Partnership for Quality Measurement
Powered by Battelle

Endorsement and Maintenance (E&M) Guidebook

Draft for Public Comment

June 2024

Prepared by:

Battelle
505 King Avenue, Columbus, Ohio 43201
Revised April 2024

The analyses upon which this publication is based were performed under Contract Number 75FCMC23C0010, entitled, "National Consensus Development and Strategic Planning for Health Care Quality Measurement," sponsored by the Department of Health and Human Services, Centers for Medicare & Medicaid Services.

Table of Contents

List of Tables

Page

List of Figures

Page

Introduction

The *Endorsement & Maintenance Guidebook (E&M Guidebook)* is a resource for measure stewards, measure developers, and organizations submitting measures to Battelle for the Partnership for Quality Measurement (PQM) for endorsement review. This guidebook provides information about the various steps of the endorsement and maintenance (E&M) process, including each phased review, possible endorsement decision outcomes, the appeals process, E&M policies and procedures, and the E&M committee structure.

The E&M Guidebook is organized to provide an overview of E&M goals, priorities, and resources; to guide measure developers and stewards through the six steps of the E&M process; and to provide key considerations for submitting measures to Battelle. The E&M Guidebook aims to do the following:

- Explain the measure submission and evaluation processes.
- Describe the expectations for measure developers and stewards as participants in the process.
- Serve as the main resource for E&M-related processes and policies.

The E&M Guidebook will be updated on a timely basis to maintain a current reference to assist measure developers and stewards in navigating the E&M process.

Who We Are

Battelle Memorial Institute (Battelle) is the world's largest independent, nonprofit, applied science and technology organization, with the objective of using science for the benefit of mankind. As a 501(c)(3) charitable trust, we are committed to translating scientific discovery and technology advances into societal benefits.

For over 20 years, we have been a leader in the science of health care quality measurement and improvement. Battelle is highly experienced in independent systematic evidence-based reviews of clinical quality measures (CQMs) and cost/resource measures.

Battelle is a certified consensus-based entity (CBE) under the Centers for Medicare & Medicaid Services' (CMS) Qualified Entity (QE) Program developed to implement Section 10332 of the Affordable Care Act and the "Medicare Program; Availability of Medicare Data for Performance Measurement" Final Rule [CMS-5061-F]. As a certified CBE, Battelle meets the criteria of an independent CBE as mandated in federal statutes (SSA Section 1890 and 1890A).

To facilitate the execution of CBE tasks, we have formed PQM, which is comprised of all interested parties (formerly referred to as multi-stakeholder groups), including but not limited to health care providers (e.g., clinicians, health plans, health systems), patients and caregivers, measure experts (e.g., developers, stewards, researchers), and health information technology specialists. Battelle's transparent, streamlined approach to consensus-building facilitates informed and thoughtful endorsement reviews of quality measures. Membership in PQM is free, and individuals must be members of PQM to serve on an E&M committee.

E&M Guidebook Updates

We are dedicated to the continued evaluation of the E&M process to address the demands of the changing health care landscape and interested party feedback. Therefore, the process and the E&M Guidebook will evolve over time. We will update the E&M Guidebook on a timely basis to maintain a current reference to assist measure developers, stewards, and E&M committee members in navigating the E&M process.

Any major changes to the E&M process, policies, or evaluation criteria undergo a formal public comment period before we implement them. We also provide additional educational resources (e.g., webinars, informational guides), as needed, on our PQM website. We will not apply any major changes in the E&M process or measure evaluation criteria to any measure that is currently going through the E&M process.

This version of the E&M Guidebook has been updated and is currently open for public comment from June 4 – June 24, 2024. All policies and procedures herein will be applied beginning with the Fall 2024 cycle. The changes to the E&M Guidebook and process include:

- Enhancements to E&M committee engagement, size, convenings, and voting (*see E&M Committee Composition, Roles, and Responsibilities, Advisory Group Meetings, and Endorsement Meeting*).
- Updates to measure public comment opportunities (*see Public Comment*).
- Added conditions for measures to be considered for endorsement (*see Requirements for Measure Consideration*).
- Updates to the Intent to Submit and Full Measure Submission forms (*see Submission Tool and Repository*).
- Changes to the PQM Measure Evaluation Rubric (*see Appendix D*).
- Updates to the conditions for the "Endorsed with Conditions" designation (see *Table 4 and Table 5 of the Overview of and Enhancements to the Endorsement Process*).
- Updates to the appeals eligibility criteria (*see Appeals Eligibility Criteria*).
- Added clarifications to Endorsement Maintenance (*see Endorsement Maintenance*).

Battelle's Portfolio of CBE Measures

Battelle organizes measures for E&M by five project topical areas. Each project topical area has an evaluation committee that oversees the portfolio of measures for the topic (Table 1). A project consists of measures submitted by measure developers/stewards and grouped by similar topic.

Table 1. Project Topical Areas

Project Title	Areas Covered	Example Measures
Primary Prevention	Education, prevention, and screening related to health status and/or health risk.	• CBE #0028 Preventive Care and Screening: Tobacco Use: Screening and Cessation Intervention • CBE #2372 Breast Cancer Screening • CBE #3620 Adult Immunization Status

Project Title	Areas Covered	Example Measures
Initial Recognition and Management	Recognition and timely diagnosis of conditions, including diagnostic accuracy, monitoring of early signs and symptoms of disease/condition.	• CBE #0058 Avoidance of Antibiotic Treatment in Adults With Acute Bronchitis (AAB) • CBE #3671 Inappropriate diagnosis of community-acquired pneumonia (CAP) in hospitalized medical patients • CBE #3663e Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Clinician Level)
Management of Acute Events, Chronic Disease, Surgery, Behavioral Health	Treatment of acute events, management of chronic disease, including structural or functional changes related to chronic disease, surgery, and related outcomes.	• CBE #0711 Depression Remission at Six Months • CBE #0729 Optimal Diabetes Care • CBE #3025 Ambulatory Breast Procedure Surgical Site Infection (SSI) Outcome Measure
Advanced Illness and Post-Acute Care	Advanced illness and/or end-stage disease management, palliative and hospice care, post-acute care, and home care.	• CBE #0384e Oncology: Medical and Radiation - Pain Intensity Quantified • CBE #2651 Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Hospice Survey • CBE #2978 Hemodialysis Vascular Access: Long-term Catheter Rate
Cost and Efficiency	The amount or frequency of health services applied to a population or event (e.g., procedures, encounters).	• CBE #2158 Medicare Spending Per Beneficiary (MSPB) – Hospital • CBE #3575 Total Per Capita Cost (TPCC) • CBE #2687 Hospital Visits after Hospital Outpatient Surgery

Current and future endorsed measures are assigned to one of the five new projects based on where the measure has the most relevance in the patient's journey through health care. Additional information for the five projects can be found on the PQM website.

Submission Tool and Repository

Key information about measure submissions, including endorsement status, is available via the Submission Tool and Repository (STAR). STAR is an online platform where developers/stewards can submit measures, and any interested party may view measure information, including the endorsement status, in the searchable repository database. The database is updated regularly as new and maintenance measures are submitted to Battelle for PQM endorsement review.

The measure submission function is available as of the Fall 2023 cycle. Measure developers/stewards must first create an account by going to the "Submit a Measure" page on the PQM website. Each cycle has a designated Intent to Submit deadline when measure developers/stewards must submit key information (e.g., measure title, type, description, specifications) about the measure (*see Intent to Submit for more details*). One month after the Intent to Submit deadline is the Full Measure Submission deadline (Table 2). During this time, measure developers/stewards submit the full measure information (*see Full Measure Submission for more details*).

Table 2. Intent to Submit and Full Measure Submission Deadlines by Cycle

E&M Cycle	Intent to Submit*	Full Measure Submission*
Fall	October 1	November 1
Spring	April 1	May 1

**Deadlines are set at 11:59 p.m. (ET) of the day indicated. If the deadline ends on a weekend or holiday, the deadline will be the next immediate business day.*

To review the measure submission items and questions within each form, Microsoft Word templates of the Intent to Submit and Full Measure Submission forms are available on the PQM website and linked below:

- Intent to Submit Form Template (Version 2.0 | 2024)
- Full Measure Submission Form Template (Version 2.0 | 2024)

Endorsement and Review Process

Overview of and Enhancements to the Endorsement Process

The E&M process ensures measures submitted for endorsement are evidence based, scientifically sound, and both safe and effective, meaning use of the measure will increase the likelihood of desired health outcomes; will not increase the likelihood of unintended, adverse health outcomes; and is consistent with current professional knowledge.

Our novel E&M process builds from the prior CBE processes and enables E&M decision-making in as few as 6 months (from the Intent to Submit deadline until the end of the project [i.e., through the end of appeal proceedings]). Under this E&M process, measures reach their endpoint when an endorsement decision is rendered. This occurs when the E&M committees reach a final endorsement decision (Table 3).

Table 3. Endorsement Decision Outcomes

Decision Outcome	Description	Maintenance Expectations
Endorsed	**Applies to new and maintenance measures.** The E&M committee agrees by 75% or more to endorse the measure.	Measures undergo maintenance of endorsement reviews every 5 years with a status report review at 3 years (*see Evaluations for Maintenance Endorsement for more details*).± Developers/stewards may request an extension of up to 1 year (two consecutive cycles), except if it has been more than 6 years since the measure's date of last endorsement.
Endorsed with Conditions*	**Applies to new and maintenance measures**. The E&M committee agrees by 75% or greater that the measure can be endorsed as it meets the criteria, but committee reviewers have conditions they would like addressed when the measure comes back for maintenance. If these recommendations are not addressed, the developer/steward should provide a rationale for consideration by the E&M committee review.	Measures undergo maintenance of endorsement reviews every 5 years with a status report at 3 years, unless the condition requires the measure to be reviewed earlier (*see Evaluations for Maintenance Endorsement for more details*). The E&M committee evaluates whether conditions have been met, in addition to

Commented [HSAG1]: Please consider adding a flag in the STAR database for measures that receive an extension, along with the due/end date of the extension.

Decision Outcome	Description	Maintenance Expectations
		all other maintenance endorsement minimum requirements.
Not Endorsed°	**Applies to new measures only**. The E&M committee agrees by 75% or greater to not endorse the measure.	None.
Endorsement Removed°	**Applies to maintenance measures only.** Either: • The E&M committee agrees by 75% or greater to remove endorsement; or • A measure steward retires a measure (i.e., no longer pursues endorsement); or • A measure steward never submits a measure for maintenance, and the steward does not respond after targeted outreach; or • There is no longer a meaningful gap in care, or the measure has plateaued (i.e., no significant change in measure results for accountable entities over time).	None.

±Maintenance measures may be up for endorsement review earlier if an emergency/off-cycle review is needed (see Emergency/Off-Cycle Reviews for more details).

**Conditions are recommended by the E&M committee, with the consideration of what is feasible and appropriate for the developer/steward to execute by the time of maintenance endorsement review.*

°Measures that fail to reach the 75% consensus threshold are not endorsed.

Endorsed with Conditions

The "Endorsed with Conditions" category serves as a means of endorsing a measure but with conditions recommended by the committee. These conditions are actions/activities that the developer/steward should undertake prior to the next maintenance cycle. When considering a condition to endorsement, the E&M committee should assess what is feasible and appropriate for the developer/steward to execute by the time of maintenance endorsement review. Table 4 lists the types of conditions that may be applied to a measure, in accordance with the respective domains of the PQM Measure Evaluation Rubric. Prior to Recommendation Group members voting during the endorsement meeting, E&M project staff will finalize conditions based on committee recommendations.

Table 4. Types of Conditions that May Be Placed on a Measure

PQM Rubric Domain/Criterion*	Condition(s)	Example
Importance	a. Conduct additional evaluation/assessment of meaningfulness to the patient	a. Developer/steward has not, or to a limited degree, provided evidence from literature, focus groups, expert panels, etc. that the target

PQM Rubric Domain/Criterion*	Condition(s)	Example
	community (e.g., patients, caregivers, advocates). b. *[For maintenance]* Expand performance gap testing to a larger population.	population (e.g., patients) values the measured outcome, process, or structure and finds it meaningful for improving health and health care. b. Maintenance measure has narrow gap, which may be due to limited data/testing within a population that may not be fully representative.
Reliability	a. Consider mitigation strategies to improve measure's reliability, such as increasing the case volume, including more than 1 year of data. For any facilities that are unable to exceed the threshold, give a rationale for why the reliability being below the threshold is acceptable for those specific facilities.	a. The developer/steward has performed measure score reliability testing (accountable entity-level reliability). The majority of facilities have a reliability that exceeds the accepted threshold of 0.6 but around 30% of facilities are below the threshold.
Feasibility	a. Provide implementation guidance or a near-term path (within 1 year) for implementing the measure. This includes providing clear system requirements for implementation of the measure.	a. Measure has experienced implementation challenges.
Use and Usability	a. Implement a systematic feedback approach to better understand if challenges exist with implementing the measure. b. Collect additional feedback from providers to ascertain the reasons why the measure is leveling off and describe appropriate mitigation approaches.	a. Measure has limited feedback due to low use and/or non-systematic feedback approach. b. Trend data show a leveling off of measure performance.

**At the time of publishing this guidebook, the Equity domain is an optional domain. Therefore, no conditions can be placed on the measure related to the Equity domain.*

If a measure with an "Endorsed with Conditions" designation is evaluated for maintenance but it has not met the prior conditions, then the committee may choose to remove endorsement, unless it agrees with any rationale provided by the developer/steward.

Battelle has identified several non-negotiable areas, meaning if a measure meets one or more of the following criteria, the measure cannot be endorsed, even with conditions (Table 5).

Table 5. Non-negotiables that Cannot Be Conditions

PQM Rubric Domain/Criterion*	Example
Importance	• Lack of or unclear business case for the measure. • Lack of evidence supporting the business case. • *[For maintenance]* Lack of sufficient evidence that a performance gap exists.
Scientific Acceptability	• Specifications, testing approach, results, or data descriptions are insufficient for the committee to apply the PQM Measure Evaluation Rubric (*Appendix D*). • Inappropriate methodology, calculations, formulas, or testing approach used to demonstrate reliability or validity. This includes not testing the measure as specified.
Feasibility	• Significantly poor feasibility of the measure (e.g., majority of test sites showed challenges with implementing the measure specifications) to be implemented due to challenges with data availability or missingness and/or due to substantial proprietary mechanisms prohibiting a measure's potential use.
Use and Usability	• No plan for use within an accountability application. Accountability applications are uses of measure performance results about identifiable, accountable entities to make judgments and decisions because of performance. This can be as confidential reporting, reward, recognition, punishment, payment, or selection (e.g., public reporting, accreditation, performance-based payment, network inclusion/exclusion).

