (or divergent) validity is the degree to which there are no cross-associations (i.e., whether concepts or measurements that are not supposed to be related are actually not related). A result > 0.85 tells us that two concepts overlap greatly and are likely measuring the same thing (confirming construct validity). If the result is < 0.85, we can conclude that discriminate validity exists.

### **Alpha Testing for Reliability and Validity Testing**

Alpha testing refers to pilot testing of a measure to inform the measure development process and refine the specifications prior to measure finalization. Alpha testing may be carried out several times in the measure development process. In contrast, beta testing refers to the final testing conducted on a measure when all necessary revisions have been made stemming from alpha testing. There are two types of alpha tests that can be used -- either questionnaire-based or focus group-based. Alpha testing questions can be sent to a participating site or practice or asked in a focus group:

#### **Inter-rater Reliability**

- Is the diagnosis entered in a specific field in a routine way?
- Are ICD-10 or CPT codes entered for all procedures and diagnoses included in the measure? If not, is another coding system used?
- Is it feasible to develop and run a report from the EHR showing all patients within a certain timeframe who were diagnosed?

- Does this report already exist at a specific practice or institution?
- Is it possible to display on the report whether they met the Measure and how?

#### **Alpha testing for Face Validity**

- Is there sufficient evidence to support the widespread use of this measure?
- Does an opportunity for improvement exist; does the measure address an area(s) where there is a substantial gap between optimal and current clinical practice?
- Are the measure rationale and results easily understood by and meaningful to users of the data?
- How often would information derived from this measure be used in everyday practice?

### **QMC Measure Selection Process for Testing**

The Quality Measures Committee will review all new and revised AGA measures on an annual basis; at a minimum, the QMC will rank the measures from highest to lowest priority for their readiness for measure testing, and select the highest-priority measures for testing. The number of measures recommended for testing will inform the annual budget process and therefore this selection process should occur no later than October of a given year. On average, the cost of testing a single measure is approximately \$15,000 depending on the complexity of the measure. Vendors with experience testing measures include PCPI, Telligen, Mathematica, and Minnesota Community Measurement. Upon recommending measures for testing, the QMC Chair and workgroup leads will author a report outlining the rationale for selecting measures to be submitted to the AGA Institute Governing Board for review and approval.

The report to the Governing Board will include the results of the Phase II ranking exercise(s) and a measure prioritization brief for each measure proposed for testing. The measure prioritization briefs will summarize each measure's Meaningfulness, Feasibility of Implementation, Magnitude of Effect, Quality Gap, and Applicability. If a measure is not recommended for testing during the measure review cycle, it may be considered for submission in future years during the AGA Institute budgetary process.

### **How Will Measure Testing Results Be Used?**

Upon receipt of the measure testing results from the vendor, the QMC will review the recommendations to determine the implementation strategy for the measure, including whether the measure should be recommended for inclusion in a national program (NQF, CMS), for private payor use, for use in a local quality improvement program, or if the measure should be retired. The workgroup assigned a measure for development will author a white paper to be submitted to one of the AGA journals for publication describing the process for developing the measure, testing results, and implementation strategy for the measure.

#### ***Use in a National Program***

To recommend that a measure be used in a national program such as the CMS Quality Payment Program (QPP) or Ambulatory Surgical Center Quality Reporting (ASCQR) Program or for endorsement consideration by NQF, the QMC will determine the specific recommended performance category. These categories include:

1. Clinical Care
2. Safety
3. Care Coordination
4. Patient and Caregiver Experience
5. Population Health and Prevention

## 6. Affordable Care

Prior to recommending a measure for submission to the Measure Applications Partnership (MAP), the QMC will conduct an analysis of the annual MAP priorities to ensure the greatest opportunity for inclusion in the QPP. <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/Pre-Rulemaking.html>

The following 2018 priorities were noted for measure implementation consideration in federal programs:

- Burden – Is there a balance between the value of the measure and the burden of collecting the measure?
- Preference for outcome measures, process measures that are proximal to important outcomes and have a solid evidence base, and composite measures that provide a comprehensive view of performance.
- Defining value from the patient perspective – measures that address the issues that patients and consumers find most important.

To meet current CMS requirements, recommendations for submission of a new measure into the Quality Payment Program must be made no later than January 1.

### ***Private Payor Use***

The Quality Measures Committee will consider recommending a measure for use by private payors if the measure is widely accepted as the standard of care for a particular medical condition and if it fully aligns with guideline recommendations. Measures recommended for use by private payors may or may not be included in the QPP or hold NQF endorsement. The QMC will recommend whether a measure is appropriate for private payor use and may suggest a measure to the Core Quality Measures Collaborative (CQMC) for inclusion in the Gastroenterology Core Measure Set or HIV/Hepatitis C Core Measure Set.

AGA will conduct annual interviews with interested payors to determine payors' measure priorities and if measures under development by the QMC align with those priorities. If measures align with payor priorities, the QMC will recommend measures for potential inclusion in payors' quality provider tiering processes and programs.

### ***Use in a Quality Improvement Program***

Measures recommended only for use in local quality improvement programs are measures that either (a) do not fully meet CMS measure priorities for submission to the Measure Applications Partnership (MAP) for inclusion in the QPP; and/or (b) are not accepted by the CQMC or other private payors as core measure(s) for a specific medical condition that can be used in reliably identifying high-quality providers for purposes of payor tiering programs. These measures may nevertheless be used locally by practices, health systems, academic medical centers, or ambulatory surgical centers in their internal quality improvement programs.

### **Recommendation of Measures for Inactive Status**

A quality measure will be recommended for inactive status if testing data indicates that the measure is not valid or reliable and the measure cannot be adequately revised to enhance its validity and reliability. Other factors that will be considered when recommending a measure for potential retirement include:

- Newly-published evidence that changes the intent of the measure and renders the measure irrelevant.
- Evidence that there is no longer a performance gap/opportunity for improvement with continued use of the measure (i.e., the measure is “topped out”).
- Removal of endorsement by NQF or from the Quality Payment Program AND the measure is not recommended for private payor use or use in a local quality improvement program, as above.