** At the time of publishing this guidebook, the Equity domain is an optional domain. Therefore, no non-negotiables are related to the Equity domain.*

Novel Hybrid Delphi and Nominal Groups

Each project uses the Novel Hybrid Delphi and Nominal Groups (NHDNG) technique[1] for measure endorsement reviews. The NHDNG is a comprehensive, adaptable tool that is employed to build consensus among E&M committee members and leverage experienced and trained facilitators. Further, the NHDNG technique is a hybrid technique, utilizing a multi-step process meant to increase engagement of all committee members and to structure meeting facilitation by using standard measure evaluation criteria and practices.

This NHDNG approach allows for efficient information exchange among E&M committee members, which is particularly important when members offer unique points of view. Furthermore, this approach ensures all members have access to the same information prior to final evaluation.

Recent Enhancements to the E&M Process

We appreciate the participation and engagement of all interested parties during the Fall 2023 cycle, including E&M committee members, developers, stewards, and members of the public. We are committed to finding new opportunities for further improvements and efficiencies to

[1] Davies S, Romano PS, Schmidt EM, Schultz E, Geppert JJ, McDonald KM. Assessment of a novel hybrid Delphi and nominal groups technique to evaluate quality indicators. Health Services Research. 2011 Dec;46(6pt1):2005-18. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-6773.2011.01297.x

enhance consensus-building and to capture increased measure review input in a less burdensome way while maintaining the diversity of voices.

After the Fall 2023 endorsement meetings, we received feedback on the novel E&M process from committee members, measure developers/stewards, and other interested parties. In response to that feedback, we implemented several changes to the E&M committee roles, voting, and meeting structure, while maintaining a 6-month timeline (Figure 1) and high standards for transparency and rigor. The proposed changes will take effect in the Spring 2024 cycle. Our goals are to:

- Enhance the engagement and participation of Advisory Group members, patient partners, and members of the public;
- Reduce burden for E&M committee members (both Advisory and Recommendation Groups); and
- Improve clarity of roles for E&M committee members by aligning processes across both E&M and Pre-Rulemaking Measure Review (PRMR).

From March 1-22, 2024, we conducted a public comment period to collect feedback on the proposed changes. The comments and our responses to the comments are available on the PQM website. We also summarized the changes and our response to public comments during an informational webinar on April 22, 2024. The changes have been incorporated throughout this version of the E&M Guidebook.

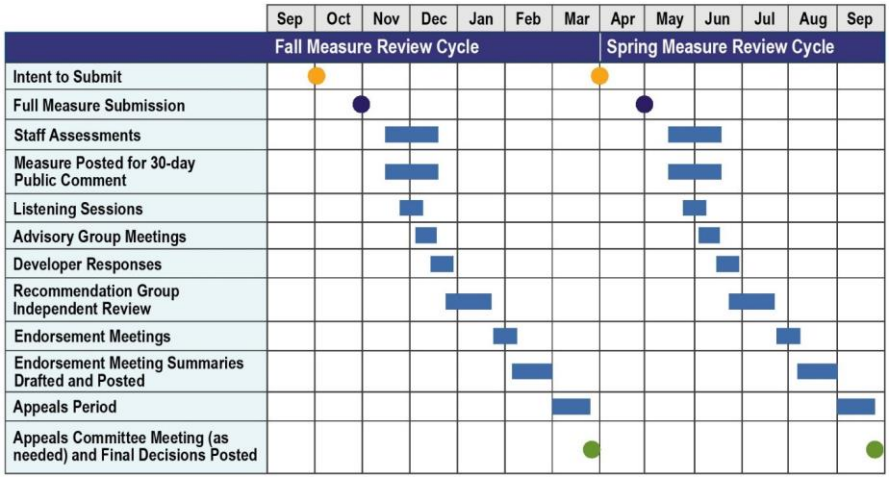

Figure 1. A 6-month Endorsement Review Process

E&M Committee Composition, Roles, and Responsibilities

We ensure a diversity of E&M committee membership through a formal nominations process (*see E&M Committee Nominations for more details*) to fill gaps in expertise and needed perspective. As a CBE, Battelle reviews nominations and selects individuals to serve on committees to participate in the E&M process. E&M committees are composed of diverse members representing all facets of the health care system. Each cycle has up to five projects, and each project has a committee that evaluates, discusses, and assigns ratings and endorsement decisions for measures under endorsement review (see Table 1).

Each E&M project committee is divided into an Advisory Group and a Recommendation Group (Figure 2), consisting of interested parties from PQM membership. This structure of membership organization enables use of the NHDNG technique, which maximizes member engagement and promotes consistent application of evaluation criteria.

- **Advisory (Delphi) Group**: The Advisory Group consists of 35-40 people. Members in this group review measures and are convened to provide feedback and questions regarding the measure(s) under review, 1-2 months prior to the Recommendation Group endorsement meeting. These inputs ensure a larger number of voices contribute to the consensus-building process.
- **Recommendation (Nominal) Group**: The Recommendation Group consists of 20-25 people and is the endorsement voting body of the committee. Members in this group also review and provide ratings and written comments on measures prior to the Recommendation Group endorsement meeting. In addition, members review the Advisory Group's feedback and questions, public comments, and respective developer/steward responses pertaining to the measure(s) under review prior to the endorsement meeting. Recommendation Group members consider and discuss these various inputs before rendering an endorsement decision via a vote during the endorsement meeting.

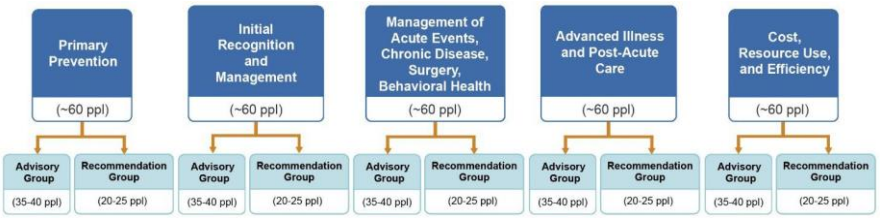

Figure 2. Recommendation and Advisory Group Structure

Each E&M project committee (the Recommendation Group plus the Advisory Group) has two co-chairs, who participate in the endorsement meeting and take part in the Recommendation Group discussions. When possible, we ensure at least one co-chair is from the patient community. The patient representative co-chair is responsible for engaging and supporting

patient representatives on their respective committee. The other co-chair is responsible for ensuring the Advisory Group's concerns and perspectives are considered by the Recommendation Group during the endorsement meeting. In addition, the co-chairs' responsibilities are to:

- Co-facilitate endorsement meetings, along with E&M project staff.
- Work with E&M project staff to achieve the goals of the project.
- Assist E&M project staff in anticipating questions and identifying additional information that may be useful to the committee.
- Participate in the Recommendation Group as a full voting member for the entirety of their term.
- Serve on the Appeals Committee.

To ensure representation of the population of interested parties, up to 60 PQM members are seated on an E&M project committee through a formal nominations process (*see E&M Committee Nominations for more details*), which is conducted annually to fill gaps in expertise and roster categories (Table 6). To serve on an E&M committee, individuals must also be PQM members. We seat PQM members based on the expertise needed for the E&M project, ensuring adequate representation and perspectives across roster categories. Additionally, all newly seated and currently serving committee members are invited to an annual E&M virtual orientation meeting in the fall, which provides an overview of the E&M process, committee roles and responsibilities, as well as a review of the PQM Measure Evaluation Rubric (*Appendix D*).

As needed, Recommendation Group membership may be augmented with individuals with specialized expertise, which is determined after each cycle's Intent to Submit deadline. For example, if a health care cost measure for a specific disease state or condition is under review by the Cost, Resource, and Efficiency committee, subject matter experts (SMEs) familiar with that disease state or condition are invited to the endorsement meeting to provide further context and relevance for the committee's consideration. These SMEs are non-voting participants and will only provide input on relevant measures.

If additional expertise is needed, we first identify if the needed expertise resides within one of the other E&M committees. If SME expertise is still absent from other E&M committees, we and our partner, the Institute for Healthcare Improvement (IHI), recruit SMEs from our combined networks, and selection is vetted against our conflict of interest policy and with input from the respective committee co-chairs. If needed, we may establish a pool of SMEs across various clinical (e.g., nephrologists, primary care providers) and methodological (e.g., psychometricians) areas. We also encourage developers/stewards invite SMEs from their technical expert panels to participate and answer committee questions during endorsement meetings.

Table 6. Roster Categories and Target Number of Individuals

Roster Category	Advisory Group Targets*	Recommendation Group Targets*
Patients, families, caregivers, patient advocates	8	4
Clinicians, including physicians, nurses, pharmacists, physical therapists, etc.	3	5
Facilities/institutions including accountable care organizations, hospitals or hospital systems, and post-acute/long-term care facilities	3	5
Purchasers and plans (state, federal, and/or private)	5	3
Rural health experts	2	2
Health equity experts	2	2
Researchers in health services, alternative payment models, population health	6	2
Other interested parties (representatives of electronic health record [EHR] vendors, provider and facility associations, and experts in areas such as quality improvement/implementation science, care coordination, patient safety, behavioral health, and national policy makers)	6	2
TOTAL±	**35**	**25**

**Note: If Battelle does not fill the number of seats listed for a given roster category, Battelle will determine if remaining seats can be distributed to other roster categories, based on the expertise needed within the committee.*

±Totals may fluctuate between 35-40 for the Advisory Group and 20-25 for the Recommendation Group.

Term of Appointment

Committee members are appointed to a 3-year term and will serve on both the Advisory and Recommendation Groups. Newly appointed committee members are initially seated to the Advisory Group for the first 2 years of their term and then move into the Recommendation Group to conclude their 3-year term. This approach ensures each member of the Advisory Group will have the opportunity to serve on the Recommendation Group within their term. Following each nominations period, Battelle will assign two co-chairs to serve on the Recommendation Group for each committee. Co-chairs are selected based on expertise and/or lived experience as well as interest in serving in this role. One of the two co-chairs will represent the patient perspective.

We ensure no more than one-third (1/3) of members roll off the committee every year. Committee members who roll off the committee, including co-chairs, may reapply to serve on

the committee during the call for nominations. Former committee members do not need to wait before reapplying.

E&M Committee Nominations

We conduct a review of committee member appointments annually, which includes internal recalibration of the membership, a call for nominations, and targeted outreach. Beginning in late spring, a call for nominations is published on the PQM website and an announcement is sent out to all PQM members. Nominations are submitted via the PQM website; self-nominations are welcome. Third-party nominations must indicate that the individual has been contacted and is willing to serve. Nominees must be PQM members (which is free), and they must complete a nomination form and a Personal/Organizational Disclosure of Interest (DOI) form (*Appendix B*). Before finalizing the appointments, a draft roster of nominees is published for public comment for transparency and for garnering input as to whether the E&M roster has the expertise needed for the given E&M project. Committee member appointments are finalized later in the summer and take effect beginning with the Fall cycle.

Nominees commit to participating in scheduled calls and meeting dates, providing timely responses to requests for feedback, and being available for ad hoc meetings and conference calls. To be eligible for participation, nominees should (1) have relevant expertise and demonstrated experience related to the use of quality and efficiency measures and/or (2) belong to at least one of the categories listed in Table 6.

Committee members are responsible for notifying the E&M project team if they:

- Change employers or contact information;
- Are unable to attend a scheduled meeting; or
- Have a prolonged conflict emerge during their term that will interfere with meeting the obligations of E&M committee membership. This information is used to determine whether ongoing membership on the committee is warranted or if inactive status can be granted for a cycle.

Inactive Status and Early Termination

We understand plans and demands of our volunteer E&M committee members change. Therefore, members may need to move to inactive status for a given review cycle or end their terms early. E&M committee members with inactive status continue with their terms, but for the cycle of interest, they are not permitted to vote and are therefore not counted in the denominator when determining meeting quorum and voting thresholds. A committee member may be granted

inactive status at any time before their respective meeting (e.g., Advisory Group meeting or Recommendation Group meeting).

If a committee member has poor attendance or participation, as determined by not attending one or more respective meetings without advanced notice and/or by not submitting independent reviews of measures (Recommendation Group only) for endorsement review (*see Independent E&M Committee Member Review and Assessment for more details*), we will contact the member and ask if he/she would like to resign. We reserve the right to remove any member from an E&M committee, including for reasons of persistent poor attendance or lack of participation.

Conflict of Interest Policy

As a CBE of contract 75FCMC23C0010 with CMS, Battelle convenes several committees of interested parties to provide input on (1) endorsement decisions on quality performance and cost/resource use measures; (2) the selection of measures for a pre-rulemaking process, which is required by Social Security Act Sections 1890(b)(7) and 1890A; and (3) a measure removal process. This Conflict of Interest Policy (the "Policy") is applicable to such committees to ensure each committee performs its functions in a manner free from bias and undue influence. All committee members must attest they will follow this policy and provide the requisite information necessary for Battelle to conduct a conflict of interest (COI) review.

The term "conflict of interest" means any financial or other interest that could actually or be perceived to (1) significantly impede your objectivity or (2) create an unfair competitive advantage for you or an organization associated with you. Disclosure of a financial interest does not automatically mean a COI exists but may warrant further discussion and review.

As part of the E&M committee nomination process, each nominee completes a Personal/Organizational Disclosure of Interest form (*Appendix B*). In addition, to complete the COI analysis, each member serving on a committee evaluating measures for endorsement and/or for providing recommendations for pre-rulemaking will be required to complete a Measure Disclosure of Interest Form for each measure, or batch of measures, assigned to that committee (*Appendix C*). This form will contain questions relevant to the specific measure(s) being reviewed. Battelle will provide the Measure Disclosure of Interest Form to committees at the start of each cycle. The form will contain questions regarding the member's financial interests and business associations, which may present a perceived or actual COI.

The questions in the Measure Disclosure of Interest Form focus on whether:

(1) You contributed directly and substantially to the development of a measure or measures being considered for endorsement or under consideration for selection or removal. For example:

- You worked on the measure as an employee of or consultant for the measure development organization.
- You directly collaborated with the measure development organization to create or refine the measure.

(2) You or your spouse, domestic partner, or child could receive a direct financial benefit from a measure being recommended for selection, removal, or endorsement. For example:

- You own stock in a company that has a financial interest in the measure being endorsed or not endorsed.

(3) In the last 5 years you have received an indirect financial benefit, i.e., not related to the measure under review, of $10,000 or more from a measure developer whose measure is under review, or an indirect financial benefit of $10,000 or more, in the aggregate, from an organization or individual that may benefit from a measure being endorsed or not endorsed or being considered for the selection or removal process. For example:

- You have received $20,000 in consulting fees from the measure developer in the last 5 years for work unrelated to the measure being reviewed.

(4) You are currently employed by the measure developer and the developer has created the measure(s) under review, has created measure(s) in the topical area under review, or has created measure(s) that compete with measure(s) created by another developer and are under review.

By participating as a committee member, each member consents to public disclosure of general information about the member's financial or business interests, professional associations, and experiences that may be of interest to the public regarding COI. Members must also disclose their organizational affiliation and any organizational conflicts of interest. Unless legally required to do so by an authoritative entity, such as CMS, specific financial information will not be provided to the public, but financial relationships may be subject to disclosure.

If you provide information that creates a perceived or actual COI, Battelle requires you recuse yourself from any voting regarding the applicable measure or measures, and in some instances, competing and related measures. However, you may still contribute to the discussion of the measure(s). Committee members who have conflicts with specific measures, as determined by the Measure Disclosure of Interest Form, must publicly recuse themselves from any voting associated with those measures.

Additionally, committee members must orally disclose relevant interests at a public committee meeting. The disclosure usually occurs at a committee's endorsement meeting. Senior Battelle staff will lead this disclosure and instruct committee members regarding information that should be disclosed. Following oral disclosure by committee members, Battelle staff will invite committee members to ask and respond to questions of each other or Battelle staff regarding any disclosures made by committee members.

Finally, all committee members have an ongoing duty to monitor their own COI issues and those of their fellow committee members and raise or disclose any issues, either in a committee meeting or to the committee chair, the Battelle program team, or the Battelle legal department. Committee members should take a proactive approach and report any instances of a fellow committee member appearing conflicted or acting in a biased manner.

Submitting Measures to Battelle

The E&M process consists of a series of stages, starting with Intent to Submit and cascading to the appeals period (Figure 3). During each stage, we work closely with developers and stewards, committee members, and other interested parties to address questions regarding process and/or criteria. We also conduct the endorsement meetings and provide all relevant materials and documentation of the endorsement deliberations and decisions, including committee rationales. We inform all interested parties of the status of measures going through the process and welcome public comment on the measures and endorsement decisions throughout the review cycle. Lastly, all information pertaining to the E&M committee meetings, the measures being reviewed, and the E&M meetings themselves are made public.

Requirements for Measure Consideration

Prior to any measure being considered and evaluated for endorsement, several requirements must be met. If any of the requirements listed below are not met, the measure will not be accepted for endorsement review. Measure developers/stewards should contact PQMsupport@battelle.org if they have questions about these requirements.

- A Quality Measure Developer and Steward Agreement form (*Appendix A*) is signed, allowing Battelle to publicize the measure, including any proprietary information associated with the measure.
- The measure must include data from the past 5 years. This includes data used for testing, performance gap and trend analyses, and stratification.

- The measure is fully specified and tested for reliability and validity.
- The measure specifies a responsible entity (i.e., accountable entity) and any analyses conducted (e.g., performance gap, reliability and validity testing, trend analyses) are performed using the data source(s) and level(s) of analysis for which the measure is specified.

- The intended use of the measure includes accountability applications[2] to achieve high-quality efficient health care.
- The measure submission information is complete and responsive to all relevant submission items so that all the information needed to evaluate the measure is provided.

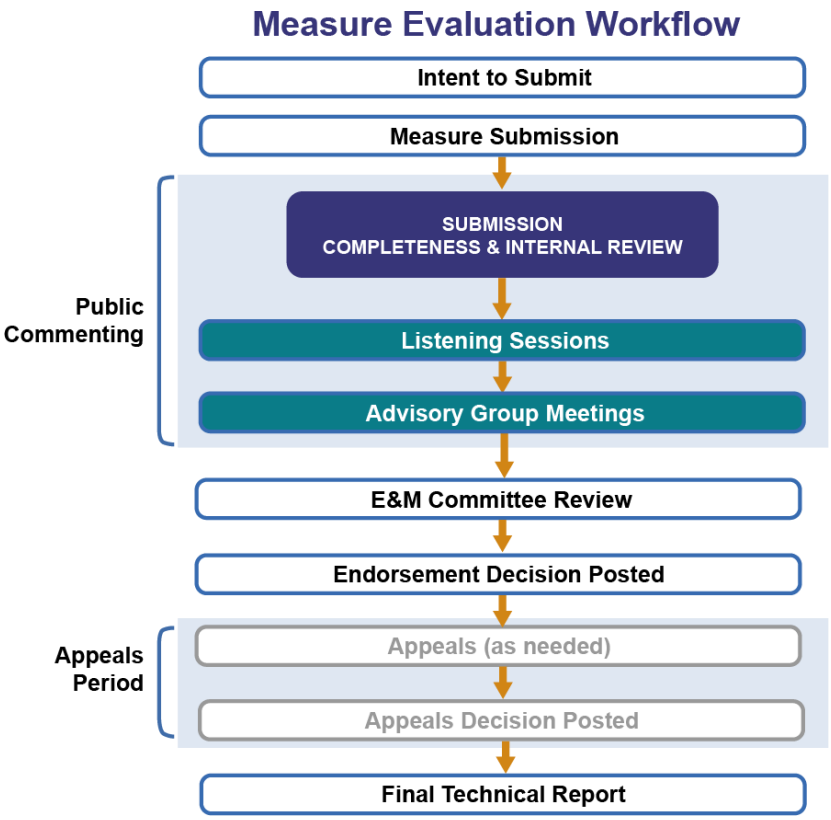

Figure 3. Measure Evaluation Workflow, 6-month E&M Cycle

[2] Accountability applications are uses of measure performance results about identifiable, accountable entities to make judgments and decisions because of performance. This can be as confidential reporting, reward, recognition, punishment, payment, or selection (e.g., public reporting, accreditation, performance-based payment, network inclusion/exclusion).

Intent to Submit

The Intent to Submit period is when measure developers/stewards submit key information about the measure via STAR at least 1 month prior to the full measure submission deadline of the intended review cycle (Fall or Spring). For all measures (new and maintenance), developers/stewards must submit the following information during Intent to Submit:

- Intended Measure Review Cycle
- Measure Title
- Measure Description
- Measure Type (e.g., structure, process, outcome)
- Measure Specifications (e.g., numerator, denominator, level of analysis, care setting)
- Intended E&M Project for Evaluation (Table 1)
- Contact Information and Affiliation
- Attestations for What is Required by Full Measure Submission

Throughout and leading up to the intended measure review cycle, developers/stewards may request technical assistance, which we provide (*see Technical Assistance for more details*).

Full Measure Submission

Completeness Checks

Within 1 month of completing the Intent to Submit, developers/stewards must submit all the measure information via STAR's online measure submission function. Requirements for initial and maintenance measure endorsement are indicated as, "*[For initial endorsement]*" or "*[For maintenance]*," within each domain of PQM Measure Evaluation Rubric (*Appendix D*). If neither distinction is listed for a rubric requirement, then the requirement applies to both initial and maintenance endorsement.

We conduct completeness checks (see *Measure Submission Completeness Checklist* below) to determine if all required responses and measure information have been submitted. We notify measure developers/stewards of any issues and request developers/stewards address the completeness check feedback by a deadline, which is no less than 2 business days from receipt of the completeness check feedback. Measures that pass the completeness check review are posted for a 30-day public comment period while simultaneously undergoing an internal measure review by E&M project staff (see *E&M Staff Preliminary Assessment* for more details). Measures that do not pass the completeness check will be pulled from consideration.

Submission of Electronic Clinical Quality Measures (eCQMs)

The following clarifications are specific to eCQMs:

- A new eCQM version of an endorsed measure is not considered an endorsed measure until it has been specifically evaluated and endorsed by Battelle. An eCQM should be submitted as a separate measure even if the same or a similar measure exists.
- Measure specifications should use the latest accepted versions of the following industry eCQM technical specifications: Health Quality Measure Format (HQMF), Quality Data Model (QDM), and Clinical Quality Language (CQL). Use of the CMS Measure Authoring Tool (MAT) ensures the measure uses these technical specifications; however, the MAT is not required to produce HQMF.
- Developers/stewards must use value sets that are published through the National Library of Medicine's Value Set Authority Center (VSAC). This helps reduce implementation issues related to value sets and code system validation and encourages the use of harmonized value sets. If an eCQM does not have a published value set, then the measure developer must look to see if there is a published value set that aligns with the proposed value set within its measure. If such a published value set does not exist, then the measure developer must demonstrate the value set is in draft form and is awaiting publication to VSAC.
- Developers/stewards must conduct testing within EHR systems from at least two EHR vendors. Beyond this minimum requirement, developers/stewards should test on the number of health systems they deem appropriate. Submission requirements for eCQMs also include a feasibility assessment, using the eCQM Feasibility Scorecard. This assessment identifies data element feasibility issues. Simulated data set results allow assessment of each branch of the measure logic to ensure the logic can be processed technically by other eCQM-capable reporting tools.
- Empirical demonstration of data element reliability is required for any unstructured data fields, and data element validation is required for all eCQMs. If the testing is focused on validating the electronic data elements, developers/stewards should analyze agreement between the electronic data obtained using the eCQM specifications and those obtained through abstraction of the entire electronic record (not just the fields used to obtain the electronic data). Developers/stewards should use statistical analyses, such as sensitivity and specificity, positive predictive value, and negative predictive value. This type of validity testing also satisfies the requirement for reliability testing. If data element testing is not possible, justification is required and must be accepted by the E&M committee. Face validity alone will not be sufficient.

Commented [HSAG2]: Please consider updating, given the decommissioning of MAT in June 2024.

Commented [HSAG3]: Please consider consistent use of terms such as "Person- or Encounter-Level Reliability" versus "data element reliability" to align with the Scientific Acceptability rubric. Additionally, please consider aligning with similar terms used in the MMS Blueprint.

Commented [HSAG4]: Please consider clarifying whether data element validity for eCQMs is required only for critical data elements, or all data elements.

Measure Submission Completeness Checklist

Developers/stewards are also encouraged to follow the checklist below to ensure the measure submission is complete and responsive prior to E&M committee review.

- ☐ The Quality Measure Developer and Steward Agreement (QMDSA) is completed and signed (*Appendix A*).
- ☐ Complete and adequate responses have been received for all relevant and required fields within the measure submission form.
- ☐ Testing is conducted for the data source(s) and level(s) of analysis for which the measure is specified; information for data source and level of analysis is consistent across the specifications and testing items.
- ☐ Attachments, including electronic clinical quality measure (eCQM) specifications, Feasibility Scorecard, and data dictionary/code lists, have been included if applicable and appropriate.
- ☐ All uniform resource locators (URLs) are active and accurate.
- ☐ All measure submission information, including attachments, is 508 compliant (*see Appendix E for more details*)
- ☐ Paired measures are submitted on separate forms.
- ☐ ICD-10 (International Classification of Diseases) codes are used and included, if applicable.

E&M Project Staff Preliminary Assessment

We review each measure submission using the PQM Measure Evaluation Rubric (*Appendix D*). Measures are evaluated on five domains (Importance, Feasibility, Scientific Acceptability [i.e., Reliability and Validity], Equity,[3] and Use and Usability). For each domain, we indicate if a measure domain has been "Met," "Not Met but Addressable," or "Not Met," based on specific evaluation considerations for each area. These preliminary assessments summarize key points of the submission as they pertain to the PQM Measure Evaluation Rubric and, when appropriate, provide additional context or interpretation for certain aspects of the submission (e.g., verifying a testing methodology is appropriate). The preliminary assessment ratings are not binding, but, instead, are meant to serve as input for committee discussion.

We share these staff preliminary assessments with developers/stewards for a factual review prior to sharing them with the Recommendation Group for endorsement consideration. Developers/stewards are asked to conduct a factual review by the requested deadline, which is

[3] Note: The Equity domain is currently optional, Battelle continues to explore this, but to align with national priorities, Battelle encourages developers and stewards to address this domain, if and when possible.

no more than 2 business days from receipt of the preliminary assessment. This factual review is to ensure the preliminary assessments include accurate results from the measure submission. For example, when summarizing the testing results of a measure, have we accurately reflected the testing results? This factual review *is not* intended to provide an opportunity for developers/stewards to disagree with the preliminary measure ratings.

After developer/steward review, we finalize the staff preliminary assessments and share them publicly on the respective measure webpage on the PQM website.

Public Comment

Once a measure submission passes the completeness check step, the full measure submission details, including all attachments, are posted to the PQM website for a 30-day public comment period. This public comment period occurs prior to the endorsement meeting and concurrently with the development of the E&M project staff preliminary assessments. The intent of this 30-day comment period is to solicit both supportive and non-supportive comments with respect to the measure(s) under endorsement review. Any interested party may submit a comment via the PQM website on any of the measures up for endorsement review for a given cycle (e.g., Fall or Spring). Public comments submitted via the PQM website are posted to the respective measure page for full transparency.

Public Comment Listening Sessions

Two to three weeks prior to the close of the public comment period, we host Public Comment Listening Sessions. Any interested party can register to attend one or more of these virtual sessions to give a brief verbal statement on one or more of the measures under endorsement review for that cycle. Commenters are kindly asked to keep their comments to 2 minutes or less.

We share transcripts of the comments from these listening sessions with developers/stewards for review and written response. No less than 1 week prior to the endorsement meetings, we make publicly available and share the developer/steward responses to the public comments with the Recommendation Group. The Recommendation Group is tasked with reviewing the comments and developer/steward responses for endorsement decision-making.

Advisory Group Meetings

Following the listening sessions, we convene public Advisory Group meetings 1-2 months prior to the endorsement meetings. The purpose of these meetings is for Advisory Group members to raise questions and share perspectives verbally regarding the measures under endorsement review for their respective E&M committee. No voting occurs during these virtual meetings.

Advisory Group members will be asked to review the measures assigned to their respective committee and come to the meeting to ask questions and provide feedback regarding the strengths and limitations of the measures. Recommendation Group members and measure developers/stewards are invited to attend these Advisory Group meetings to listen to the discussion. In addition, developers/stewards will have the opportunity to respond to Advisory Group questions during the meeting.

We summarize the feedback and questions received from the Advisory Group members and share this with developers/stewards for review and written response. No less than 1 week prior to the endorsement meetings, we share the Advisory Group feedback and questions, along with the developer/steward responses, with the respective Recommendation Group for endorsement consideration.

Independent E&M Committee Member Review and Assessment

At least 3 weeks prior to an E&M committee endorsement meeting, the Recommendation Group of each E&M committee receives the full measure submission details for each measure up for review, including all attachments, the PQM Measure Evaluation Rubric, and the E&M project staff preliminary assessments (*referred to as "E&M Committee Review" in Figure 3*).

Recommendation Group members are asked to review each measure, independently, against the PQM Measure Evaluation Rubric. Committee members assign a rating of "Met," "Not Met but Addressable," or "Not Met" for each domain of the PQM Measure Evaluation Rubric. In addition, committee members provide associated rationale for each domain rating, which is based on the rating criteria listed for each domain (*see Appendix D for more details*). We aggregate and summarize the results, which are then used by Battelle facilitators and committee co-chairs to guide measure discussions during the endorsement meetings (*see Endorsement Meeting section for more details*). The use of these independent committee reviews anchors opinions based on each individual's knowledge and limits the likelihood that a vocal few impart too much bias on the results.

Endorsement Meeting

We convene the Recommendation Group of each E&M committee for an endorsement meeting each cycle. Spanning 1-2 days, all meetings are held virtually and are open to the public. Measure developers/stewards are also invited to the endorsement meeting to introduce their measure(s), to provide further context and rationale for their measure(s), and to answer questions posed by the Recommendation Group during designated times.

During the endorsement meeting, the E&M team and committee co-chairs focus the Recommendation Group discussions on the identified strengths and limitations of the measure(s) under review. This is achieved by prioritizing the findings from the public comments received, the Advisory Group meetings, and the associated developer/steward responses. The E&M project staff preliminary assessments and results are also taken into consideration in these facilitated discussions.

After the discussions conclude for a measure, the E&M committee co-chairs summarize the deliberations of the Recommendation Group before moving to an endorsement vote. Within this summary, co-chairs draw attention to the issue(s) discussed, noting any endorsement conditions and clearly capturing the committee rationales for supporting and not supporting the measure. These rationales are based on whether PQM Measure Evaluation Rubric domains are "Met," "Not Met but Addressable," or "Not Met."

The committee does not vote on each domain of the PQM Measure Evaluation Rubric; rather, the Recommendation Group members only vote on the endorsement decision (Table 3). The committee provides its rationale for each endorsement decision, noting any deficiencies in the submission, the measure specifications, or the failure to identify deficiencies in the submission. For example, if committee members determine the Scientific Acceptability is "Not Met" due to concerns with the reliability testing results being low, then this rationale, plus any other issues identified, is clearly stated with the endorsement vote. If the committee does not reach consensus on an endorsement vote, then the measure is not endorsed.

Defining Consensus

Consensus is determined to be 75% or greater agreement among members. Battelle established the 75% threshold of consensus to be consistent with the goal of adding rigor to all aspects of the consensus development process. The threshold is based on an evidence-based index (Appendix F). Analogous to inter-rater reliability statistics, the evidence-based index assess the degree of disagreement (or lack of consensus) amongst the independent committee reviews and the committee endorsement votes. The evidence-based index is one of agreement, where the closer to 1.0, the more there is agreement or consensus. From the table in Appendix F, when the index is 0.95 or greater, the corresponding threshold of consensus is 75%. This approach is advantageous compared to other metrics based on variance, in that it takes into consideration the different sizes of the voting groups and different rating options (*see Appendix F for more details*). Figure 4 below depicts how endorsement decisions are reached based on the 75% consensus threshold.

A measure is endorsed when 75% or more of committee members vote to endorse the measure (Scenario 1 in Figure 4). A measure receives an endorsed with conditions decision if 75% or more of committee members vote to endorse the measure with conditions or if 75% or more of committee member votes are distributed across endorse and endorse with conditions (Scenario 2). Lastly, a measure is not endorsed (new measures only) or its endorsement is removed (maintenance measures only) if 75% or more of committee members vote to not endorse the measure (Scenario 3) or if the committee does not reach consensus (Scenario 4).

Scenario	Endorse (A)	Endorse with Conditions (B)	Do Not Endorse (C)	Consensus Voting Status
1	**75% or More**	0%	Less than 25%	A
2	**75% or More**		Less than 25%	B
3	Less than 25%		**75% or More**	C
4	26% to 74%		26% to 74%	No consensus

Figure 4. Consensus Voting for Final Endorsement Decisions

Quorum

A crucial aspect of a successful consensus-based process is effective and organized meeting facilitation to ensure discussions remain productive, within scope, and inclusive of all voices. At the beginning of the endorsement meetings, Battelle facilitators confirm quorum and engage committee members in robust discussion to build consensus recommendations about each measure under review.

The **consensus-based process** ensures:

- Productive discussions
- Discussions within scope
- Inclusion of all voices
- Increased engagement
- Efficient information exchange

Having a quorum for meeting attendance and for voting is critical to ensuring the Recommendation Group discussions and the endorsement vote are robust and reflective of all perspectives represented on the E&M committee. Meeting quorum requires at least 60% of only the Recommendation Group members to be present during roll call at the beginning of the meeting (Figure 5).

Figure 5. Quorum Thresholds

If less than 60% of the Recommendation Group members are in attendance, then the Recommendation Group will not discuss the measures and a back-up meeting will be held. If meeting quorum is lost during the meeting, the measure evaluation discussions will cease and an alternative meeting will be held to complete the measure review.

Voting quorum is at least 80% of all active Recommendation Group members present who have not been recused (*see Conflict of Interest Policy for more details*). If the voting quorum is not met at committee roll call but meeting quorum is achieved, the Recommendation Group will proceed with discussing the measures but endorsement voting will not occur during the meeting. After the endorsement meeting has ended, the E&M team will share the meeting recording with those Recommendation Group members not in attendance and request they submit their endorsement vote via an offline voting tool. Recommendation Group members will have 2 business days to submit their votes.

We monitor attendance throughout the endorsement meeting to ensure both meeting quorum and voting quorum are maintained. We promote high attendance among voting committee members by engaging them early and often, including providing meeting notices well in advance of scheduled meetings and sending detailed meeting agendas and measure submission information with sufficient time for review.

We acknowledge committee member priorities may change from time to time, impacting their ability to meaningfully participate in the E&M process. We employ a policy allowing committee members to be inactive for a given cycle, based on competing priorities, or end a committee member's term early if there is a consistent lack of participation and/or engagement (*see E&M Committee Composition, Roles, and Responsibilities section for more details*).

Endorsement Decision Posted

After the endorsement meetings, we publicly share all measure endorsement decisions and associated rationales on the PQM website. This starts the 3-week appeals period, during which any interested party may request an appeal regarding any endorsement decision rendered by the Recommendation Group (Table 3).

Appeals

When an appeal is received, we conduct a preliminary review to determine its eligibility based on the criteria for the respective endorsement decision. If an appeal is not eligible, we notify the appellant, noting the eligibility criterion/a not met. However, if an appeal is eligible, an ad hoc Appeals Committee is convened to review and discuss the appeal, followed by a vote to uphold (i.e., overturn a committee endorsement decision) or deny (i.e., maintain the endorsement decision) the appeal. Consensus is determined to be 75% or greater agreement among the Appeals Committee.

The Appeals Committee consists of all co-chairs from each of the five E&M project committees from the respective endorsement cycle. If additional perspectives are needed, we send ad hoc requests to the PQM membership. This structure ensures these meetings can be convened quickly and as needed, and the inclusion of E&M project staff and committee chairs reduces the risk of duplicative or contradictory discussions. If needed, SMEs may be recruited, as non-voting participants, to support the Appeals Committee discussions. We employ the same SME-recruitment approach as with E&M committees. However, the co-chairs of the committees that did not receive an appeal will vet the SMEs.

To promote transparency and accountability, Appeals Committee meetings are open to the public, and a meeting summary is shared publicly via the PQM website. All Appeals Committee decisions are final.

Appeals Eligibility Criteria

If a measure's endorsement is being appealed, including an "Endorsed with Conditions" decision, the appeal must cite evidence that the appellant's interests are directly and materially affected by the measure, and the CBE's endorsement of the measure has had, or will have, an adverse effect on those interests. The appeal must also include one of three rationales:

- Evidence exists that was available by the cycle's Intent to Submit deadline but was not considered by the E&M committee at the time of the endorsement decision and is reasonably likely to affect the outcome of the original endorsement decision.

- The CBE's measure evaluation criteria were not applied appropriately. The appellant must specify the evaluation criterion that they believe was misapplied and why.
- The CBE executed a procedural error (i.e., CBE's E&M process was not followed). The appellant must specify the error/process step, how it was misapplied/not followed properly, and how this resulted in the measure being endorsed.

In the case of a measure not being endorsed (new measure) or its endorsement removed (maintenance measure), the appeal must be based on one of two rationales:

- The CBE's measure evaluation criteria were not applied appropriately. The appellant must specify the evaluation criterion that they believe was misapplied and why.
- The CBE executed a procedural error (i.e., CBE's E&M process was not followed). The appellant must specify the error/process step, how it was misapplied/not followed properly, and how this resulted in the measure not being endorsed.

Final Technical Report

The E&M team develops and publishes a technical report for each project upon completion. Each technical report includes the following information:

- A summary of the scope of review conducted under the E&M project.
- A list of the performance measures submitted and evaluated under the E&M project.
- A list of the performance measures endorsed and not recommended for endorsement under the E&M project.
- A list of measure concepts submitted during Intent to Submit for measures under the E&M project.
- A summary of the public comments received during the E&M process for the E&M project.
- A summary of any potential high-priority gap areas identified during the E&M project for measure developers to consider for future development.
- A summary of any major concerns or methodological issues raised during performance measure evaluation of the E&M project.

Harmonization

The current health care quality landscape contains a proliferation of measures, including some that could be considered duplicative or overlapping and others that measure similar but nonidentical concepts and/or define patient populations differently. Such duplicative measures and/or those with similar but not identical specifications may increase data collection burden and create confusion or inaccuracy in interpreting performance results for those who implement and use performance measures. Resolving issues around harmonizing measures and handling

competing measures is one of the key challenges. Developers/stewards must respond to the questions about harmonization in their measure submission.

ICD-10

The Department of Health and Human Services implemented conversion to ICD-10 coding on October 1, 2015. Further details explaining the changes can be accessed at https://www.hhs.gov/guidance/document/icd-10-general-equivalence-mapping. Battelle requires ICD-10 codes to replace any ICD-9 codes for all new submissions, measures undergoing endorsement maintenance, and measures with annual updates. If measures require the use of ICD-9, the measure should include ICD-9 codes with a description of the transition process used including a crosswalk of ICD-9 to ICD-10 codes and intent of the submission. Measures that are specified to capture data retrospectively may continue to be specified in ICD-9 depending on the look-back period. Some measures may be specified to capture data retrospectively and prospectively and therefore may be specified using both ICD-9 and ICD-10.

Additional Developer Resources

We engage measure developers/stewards extensively regarding the E&M process. We provide educational materials and events (e.g., webinars) for measure developers/stewards to promote transparency and a collaborative environment benefiting all interested parties.

As a CBE, Battelle cannot engage in measure development. However, each year we host a virtual Measure Developer Workshop, with the intent of engaging measure developers/stewards in cutting-edge topics relevant to measurement and E&M. For example, we share recommendations about measure evaluation criteria or testing requirements with measure developers to (1) obtain feedback on the recommendations and (2) make developers aware of potential changes to future cycles. This more deeply engages measure developers/stewards in the refinement of processes and requirements; gives interested parties a “heads-up” as to what is coming at every stage, with the intention of improving overall openness to the changes; and contributes to consensus-building by providing an opportunity for us to gather in-depth input and recommendations on improvements to the process.

In addition, E&M project staff who have measurement expertise can assist developers through the submission and review process. The E&M project staff can serve as a resource to developers through deep and nuanced understanding of the quality measure lifecycle, the tools and resources required to develop measures, the underlying measure science that guides measure development, and the time and resource constraints that impact measure development.

Technical Assistance

The E&M project staff provides technical assistance to measure developers and stewards at any time before or during the measure submission process. Contact PQMsupport@battelle.org with any questions about PQM's Measure Evaluation Rubric, how to answer the questions in the form, any technical issues with the online submission process, or anything else.

Endorsement Maintenance

Maintenance of endorsement encompasses several processes: (1) evaluations for endorsement maintenance, (2) annual updates to measure specifications of endorsed measures, (3) emergency/off-cycle reviews (i.e., early maintenance review), and (4) education and technical assistance to measure developers on endorsement maintenance activities.

Evaluations for Endorsement Maintenance

Once a measure is endorsed, it will enter a 5-year maintenance cycle, at which time the measure is resubmitted to Battelle for PQM endorsement review. However, prior to the 5-year maintenance review, at 3 years since the measure's endorsement, developers/stewards provide a status report indicating whether any changes to the measure specifications are needed. The developer/steward may also attest if no changes are required. Once the 3-year status report is submitted, we will review to confirm if any indicated changes require the measure to be submitted for endorsement review before the 5-year maintenance cycle (*see Emergency Review/Off-Cycle Reviews below*).

Annual Updates

Every year, when an endorsed measure is not being re-evaluated for continued endorsement, measure stewards have the option to submit an annual update of the measure specifications to Battelle. This report either reaffirms the measure specifications remain the same as those at the time of endorsement or last update or outline any changes or updates made to the endorsed measure, including the purpose for the changes.

If changes occur to a measure at any time in between the measure's last endorsement review and its scheduled maintenance endorsement review, the measure steward is responsible for informing Battelle immediately by submitting a status report. An early maintenance review is conducted if the changes materially affect the measure's original concept or logic (*see Emergency Review/Off-Cycle Reviews below*).

Emergency/Off-Cycle Review

Prior to the 5-year scheduled maintenance of endorsement date, a measure may require an early maintenance review due to evidence of unintended consequences (emergency review) or due to material changes to the measure specifications (off-cycle review). Early maintenance reviews are formal endorsement evaluations and follow the same processes as a maintenance of endorsement evaluation.

An early maintenance review (e.g., emergency or off-cycle review) is triggered by:

- A request by a developer/steward due to a material change to an endorsed measure during an annual update. A material change is defined as any modification to the measure specifications that significantly affects the measure results such as:
    - Changes to the population being measured (e.g., changes in age inclusions, changes in diagnoses or other inclusion criteria, changes in excluded populations, change from one type of insured population to another population);
    - Changes to what is being measured (e.g., changes in target values such as blood pressure or lipid values);
    - Inclusion of new data source(s); or
    - Expansion of the level or changing unit of analysis or care setting(s) (e.g., adding clinician level to a measure currently endorsed at practice level).
- A request by an interested party because of a perceived unintended negative consequence associated with the measure, a change in the clinical guideline driving the measure, or a significant implementation issue. The request can come from a measure developer/steward, E&M committee member, or any other type of interested party.

Battelle restricts the scope of early maintenance reviews to the immediate issue (i.e., concern with the measure's evidence, updated measure specifications and testing). An early maintenance review can be requested by any party, if there is adequate, credible, and consistent evidence to justify the review. To initiate the review, the interested party must send an email to PQMSupport@battelle.org with the subject "Emergency/Off-Cycle Review Requested," which alerts the E&M project team. The project team and respective E&M committee co-chairs review the request to see whether it is significant and emergent; for example, if the clinical practice underlying the measure is causing harm to patients directly or due to an unintended consequence. If deemed significant and emergent, the project team notifies the developer/steward (if they are not the requester of the emergency review) and pulls the measure off its maintenance cycle early to be reviewed by the E&M committee during the next immediate cycle.

The E&M team recruits additional SMEs, as needed, ensuring an appropriate combination of perspectives, from PQM. The E&M committee determines whether the measure needs immediate attention, such as a change to the specifications and shares this information with the measure developer/steward. If the change is not feasible, the committee may decide to remove

Commented [HSAG5]: Please consider clarifying whether the expansion of a measure's level of analysis is considered to be the same measure that is to be assessed as part of maintenance review and not as a new measure submission.

We would recommend that the new testing supporting the expanded level of analysis would remain under the same CBE ID.

the measure's endorsement. If the measure does not need immediate attention, the measure developer/steward should document the issue for consideration in the next round of full review. The E&M team informs the requester of the final decision with justification.

Appendix A: Quality Measure Developer and Steward Agreement

Each candidate measure or set of measures has a measure steward who assumes responsibility for the submission of the measure to Battelle for potential endorsement. The measure steward is responsible for making necessary updates to the measure and informing Battelle about any changes made to the measure. In addition, the measure steward is responsible for providing the required measure information during the measure maintenance process:

- The measure steward organization is required to identify a single point of contact who will be notified of any upcoming maintenance deadlines or requirements related to the endorsed measure(s). If the point of contact changes, then the steward should notify Battelle of the new point of contact.
- Stewards may be contacted by PQM members or other members of the public with inquiries about specifications, updates, and implementation of the endorsed measure(s).
- Stewards are also responsible for maintaining measure details and specifications on any publicly available website.

Each steward who submits a fully specified and tested measure to Battelle must submit a completed and signed Quality Measure Developer and Steward Agreement (QMDSA) on or before the project's measure submission deadline in order for the measure to be considered by the committee. The agreement is between Battelle and the measure steward and only shared between these parties.

- For new measure stewards, the QMDSA should be accompanied by the completed addendum, in which the steward must list all the measures (measure number and measure title) being submitted for review.
- For existing measure stewards, only a signed addendum is needed and will be appended to the existing QMDSA; a new QMDSA is not required. Contact E&M project staff to obtain the addendum.

Only one QMDSA is necessary per measure steward. If the steward is a governmental organization, a QMDSA is not required. For more information about how to complete the QMDSA, please see the QMDSA Submission Instructions.

Battelle will work with all measure stewards to transition to this QMDSA. Those who have measures up for maintenance or wish to add additional measures to their current QMDSA will need to complete an Additional and Maintenance Measures Form. Each QMDSA will remain effective for 5 years from the date it is signed.

The QMDSA and Additional and Maintenance Measures Forms are contractual agreements that must be signed by Battelle and any measure steward that is submitting one or more measures to be evaluated for endorsement.

Appendix B: Personal/Organizational Disclosure of Interest Form

NOTE: *You will be asked to complete this form annually. Please denote the year this disclosure will cover (ex. 2023).*

1. **Name:**

   Organization Affiliation:

   Committee Name:

   Year:

2. Describe any personal or organizational relationships subject to disclosure. If none, check here: ☐

3. Describe any personal or organizational financial interests subject to disclosure. If none, check here: ☐

4. Electronic Certification
   By executing this Electronic Certification, I certify that I have reviewed the Personal/Organizational Disclosure of Interest Form, and the information given above is true to the best of my knowledge.

Name: _____ **Signature:** _____

Date: _____

All persons and organizations must be free of any financial conflicts of interest for this effort. If at any time you believe that a potential or actual conflict exists, you must notify Battelle immediately. “Conflict of Interest” means: because of other activities or relationships with other persons or organizations you are unable or potentially unable to (1) render impartial assistance or advice; (2) execute your duties as a committee member due to the impairment of or the possibility of the impairment of your objectivity; (3) engage in this effort because you have or might acquire an unfair competitive advantage.

Appendix C: Measure Disclosure of Interest Form

1. **Name:**

   Organization Affiliation:

   Committee Name:

   Cycle (ex. Fall 2025):

2. Describe any personal or organizational measure conflicts. If none, check here: ☐
   a. Measures under review:

CBE #	Measure Title	Measure Developer/Steward
####	[insert title]	[insert developer and steward]
####	[insert title]	[insert developer and steward]
####	[insert title]	[insert developer and steward]

   i. If you have worked as an employee, collaborator, or consultant of the measure developers/stewards listed *OR* contributed to the development of the measures listed, in any capacity, in the past five (5) years, check here: ☐

   b. Competing measures:

CBE #	Measure Title	Measure Developer/Steward
####	[insert title]	[insert developer and steward]
####	[insert title]	[insert developer and steward]
####	[insert title]	[insert developer and steward]

   i. If you have worked as an employee, collaborator or consultant of the measure developers/stewards listed *OR* contributed to the development of the measures listed, in any capacity, in the past five (5) years, check here: ☐

   c. If you checked either box under 2a. or 2b. above, please provide a detailed description of the involvement. (Include CBE ID number, Measure Title, Cycle, and Steward Name:)

(continued on next page)

3. Electronic Certification

   By executing this Electronic Certification, I certify that I have reviewed the Measure Disclosure of Interest Form, and the information given above is true to the best of my knowledge.

Name: _____ **Signature:** _____

Date: _____

All persons and organizations must be free of any conflicts of interest for measures under this effort. If at any time you believe that a potential or actual conflict exists, you must notify Battelle immediately. "Conflict of Interest" means because of other activities or relationships with other persons or organizations you are unable or potentially unable to (1) render impartial assistance or advice; (2) perform due to the impairment of or the possibility of the impairment of your objectivity; or (3) perform because you have or might acquire an unfair competitive advantage.

Appendix D: PQM Measure Evaluation Rubric

Rubric items correspond to items in the measure submission form and provide the information needed to evaluate each of the five rubric domains.

The requirements for initial and maintenance measure endorsement are indicated as, "*[For initial endorsement]*" or "*[For maintenance],*" within each domain of PQM Measure Evaluation Rubric. If neither distinction is listed for a rubric requirement, then the requirement applies to both initial and maintenance endorsement.

The PQM Measure Evaluation Rubric does not include must-pass criteria nor algorithms for assigning a rating; rather, it guides reviewers to a rating of "Met," "Not Met but Addressable," or "Not Met" based on the criteria listed for each. As part of its continuous quality improvement of the E&M process, Battelle considers whether changes to the domains, criteria, and/or additional guidance, such as an algorithm, are needed.

Note on instrument-based clinical quality measures: Instrument-based clinical quality measures are measures that are derived from instruments or surveys. As a CBE, Battelle does not review or endorse instruments or surveys. Rather, the CBE reviews and endorses clinical quality measures derived from instruments or surveys. There are no differences in the requirements or criteria for endorsement & maintenance between instrument-based clinical quality measures and other clinical quality measures. Each clinical quality measure derived from an instrument or survey is reviewed and endorsed separately. Measure developers/stewards are encouraged, where appropriate, to combine individual instrument or survey items into a person/respondent-level "composite," which may then be aggregated to the accountable entity level. Such a measure would be reviewed and endorsed as a single measure. For more information, please see our CBE Policy on Instrument-based Clinical Quality Measures (Appendix G).

1. Importance

***Description**: Extent to which the measure is important for making significant gains in health care quality or cost where there is variation in or overall less-than-optimal performance. The measure focus is associated with a material outcome.*

Importance Items
Attach a logic model depicting the relationship between structures and processes and the desired outcome.
Summarize evidence of measure importance from the literature linking the structure/process/intermediate outcome to the outcome.
[For initial endorsement] If implemented, what is the measure's anticipated impact on important outcomes?
[For maintenance] Provide evidence of performance gap or measurement gap by providing performance scores on the measure as specified (current and over time) at the specified level of analysis.

Commented [HSAG6]: Please consider requiring this criterion for both initial endorsement and maintenance. If the developer has collected data to test reliability and validity, they should be able to provide at least some data on the estimated performance gap, even if it is only from a few measured entities.

Importance Items
Explain why existing measures/quality improvement programs are insufficient for addressing this health care need.
Provide evidence the target population (e.g., patients) values the measured outcome, process, or structure, and finds it meaningful. Describe how and from whom you obtained input.

Not Met:

- Evidence is about something other than what is measured; OR
- Empirical evidence submitted is without literature review or grading; OR
- Empirical evidence includes only selected studies from the literature review;[4] OR
- Evidence is not graded high quality or strong recommendation; OR
- Literature review conclusion is that consistency is low or controversial; moderate/high certainty that the net benefit (i.e., improved outcomes, adverse events, and/or costs avoided due to the measure's anticipated impact) is null or small or grade of weak; OR
- There is low confidence/certainty that the business case[5] is adequate (the anticipated impacts of the measure on patient outcomes and/or costs/resource use justify the measure and its use), where "adequate" means there is a net benefit to measurement; OR
- *[For initial endorsement]*
    - There is no description of other existing measures or programs or no search conducted to identify other existing measures or programs; OR
- *[For maintenance]*
    - There is low confidence/certainty that there is evidence of a performance gap, as determined by variation in performance or less-than-optimal performance for the overall target population and/or subpopulations; OR
- Proposed measure has the same measure focus and target population as existing measure(s) and offers no advantage in terms of addressing disparities, feasibility, potential use, or scientific acceptability; OR

[4] A literature review could include a systematic review, clinical practice guidelines, observational studies, case studies, etc. The purpose of the literature review is to identify relevant studies to support the measure's logic model. Developer/stewards should provide a summary of the evidence for the committee's consideration. An evaluation of the quality of evidence should also be conducted. Often clinical practices guidelines conduct systematic reviews. If a literature review is not possible, the committee a would consider a rationale as to why.

[5] For more information on how to consider the business case for a measure, please refer to the business case development page of the CMS website.

- Patient input does not support the conclusion that the measured outcome, process, or structure is meaningful or it does so with a low degree of certainty.

Not Met but Addressable:

- Criterion is not met (see above), but the reviewer can identify changes to specifications that may strengthen the measure's importance such that the criterion could be met.

Met:

- Literature review concludes with at least moderate certainty that a net benefit (i.e., improved outcomes, adverse events, and/or costs avoided due to the measure's anticipated impact) is at least moderate; AND
- There is at least moderate confidence/certainty that the business case is adequate (i.e., the anticipated impacts of the measure on patient outcomes and/or costs/resource use justify the measure and its use), where "adequate" means there is a net benefit to measurement; AND
- *[For maintenance]*
    - There is at least moderate confidence/certainty that there is evidence of a performance gap, as determined by variation in performance or less-than-optimal performance for the overall target population and/or subpopulations; AND
- Description of existing measures or programs justifies the proposed measure's focus among the proposed measure's target population and/or the proposed measure is superior[6] to identified related or competing measures; AND
- Description of patient input supports the conclusion that the measured outcome, process, or structure is meaningful with at least moderate certainty.

2. Equity

Description: Extent to which the measure can identify differences in care for certain patient populations, which can be used to advance health equity and reduce disparities in care.

[6] Measure developers/stewards must document why the proposed measure is superior to any identified and/or competing measures and should include any literature used to support this position. For instance, clinical practice guidelines supporting the proposed measure do not support any existing measures identified; or the proposed measure's intentions vary across programs/payors, which requires the measure to be distinct from other existing measures; or the proposed measure captures a target population at higher risk such that the use of the proposed measure may close care gaps for a higher-risk population.

Equity*
Describe how this measure contributes to efforts to address inequities in health care. Provide a description of your methodology and approach to empirical testing of differences in performance scores across multiple sociocontextual variables (e.g., race, ethnicity, urbanicity/rurality, socioeconomic status (SES), gender, gender identity, sexual orientation, age). Provide an interpretation of the results, including interpretation of any identified differences and consideration of negative impact or unintended consequences on subgroups.

**The Equity domain is optional, as Battelle recognizes some measures are not designed to advance health equity. Battelle continues to explore this, but to align with national priorities, Battelle encourages developers and stewards to address this domain when possible.*

Not Met:

- Reviewer determines equity is not sufficiently assessed OR the measure does not contribute to efforts to address inequities in health care.

Not Met but Addressable:

- Criterion is not met but reviewer can identify changes to the assessment of equity OR changes to the measure specifications that would address inequities in health care.

Met:

- Reviewer determines sufficient assessment of equity was conducted (i.e., methodology provided, differences in scores tested across multiple categories, and interpretation of results) AND the measure contributes to efforts to address inequities in health care.

3. Feasibility

Description: Extent to which the measure specifications (i.e., numerator, denominator, exclusions) require data that are readily available OR could be captured without undue burden AND can be implemented for performance measurement. There is an explicit articulation of the people, processes, and technology required for data collection and reporting.

Feasibility
Describe the feasibility assessment showing you considered the people, tools, tasks, and technologies necessary to implement this measure. If an eCQM, please attach your completed eCQM Feasibility Scorecard. • *[For initial endorsement]* Describe the extent to which the required data elements: 1. Are routinely generated and used during care delivery, AND 2. Are available in electronic health records or other electronic sources or provide a credible near-term path (within 1 year) to electronic collection, AND 3. Have a data collection strategy that can be implemented. • *[For maintenance]* If changes to the measure's specifications have occurred: 1. Describe the extent to which those changes impact 1-3 above, AND

2. Describe the extent of measure implementation challenge(s)/barrier(s) that occurred as a result of the data elements and provide the mitigation strategy used or a near-term path (within 1 year) to overcome the challenge(s)/barrier(s). If no changes to the measure's specifications have occurred: 1. Describe the extent of measure implementation challenge(s)/barrier(s) that occurred as a result of the data elements and provide the mitigation strategy used or a near-term path (within 1 year) to overcome the challenge(s)/barrier(s).
Describe how the feasibility assessment informed the final measure, indicating any decisions made to adjust the measure in response to data availability.
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).

Not Met:

- Feasibility assessment not systematically conducted or described; OR

[For initial endorsement]

- No near-term path (within 1 year) is specified to support routine and electronic data capture with an implementable data collection strategy.

[For maintenance]

- No near-term path (within 1 year) is specified to overcome the challenge(s)/barrier(s) identified due to implementation of the measure's data elements.

Not Met but Addressable:

- Criterion is not met (see above), but the reviewer can identify changes to specifications that may improve feasibility such that the criterion could be met.

Met:

- Feasibility assessment systematically conducted or described; AND

[For initial endorsement]

- Near-term path (within 1 year) is specified to support routine and electronic data capture with an implementable data collection strategy; AND
- Required data are routinely generated and used during care, required data are available in EHRs or other electronic sources, and the data collection strategy can be implemented.

[For maintenance]

- No feasibility challenge(s)/barrier(s) identified due to due to implementation of the measure's data elements; OR
- Near-term path (within 1 year) is specified to overcome the challenge(s)/barrier(s) identified due to implementation of the measure's data elements.

4. Scientific Acceptability

Description: Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented.

Scientific Acceptability
Describe the data or sample used for testing (include dates, source).* • If you used a sample, describe how you selected the patients for inclusion in the sample and the representativeness of the sample. • If you used multiple data sources for different aspects of testing (e.g., reliability, validity, risk adjustment), identify how the data or sample are different for each aspect of testing. ***Note:** The measure must include data from the past 5 years. This includes data used for testing, performance gap and trend analyses, and stratification.*
Provide descriptive characteristics of measured entities included in the analysis (e.g., size, location, type). If you used a sample, describe how you selected entities for inclusion in the sample.
Identify the number and descriptive characteristics (e.g., age, sex, race, diagnosis) of the unit of analysis (e.g., patient, encounter or episode, separated by level of analysis and data source). If you used a sample, describe how you selected the patients for inclusion in the sample. If there is a minimum case count used for testing, you must reflect that minimum in the specifications.
If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), please identify how the data or sample are different for each aspect of testing.
Select the level of reliability testing conducted.± ☐ Person- or Encounter-Level (e.g., inter-abstractor reliability) ☐ Accountable Entity-Level (e.g., signal-to-noise analysis) ±***Note:** [For initial endorsement] Person- or encounter-level empirical testing is required or existing evidence (e.g., prior research, literature) is presented to support testing of all critical data elements (numerator, denominator, exclusions).* *[For maintenance] Accountable entity-level empirical testing is required.*
For each level of reliability testing conducted, describe the method of reliability testing and what it tests.
Provide the statistical results from each level of reliability testing conducted and at the measure's level of analysis (e.g., clinician, health plan, facility).

Scientific Acceptability
Provide your interpretation of the results in terms of demonstrating reliability (i.e., How do the results support an inference of reliability for the measure?).
Select the level of validity testing conducted.° ☐ Person- or Encounter-Level (e.g., sensitivity and specificity) ☐ Accountable Entity-Level (e.g., criterion validity) °***Note:*** *[For initial endorsement] Person- or encounter-level empirical testing is required or existing evidence (e.g., prior research, literature) is presented to support testing of all critical data elements (numerator, denominator, exclusions).* *[For maintenance] Accountable entity-level empirical testing is required. Face validity testing of the measure score (i.e., accountable entity level) is not acceptable for maintenance endorsement.*
If accountable entity-level validity testing was performed, select the type of validity testing conducted.^ ☐ Empirical validity testing (e.g., empirical testing of measure score; association and mechanism studies). ☐ Systematic assessment of face validity of performance measure score as an indicator of quality or resource use (i.e., the score is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance). ^***Note:*** *[For maintenance] Accountable entity-level empirical testing is required. Face validity testing of the measure score (i.e., accountable entity level) is not acceptable for maintenance endorsement.*
For each level of testing conducted, describe the method of validity testing and what it tests. For accountable entity-level testing, provide a narrative describing the hypothesized relationships, why examining these relationships would validate the measure, the expected direction of the correlations of those relationships, and the strength of those associations.
Provide the statistical results from each level of validity testing conducted and at the measure's level of analysis (e.g., clinician, health plan, facility).
Provide your interpretation of the results in terms of demonstrating validity (i.e., How do the results support an inference of validity for the measure? How do the results relate to the hypothesis? If the results are not what were expected, why?).
Check all methods used to address risk factors. ☐ Statistical risk model with risk factors (___Specify number of risk factors) ☐ Stratification by risk category (___Specify number of categories) ☐ Other (___Specify) ☐ No risk adjustment or stratification
Attach a conceptual model illustrating the pathway between patient risk factors (including social, functional status-related, and clinical factors), the quality of care, and the measured outcome. Explain the rationale for the model.
Provide descriptive statistics on the distribution across the measured entities of the risk variables identified in the conceptual model.

Scientific Acceptability
If using statistical risk models or stratification, provide detailed risk adjustment model and/or stratification specifications, including the method(s), risk factor data sources, and equations, as applicable. List all risk factors in your conceptual model, clearly indicating which factors were available/tested and which (if any) were retained in final model and/or stratification plan. Also include the data source, code with descriptor, and coefficient for each risk factor in the final risk adjustment model or stratification plan, as appropriate.
Detail the statistical results of the analysis used to test and select risk factors for inclusion in or exclusion from the risk model/stratification.
If using statistical risk models, provide the approach and results of calibration and discrimination testing. Describe any over- or under-prediction of the model for important subgroups.
If using statistical risk models or stratification, provide your interpretation of the results, in terms of demonstrating adequacy of controlling for differences in patient characteristics (i.e., case mix). Clearly describe the rationale for why each risk factor tested WAS or WAS NOT included in the final model/stratification specifications. Describe what the results mean, including what is normally expected in relation to the test conducted.
If an outcome or resource use measure is not risk-adjusted or stratified, provide rationale and analyses to demonstrate there is no need to control for differences in patient characteristics (i.e., case mix) to achieve fair comparisons across measured entities.

Not Met:

Sampling

- Sampling is used and sampling strategy is not determined by the measure's analytic unit; OR sample does not represent variety of entities whose performance will be measured; OR sample does not include adequate numbers of units of measurement for the selected statistical method; OR

Reliability

- For Person- or Encounter-Level Reliability[7]
    - Internal consistency < 0.7; OR
    - Inter-rater agreement < 0.4; OR
    - Test-retest reliability (Intraclass correlation or Pearson correlation) < 0.5; OR
    - Linear relationship < 0.6; OR
    - *[For initial endorsement]*

[7] Reliability thresholds were established by the Scientific Methods Panel and confirmed at the June 14, 2022, advisory meeting.

        - No person- or encounter-level reliability testing or existing evidence provided of all critical data elements (numerator, denominator, exclusions).

- For Accountable Entity-Level Reliability[7,8]
    - Signal to noise/inter-unit reliability < 0.6; OR
    - Split-half reliability (ICC) < 0.6; OR
    - *[For maintenance]*
        - No accountable entity-level reliability testing provided.

Validity

- For Person- or Encounter-Level Validity
    - Reviewer determines the methodology to assess validity is inadequate/inappropriate;[9] OR the analytic approach is inadequate/inappropriate; OR
    - Evidence of validity testing for all critical data elements (numerator, denominator, exclusions) was not provided; OR
    - Reviewer disagrees with the assertion that all the measure's critical data elements are valid with limited or no threats to validity present.
- For Accountable Entity-Level Validity
    - Reviewer determines the methodology to assess validity is inadequate/inappropriate; OR the analytic approach is inadequate/inappropriate; OR
    - Reviewer disagrees with the assertion that the measure can distinguish quality with limited or no threats to validity present; OR
    - *[For initial endorsement]*
        - Face validity is inadequate.[10]

Commented [HSAG7]: Please consider clarifying whether for maintenance of endorsement if an exception will still be allowed if there is justification for not being able to assess empiric validity (e.g., no data on conceptually related measures). The prior CBE criteria stated that face validity may be accepted if the justification for why empirical testing was not conducted was acceptable to the reviewers.

Commented [HSAG8]: Please consider clarifying the acceptable threshold for face validity. We recommend Battelle consider a threshold of agreement of greater than or equal to 60%.

[8] For accountable entity-level reliability testing, the associated thresholds apply to the accountable entity (e.g., facility, clinician, health plan), not the mean or median across all entities.

[9] As part of the validity testing methodology, developers/stewards should empirically assess, as appropriate, the impact of missing data and/or measure exclusions.

[10] Face validity is accomplished through a systematic and transparent process in which developers/stewards disclose identified relevant experts (e.g., clinicians, accountable entity representatives, those with lived experience [patient, caregivers]) 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.

- *[For maintenance]*
    - Face validity is the only type of validity discussed and the measure is undergoing maintenance review.

Risk Adjustment

- Factors in the risk model do not influence the measured outcome; OR are not present at the start of care; OR the risk model includes factors that are associated with differences or inequities in care without sufficient rationale based on the conceptual model; OR
- Analysis does not demonstrate:
    - Variation in prevalence of risk factors across measure entities; AND
    - Contribution to unique variation in the outcome; AND
    - Impact of risk adjustment for providers at high or low extremes of risk; OR
    - Results do not demonstrate acceptable model performance.

Not Met but Addressable:

- Criterion is not met but the reviewer can identify:
    - Improvements to the sampling methodology; OR
    - Changes to the methodology/analytic approach that could improve assessment of reliability; OR
    - Changes to the methodology/analytic approach that could improve assessment of validity; OR
    - Changes to the specifications that could improve validity and/or address threats to validity; OR
    - Changes to the risk model that could improve model appropriateness or performance.

Met:

Sampling

- If a sample is used, the sampling strategy is determined by the measure's analytic unit; AND sample represents the variety of entities whose performance will be measured; AND sample includes adequate numbers of units of measurement for the selected statistical method; AND

Reliability

- For Person- or Encounter-Level Reliability[7]
    - Internal consistency $\geq$ 0.7; OR

- Inter-rater agreement ≥ 0.4; OR
- Test-retest reliability (ICC or Pearson correlation) ≥ 0.5; OR
- Linear relationship ≥ 0.6; AND
- *[For initial endorsement]*
  - Person- or encounter-level reliability testing or existing evidence provided of all critical data elements (numerator, denominator, exclusions).

- For Accountable Entity-Level Reliability[7,8]
  - Signal to noise/inter-unit reliability ≥ 0.6; OR
  - Split-half reliability (ICC) ≥ 0.6; AND
  - *[For maintenance]*
    - Accountable entity-level reliability testing provided.

Validity

- For Person- or Encounter-Level Validity
  - Reviewer determines methodology employed[Error! Bookmark not defined.] and the analytic approach presented are appropriate and thorough; AND
  - Reviewer determines results of empirical testing or prior evidence adequately demonstrates that all critical data elements (numerator, denominator, exclusions) are valid with limited or no threats to validity present.
- For Accountable Entity-Level Validity
  - Reviewer determines methodology employed[Error! Bookmark not defined.] is adequate a nd the analytic approach presented is appropriate and thorough; AND
  - Reviewer determines results of empirical testing adequately demonstrate that the measure is valid with limited to no threats to validity; AND
  - Reviewer determines the interpretation of the empirical results supports an inference of validity; AND
  - *[For initial endorsement]*
    - Face validity is adequate[Error! Bookmark not defined.] and the measure is u ndergoing initial review.

Risk Adjustment

- Factors in the risk model influence the measured outcome; AND are present at the start of care; AND the risk model does not include factors that are associated with differences or inequities in care unless justification is provided based on the conceptual model; AND

- Analysis demonstrates:
    - Variation in prevalence of risk factors across measured entities; AND
    - Contribution to unique variation in the outcome; AND
    - Impact of risk adjustment for providers at high or low extremes of risk; AND
    - Results demonstrate acceptable model performance.

5. Use and Usability

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

Use and Usability
[For initial endorsement] Check all current or planned uses and provide the name of the program and sponsor, URL, purpose, geographic area, and percentage of accountable entities and patients included, and level of analysis and care setting. ☐ Public Reporting ☐ Public Health/Disease Surveillance ☐ Payment Program ☐ Regulatory and Accreditation Programs ☐ Professional Certification or Recognition Program ☐ Quality Improvement with Benchmarking (external benchmarking to multiple organizations) ☐ Quality Improvement (internal to the specific organization) ☐ Other (___Specify)
[For maintenance review] Check all current uses and provide the name of the program and sponsor, URL, purpose, geographic area, and percentage of accountable entities and patients included, and level of analysis and care setting. ☐ Public Reporting ☐ Public Health/Disease Surveillance ☐ Payment Program ☐ Regulatory and Accreditation Programs ☐ Professional Certification or Recognition Program ☐ Quality Improvement with Benchmarking (external benchmarking to multiple organizations) ☐ Quality Improvement (internal to the specific organization)

Use and Usability
☐ Other (___Specify) ☐ Not in use
What are the actions measured entities can take to improve performance on this measure? How difficult are those actions to achieve?
[For maintenance only] Summarize the feedback on measure performance and implementation from the measured entities and others. Describe how you obtained feedback.
[For maintenance only] Describe how you considered the feedback when developing or revising the measure specifications or implementation, including whether you modified the measure and why or why not.
[For maintenance only] Discuss any progress on improvement (trends in performance results, including performance among sub-populations, if available, number and percentage of people receiving high-quality health care, geographic area, number and percentage of accountable entities and patients included). If use of the measure demonstrated no improvement, provide an explanation.

Not Met:

[For initial endorsement]

- There is no plan for use in at least one accountability application after initial endorsement but before the measure's first maintenance review; OR
- Performance scores do not yield actionable information that can be used to improve performance among measured entities.

[For maintenance]

- The measure is not currently in use in at least one accountability application[11] and has no short-term plan (i.e., within 1 year) for such use; OR
- Performance scores do not yield actionable information that can be used to improve performance among measured entities; OR
- Reviewer determines, based on the information provided regarding feedback on measure performance, the measure is not usable.

Not Met but Addressable:

[11] Accountability applications are uses of measure performance results about identifiable, accountable entities to make judgments and decisions because of performance. This can be as confidential reporting, reward, recognition, punishment, payment, or selection (e.g., public reporting, accreditation, performance-based payment, network inclusion/exclusion).

- Criterion is not met (see above), but the reviewer can identify changes to specifications that may strengthen the measure's ability to yield actionable information or usability.

Met:

[For initial endorsement]

- There is a plan for use in at least one accountability application after initial endorsement but before the measure's first maintenance review; AND
- Performance scores yield actionable information that can be used to improve performance among measured entities.

[For maintenance]

- The measure is currently in use in at least one accountability application; AND
- Performance scores yield actionable information that can be used to improve performance among measured entities; AND
- Reviewer determines, based on the information provided regarding feedback on measure performance, the measure is usable.

Appendix E: Guidance to Make Submissions 508 Compliant

Battelle ensures all public-facing materials are 508 compliant. Section 508 of the Rehabilitation Act of 1973 (29 U.S.C. 794d), as amended by the Workforce Investment Act of 1998 and the Architectural and Transportation Barriers Compliance Board Electronic and Information (EIT) Accessibility Standards (36 CFR part 1194), ensures those with disabilities have equal access to government information as contained on information and communications technology (ICT), and thereby to the government employment, programs, and services to which all citizens are entitled. The following steps should be taken during the measure submission process to maintain Section 508 compliance:

- Creating tables with row and column headers and proper reading order. Tables must be properly created to be Section 508 compliant. The table feature of the software must be employed, rather than using tabs and drawn lines. Row and column headers must be identified as such. Tables must be created so that reading order is left to right and top-down to be read correctly by read-aloud software. Tables should not contain any merging or splitting of cells. Table rows should not split/break across pages. Repeat the column and row headers to avoid merging cells and issues with splitting/breaking rows across pages.
- Providing alternative text (alt-text) to describe images, graphics, and exhibits that can be used by text-to-speech programs. Developers/stewards should provide alt-text of the image, chart, or graphic that provides all the necessary information for the visually impaired web user to understand the image. All the relevant information in charts, graphs, and diagrams should be included in the alt-text. Images or graphics that are added for design or layout only (i.e., add no meaning to the document) can be described briefly, as in “bullet” or “empty cell.”
- Using color appropriately. There must be enough color contrast in graphics to prevent those individuals with color vision deficiencies from having problems understanding the graphic. Color alone cannot be used to convey information or meaning.
- Creating hyperlinks using a description of the link destination rather than vague or confusing text such as “click here.”

The E&M team provides a checklist of 508 compliance criteria for developers/stewards to consider when submitting measures to Battelle (see below).

508 GUIDANCE CHECKLIST

These guidelines apply to all parts of your measure submission including all fields and attachments used within the measure submission forms.

Text

- ☐ Is all my text black, not using any other colors?
- ☐ Am I reserving underlined text for **hyperlinks** only and creating emphasis using *italic,* **bold,** and ***bold-italic*** text instead of using underlining?
- ☐ Am I avoiding multiple hard and soft returns?
- ☐ Are all my hyperlinks working, linked to their correct destination, and using a distinct style to set them off from regular text?
- ☐ Do all my bulleted or numbered lists use the built-in bulleting or numbering options?

Tables

- ☐ Are my tables actual tables and **not** images or screenshots of a table?
- ☐ Am I using a table creation tool or attaching a Word *Table Design Style* table?
- ☐ Am I repeating the column and row headers in individual cells to avoid merged table cells?
- ☐ Do my empty table cells contain a symbol like * with the note: **Cells intentionally left empty,* at the bottom outside of my table?
- ☐ Did I write a brief description of what the table conveys using the Table Caption option?
- ☐ Is the table converted to paragraph text if it is too long to fit all of one column on a single page and flows over to the next page?
- ☐ Does my attached Word table have *Allow row to break across pages* turned **off** for all rows and *Repeat as header row at the top of each page* turned **on** for the first row?

Images, Figures, Graphs, Charts, and Pictures

- ☐ Do my images include clear, concise alt-text descriptions of what they represent using the image caption option or Edit/Alt-text option for Word attachments?

Appendix F: Measure of Consensus

Variance is used as a metric to assess disagreement (lack of consensus). However, variance alone is insufficient when comparing different sizes of groups or groups with different means.

The measure of consensus is the complement of the index of disagreement, which is based on the variance of the responses scaled by the total available range of variance conditional on the mean response.

Table F1. Measure of Consensus

Number of respondents	Endorse	Endorse with Conditions	Not Endorse/Remove Endorsement	Measure of Consensus
40	0.000	0.250	0.750	1.00000
40	0.125	0.125	0.750	0.99429
20	0.000	0.250	0.750	1.00000
20	0.125	0.125	0.750	0.95170
40	0.125	0.750	0.125	0.99707
20	0.150	0.750	0.100	0.97065
40	0.250	0.000	0.750	0.94527
20	0.250	0.000	0.750	0.95110
40	0.500	0.000	0.500	0.81789
20	0.500	0.000	0.500	0.80713

Measure of Consensus (far-right column) = 1 minus the Index of Disagreement

Index of Disagreement (not shown) = Response variance / Total available range of variance

At 0.95000 for the Measure of Consensus (far-right column), at least 75% of respondents are in agreement (i.e., 75% of the respondents [blue shading] voted to Endorse, Endorse with Conditions, or Not Endorse/Remove Endorsement). As the response variance increases, the more disagreement there is amongst respondents, and the Measure of Consensus (far-right column) decreases. Perfect agreement would mean there is zero variance, which may be insurmountable given the differences of opinions, expertise, and/or experience of respondents. Therefore, the E&M process employs a Measure of Consensus of 95%, which corresponds to a consensus threshold of 75% amongst respondent votes.

References

M. A. Rahem and M. Darrah, “A geometric approach for computing a measure of consensus for groups,” International Mathematical Forum, vol. 11, pp. 961–973, 2016.

M. A. Rahem and M. Darrah, "Using a Computational Approach for Generalizing a Consensus Measure to Likert Scales of Any Size," International Journal of Mathematics and Mathematical Sciences, 2018.

Y. Akiyama, J. Nolan, M. Darrah, M. A. Rahem, and L. Wang, “A method for measuring consensus within groups: an index of disagreement via conditional probability,” Information Sciences, vol. 345, pp. 116–128, 2016.

Y. Tsuchiya and N. Hiramoto, “Measuring consensus and dissensus: a generalized index of disagreement using conditional probability,” Information Sciences, vol. 439/440, pp. 50–60, 2018.

Appendix G: CBE Policy on Instrument-based Clinical Quality Measures

Overview

Instrument-based clinical quality measures are measures that are derived from instruments or surveys, such as various versions of the Consumer Assessment of Healthcare Providers and Systems (CAHPS), the Hospice Outcomes and Patient Evaluation (HOPE), or End-Stage Renal Disease (ESRD) Patient Life Goals Survey (PaLS).

Policy

The following is the policy of the CBE with respect to instrument-based clinical quality measures:

- The CBE does not review or endorse instruments or surveys. Rather, the CBE reviews and endorses clinical quality measures derived from instruments or surveys.
- Clinical quality measures derived from instruments or surveys must be specified and tested at the accountable entity level (e.g. clinician or facility).
- There are no differences in the requirements or criteria for endorsement & maintenance between instrument-based clinical quality measures and other clinical quality measures. Specifically, all measures are evaluated based on data element-level (i.e., person- or encounter-level) reliability and validity, and accountability entity-level reliability and validity.
- For data element-level reliability and validity, measure developers/stewards may cite existing literature to substantiate those properties.
- Measures developers/stewards are also encouraged to attest that the instrument or survey was developed using a best practice protocol (e.g., Holmbeck, 2009).
- Each clinical quality measure derived from an instrument or survey is reviewed and endorsed separately.
- Measure developers/stewards are encouraged, where appropriate, to combine individual instrument or survey items into a person/respondent-level "composite," which may then be aggregated to the accountable entity level. Such a measure would be reviewed and endorsed as a single measure.
- E&M project staff are available for technical assistance to measure developers/stewards in the application of this policy.

Reference

G.N. Holmbeck and K.A. Devine, "Editorial: an author's checklist for measure development and validation manuscripts," Journal of Pediatric Psychology, vol. 34, pp. 691-696, 2009.


Partnership for
Quality Measurement
Powered by Battelle

Full Measure Submission

Instructions: This form can be used as a worksheet to assist you in developing your **Full Measure Submission (FMS)** for a new or maintenance measure. When you have received the approval for full measure submission, navigate to https://p4qm.org/ and log into your PQM account. Once logged in, click "My Account" to go to your dashboard, then scroll to the bottom of the page and select *Approved for Full Measure Submission* from the "Endorsement Cycle Status" drop-down list and click "Apply" to see your measures ready for FMS. To return to an FMS draft in progress, select *Full Measure Submission Draft* from the drop-down list and click "Apply". Click here for more information on the Endorsement & Maintenance measure submission process.

- You must complete all required fields (denoted by *) to submit the final FMS
- You may save a draft of the FMS form before completing all required fields
- If you would like to make changes to information submitted via the Intent to Submit (ITS), you may edit the original content in the FMS form

Required fields vary depending on whether your measure is an electronic Clinical Quality Measure (eCQM), or an initial (new) measure versus a maintenance measure, or for selected other situations. Conditional fields are indicated in this template with brackets before each field (e.g., *[If the measure is an eCQM]* **Attach MAT Output** *).

Section 1. Measure Specifications

[NOTE: Items 1.1-1.9, 1.14, and 1.15 were entered in the ITS, and can be edited in the FMS]

1.10 Measure Rationale *

Provide a rationale for why measured entities should report this measure, including how the measure will improve the quality of care for patients and/or any associated health care costs, and what are the benefits or improvements in quality envisioned by use of this measure.

1.11 Measure Webpage *

Provide a URL to a webpage, specific for this measure, containing current detailed specifications, including code lists, risk model details, and supplemental materials. Do not enter a URL to a home page or to general information. The webpage must be publicly accessible. If no URL is available, copy and paste this example: http://example.com.

1.12 *[If the measure is an eCQM]* **Attach MAT Output**

Attach the zipped output from the Measure Authoring Tool (MAT). If you did not use the MAT,

Commented [HSAG9]: Please consider aligning the fields and definitions of the Consensus-Based Entity (CBE) Measure Submission Form with corresponding fields within the Measures Under Consideration Entry/Review Information Tool (MERIT), where feasible. This will reduce measure developer burden by consolidating or streamlining the required measure information documentation across the CBE endorsement submission form, the Measures Management System (MMS) Blueprint templates, and the Centers for Medicare & Medicaid Services (CMS) MERIT Data Template.

please contact PQM Support. Use the measure specification fields (e.g., 1.14a – 1.15c) for the plain-language description of the specifications. One file only; 256 MB limit; Allowed file types: .zip.

1.13 Attach Data Dictionary

Attach a data dictionary, code table, and/or value sets (include variables in the final risk model or stratification plan, if applicable). Attachment should include variables used in the final risk model and/or stratification, if applicable.
One file only; 256 MB limit; Allowed file type: .xls; .xlsx; .csv (please clearly label sheets).

☐ **1.13a Data dictionary not attached**
I attest that all information will be provided in relevant fields where code and/or value sets are needed (e.g., 1.14a – 1.15b).

1.14a Numerator Details *

Provide details needed to calculate the numerator. All information required to identify and calculate the cases from the target population (denominator) with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets. If your list of codes with descriptors is greater than will fit in this text box, you must attach an Excel or csv file in the previous question. If the numerator includes a list (or lists) individual codes with descriptors that exceeds one page, please provide this information in an xls; .xlsx; .csv file as part of the data dictionary attachment.

Commented [HSAG10]: Please consider adding guidance for how continuous variable measures should be entered since these types of measures do not have a traditional numerator/denominator similar to proportion measures.

Please also refer to the guidance for measure specifications by measure category as outlined in the CMS MMS Blueprint related to measures specified as continuous variable: https://mmshub.cms.gov/measure-lifecycle/measure-specification/specifications-measure-category

1.15a Denominator Details *

Provide details needed to calculate the denominator. 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. If the list(s) of individual codes with descriptors exceeds one page, please provide this information in an Excel or .csv file as part of the data dictionary attachment.

1.15b Denominator Exclusions *

Briefly describe exclusions from the denominator cases, if any. Enter "None" if the measure does not have denominator exclusions.

1.15c Denominator Exclusions Details *

Provide details needed to calculate denominator exclusions. Enter "None" if the measure does not have denominator exclusions. 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. If the list(s) of codes with descriptors exceeds one page, please provide this information in an Excel or .csv file as part of the data dictionary attachment.

1.16 Type of Score *
Select the most relevant type of score.

☐ Categorical, e.g., yes/no
☐ Continuous variable, e.g., average
☐ Count
☐ Rate/proportion
☐ Composite scale
☐ Other scoring method
1.16a Describe other scoring method *

1.17 *[If Measure Type (1.5) IS NOT "Cost/Resource Use"]* **Measure Score Interpretation** *
Select the appropriate interpretation of the measure score

☐ Better quality = Higher score
☐ Better quality = Lower score
☐ Better quality = Score within a defined interval
☐ Passing score defines better quality
☐ Other
1.17a Describe Other measure score interpretation *

1.17 *[If Measure Type (1.5) IS "Cost/Resource Use"]* **Select the type of cost measure** *

☐ Per capita (population- or patient-based)
☐ Per episode
☐ Per procedure
☐ Other
1.17a Specify other cost measure *

1.18 Calculation of Measure Score *
Diagram or describe the calculation of the measure score as an ordered sequence of steps. Identify the denominator, denominator exclusions (if any), numerator, time period of data collection, risk adjustment and/or stratification, and any other calculations.

1.18a Attach measure score calculation diagram
Attach a measure score calculation diagram, if desired.
One file only; 256 MB limit; Allowed file types: .pdf; .jpg; .png.

1.19 Measure Stratification Details *
Provide all information required to stratify the measure results, if necessary. Include the stratification variables, definitions, code/value sets, and if appropriate, the risk-model covariates and coefficients for the clinically-adjusted version of the measure. If the list(s) of codes with descriptors exceeds one page, please provide this information in an Excel or .csv file as part of the data dictionary attachment. If the measure is not stratified, please state "The measure is not stratified." If the information is included within the data dictionary attachment, please state "See data dictionary attachment."

1.20 Testing Data Sources *
Select the data sources for which you have tested and specified the measure. Choose all that apply.

☐ Administrative Data
☐ Claims Data
☐ Electronic Health Records
☐ Paper Patient Medical Records
☐ Registries
☐ Standardized Patient Assessments
☐ Patient-Reported Data and/or Survey Data *[Answer questions 1.21-1.24]*
☐ Non-Medical Data
☐ Other Data Source

1.20a Specify other data source *

1.21 *[If "Patient-Reported Data and/or Survey Data" was selected above]* **Patient reported data collection tools**
Choose one (1.21a or 1.21b). If the measure requires patient-reported data to collect stratification and/or risk adjustment variables, please include this information as well.

1.21a Data Source URL(s)
Provide link to the survey, tool, questionnaire, or scale used as a data source for your measure. This must be an external URL such as http://example.com. If no URL is available, copy and paste the example: http://example.com. Click "Add Another Item" to enter multiple URLs.

1.21b Attach Data Collection Tool(s)
Attach the survey, tool, questionnaire, or scale used as a data source for your measure.
One file only; 256 MB limit; Allowed type: .zip.

1.22 *[If "Patient-Reported Data and/or Survey Data" was selected for 1.20]* **Proxy Responses** *

Are proxy responses allowed?

☐ No ☐ Yes

1.23 *[If "Patient-Reported Data and/or Survey Data" was selected for 1.20]* **Survey Respondent** *

Please indicate the respondent for your survey, tool, questionnaire, or scale. Select all that apply.

☐ Patient
☐ Family or other caregiver
☐ Clinician
☐ Other

1.23a Specify other survey respondent *

1.24 *[If "Patient-Reported Data and/or Survey Data" was selected for 1.20]* **Data Collection and Response Rate** *

For survey/patient-reported data, provide instructions for data collection (e.g., modes of collection, languages of administration), including disclosing minimum response rates and guidance on improving response rates. In addition, specify how to calculate response rates for reporting with performance measure results.

1.25 Data Sources *

Identify the specific data source(s), other than or in addition to any patient-reported data and/or survey data collection instrument(s) indicated for the measure. For example, provide the name of the database, clinical registry, etc. and describe how the data are collected. Please discuss any data feasibility, reliability, and/or validity challenges and how this has been mitigated.

1.26 Minimum Sample Size *

Indicate whether the measure has a minimum sample size to calculate the performance score and provide any instructions needed for obtaining the sample and guidance on minimal sample size.

Partnership for Quality Measurement
Powered by Battelle

Section 2. Importance

2.1 Attach Logic Model *

Attach a logic model depicting the relationship between structures and processes and the desired outcome. Briefly describe the steps between the health care structures and processes (e.g., interventions, or services) and the desired health outcome(s). Identify the relationships among the inputs and resources available to create and deliver an intervention, the activities the intervention offers, and the expected results (i.e., desired outcome). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process, or outcome being measured.
One file only; 256 MB limit; Allowed file types: .pdf; .doc; .docx.

2.2 Evidence of Measure Importance *

Summarize evidence of the measure's importance from the literature, linking the structure/process/intermediate outcome to the desired health outcome. Please provide references for supporting evidence.

2.3 *[If initial endorsement]* Anticipated Impact *

If implemented, what is the measure's anticipated impact on the desired outcomes, such as those listed in the logic model? Please cite evidence to identify adverse events and costs avoided and provide references. Describe how the benefits of the measure's impact will outweigh any potential unintended consequences.

2.4 Performance Gap

If available, provide evidence of performance gap or measurement gap by providing performance scores on the measure as specified at the specified level(s) of analysis. Please include mean, minimum, maximum, and scores by deciles by using the table below or upload an attachment. In the text field here, describe the data source, including number of measured entities, number of patients, dates of data. If a sample was used, provide characteristics of the entities included. If performance scores are unavailable for the measure, please explain.

Table 1 Performance Scores by Decile

Enter the overall mean, minimum, and maximum scores, and mean scores by decile. Enter the number of measured entities and persons/encounters/episodes overall and within each decile.

	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Performance Score													

N of Entities													
N of Persons / Encounters / Episodes													

2.5 *[If initial endorsement]* **Health Care Quality Landscape** *

Please explain why existing measures/quality improvement programs are insufficient for addressing this health care need.

2.6 Meaningfulness to Target Population *

Provide evidence the target population (e.g., patients) values the measured outcome, process, or structure, and finds it meaningful. Please describe how and from whom you obtained input.

Partnership for Quality Measurement
Powered by Battelle

Section 3. Feasibility

3.1 Feasibility Assessment *

Describe the feasibility assessment conducted showing you considered the people, tools, tasks, and technologies necessary to implement this measure. For maintenance measures, describe whether feasibility issues due to implementation might have arisen and the near-term (i.e., within one year) mitigation approaches
The feasibility assessment should address:

- *Whether all required data elements are routinely generated and used during care delivery*
- *The extent of any missing data, measure susceptibility to inaccuracies, and the ability to audit data to detect problems*
- *Estimates of the costs or burden of data collection, data entry, and analysis including the impact on clinician workflow, diagnostic thought processes, and patient-physician interaction*
- *Barriers encountered or that could be encountered in implementing the measure specifications, data abstraction, measure calculation, or performance reporting*
- *Ability to collect information without violation of patient confidentiality, including circumstances where measures based on patient surveys or the small number of patients may compromise confidentiality*
- *Identification of unintended consequences*

3.2 *[If an eCQM]* **Attach Feasibility Scorecard** *

Attach your completed feasibility scorecard; please create the scorecard using the approved template [link].
One file only; 256 MB limit; Allowed types: xlsx.

3.3 Feasibility Informed Final Measure *

Describe how the feasibility assessment informed the final measure specifications, indicating any decisions made to adjust the measure in response to feasibility assessment.

3.4 Proprietary Information *

Indicate whether your measure or any of its components are proprietary, with or without fees (choose one).

☐ Proprietary measure or components (e.g., risk model, codes), without fees

☐ Proprietary measure or components with fees
☐ Not a proprietary measure and no proprietary components

3.4a *[If any proprietary components for 3.4]* **Fees, Licensing, or Other Requirements** *

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).

Partnership for Quality Measurement
Powered by Battelle

Section 4. Scientific Acceptability

4.1 Data and Samples

4.1.1 Data Used for Testing *

Describe the data used for testing (include dates, sources).

4.1.2 Differences in Data *

If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), clearly identify which data source/sample is used for each aspect of testing, including the years of data used in each. If there are no differences to report, enter "None."

4.1.3 Characteristics of Measured Entities *

Describe characteristics of measured entities included in the analysis (e.g., number, size, location, type). If you used a sample, describe how you selected measured entities for inclusion in the sample and the representativeness of the sample.

4.1.4 Characteristics of Units of the Eligible Population *

Describe characteristics of the patients, encounters, episodes, etc., including numbers and percentages by factors such as age, sex, race, or diagnosis. Provide descriptive statistics separately by each specified level of analysis and data source. If you used a sample, describe how you selected the patients for inclusion in the sample and the representativeness of the sample. If there is a minimum case count used for testing, you must reflect that minimum in the specifications in Minimum Sample Size in Section 1.

4.2 Reliability

4.2.1 Level(s) of Reliability Testing Conducted *

Choose all that apply.

☐ Patient- or Encounter-Level (e.g., inter-abstractor reliability)
☐ Accountable Entity-Level (e.g., signal-to-noise analysis)

☐ Not applicable/reliability testing not conducted

4.2.1a *Please explain why reliability testing was not conducted*

4.2.2 *[If reliability testing was conducted]* **Method(s) of Reliability Testing** *

For each level of reliability testing conducted, describe the method(s) of reliability testing and explain what each tests. Describe the steps, do not just name a method. What type of error does it test? Provide the type of statistical analysis used. Describe proportion of missing data, how missing data was analyzed and/or excluded, and any sensitivity analysis conducted.

Note: Testing at the patient- or encounter-level requires that all critical data elements be tested (not just agreement of one final overall computation for all patients). At a minimum, the numerator, denominator, and exclusions must be assessed and reported separately. Prior evidence of reliability of data elements for the data type specified in the measure (e.g., hospital claims) can be used as evidence for those data elements. Prior evidence could include published or unpublished testing that: includes the same data elements, uses the same data type (e.g., claims, chart abstraction), and is conducted on a sample as described above (i.e., representative, adequate numbers, and randomly selected, if possible).

4.2.3 *[If reliability testing was conducted]* **Reliability Testing Results** *

Provide the statistical results from reliability testing for each level and type of reliability testing conducted. Where applicable, include results from accountable entity-level reliability testing (e.g., signal-to-noise testing) in the table below.

Table 2 *[If accountable entity-level testing was conducted, i.e., if 4.2.1 includes "Accountable Entity-Level")]* **Accountable Entity-Level Reliability Testing Results**

Enter the overall reliability, minimum, maximum, and mean reliability by decile. Enter the number of measured entities and persons/encounters/episodes overall and within each decile. If a sample, provide characteristics of the entities included. Note that the mean performance score should be the same as what was entered in the performance score table in Section 2 for this level of analysis and year(s).

	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Reliability													
Mean Performance Score													
N of Entities													
N of Persons / Encounters / Episodes													

Commented [HSAG11]: Please consider clarifying the meaning of the deciles (e.g., deciles of reliability scores across providers or deciles of provider volume). Note that some approaches to measure score reliability (e.g., random split half reliability), only result in one summary reliability statistic rather than an estimate for each provider.

4.2.4 *[If reliability testing was conducted]* **Interpretation of Reliability Results** *

Provide your interpretation of the results in terms of demonstrating reliability for each level and type of reliability testing conducted. How do the results support an inference of reliability for the measure?

4.3 Validity

4.3.1 Level(s) of Validity Testing Conducted *

Choose all that apply.

☐ Patient- or Encounter-Level (e.g., sensitivity and specificity)
☐ Accountable Entity-Level (e.g., criterion validity)
☐ Not applicable/validity testing not conducted

4.3.1a Provide a rationale for why validity testing is not applicable/was not conducted

4.3.2 Type of accountable entity-level validity testing conducted *

Choose all that apply.

☐ Empirical validity testing at the accountable entity-level (e.g., criterion validity, construct validity, known groups analysis)
☐ Systematic assessment of face validity of the measure's performance score as an indicator of quality or resource use (i.e., the score is an accurate reflection of the effect of performance on quality or resource use and can distinguish good from poor performance).
☐ Not applicable/accountable entity-level validity testing not conducted

4.3.2a *[If a maintenance measure]* **Provide a rationale for why accountable entity-level validity testing was not conducted**

4.3.3 *[If validity testing was conducted]* **Method(s) of Validity Testing** *

For each level of testing conducted, describe the method(s) of validity testing and what each tests. Describe the steps (do not just name a method) and explain what was tested (e.g., accuracy of data elements compared with authoritative source, relationship to another measure as expected). What statistical analysis did you use? Describe proportion of missing data, how missing data was analyzed and/or excluded, and any sensitivity analysis conducted.

Note: Testing at the patient- or encounter-level requires that all critical data elements be tested (not just agreement of one final overall computation for all patients). At a minimum, the numerator, denominator, and exclusions must be assessed and reported separately. For patient- or encounter-level testing, prior evidence of validity of data elements for the data type specified in the measure (e.g., hospital claims) can be used as evidence for those data

elements. Prior evidence could include published or unpublished testing that: includes the same data elements, uses the same data type (e.g., claims, chart abstraction), and is conducted on a sample as described above (i.e., representative, adequate numbers, and randomly selected, if possible).

For empirical accountable entity-level testing, the following should be included:

- *Narrative describing the hypothesized relationships*
- *Narrative describing why examining these relationships (e.g., correlating measures) would validate the measure*
- *Expected direction of the association*
- *Expected strength of the association*

4.3.4 *[If validity testing was conducted]* **Validity Testing Results** *

Provide the statistical results from validity testing for each level and type of validity testing conducted.

4.3.5 *[If validity testing was conducted]* **Interpretation of Validity Results** *

Provide your interpretation of the results in terms of demonstrating validity for each level and type of validity testing conducted. How do the results support an inference of validity for the measure? For accountable entity-level testing, discuss how the results relate to the hypothesis? If the results are not what were expected, why?

4.4 Risk Adjustment

4.4.1 Methods Used to Address Risk Factors *

*What methods or approaches were used to explore the effects of risk factors on this measure? (**Note:** If you tested for the effects of risk factors and ultimately determined that risk adjustment or stratification was not warranted, please select the method(s) used and provide details of the testing and your rationale in 4.4.2 through 4.4.6; the measure's ultimate status will be reported in 4.4.7).*
Choose all that apply.

☐ Statistical risk adjustment model with risk factors
☐ Stratification by risk factor category
☐ Other

4.4.1a Describe other method(s) used

☐ No risk adjustment or stratification.

4.4.1b *[If Measure Type is outcome or cost/resource]*
Provide a rationale for why there is no need to address differences in patient characteristics (i.e., case mix) to achieve fair comparisons across measured entities for your outcome or resource measure.

4.4.2. *[If risk factors are addressed by any method (4.4.1)]* **Conceptual Model Rationale** *

Explain the rationale for the risk approach, including reasons for risk adjustment and/or stratification. Describe the sources that inform the conceptual model, e.g., scientific literature, unpublished findings, TEP. Consider age, gender, race, ethnicity, urbanicity/rurality, Medicare/Medicaid dual eligibility status, indices of social vulnerability (e.g., Centers for Disease Control and Prevention Social Vulnerability Index), and markers of functional status-related risk (e.g., cognitive or physical function) in the conceptual model, using evidence to support the model, with references. If risk factors (e.g., social, functional status-related, clinical) are included in the conceptual model but data are not available for all factors, describe any potential bias, as a result of not including the risk factor(s) in the final risk adjustment model or stratification. Address the validity of the measure in light of this bias.

4.4.2a *[If risk factors are addressed by any method (4.4.1)]* **Attach Conceptual Model** *

Attach a figure of the conceptual model that illustrates the hypothesized pathway between the social and/or functional status-related risk factors, patient clinical factors, quality of care, and the measured outcome.
One file only; 256 MB limit; Allowed types: .pdf, .jpg, .png, .zip

4.4.3 *[If risk factors are addressed by any method (4.4.1)]* **Risk Factor Characteristics Across Measured Entities** *

Provide descriptive statistics showing how the risk variables identified from the conceptual model are distributed across the measured entities. Indicate which risk factors were tested in the risk adjustment model and which were tested for stratifying the measure, as applicable.

4.4.4 *[If risk factors are addressed by any method (4.4.1)]* **Risk Adjustment Modeling and/or Stratification Results** *

Describe the statistical results of the analyses used to test and select risk factors for inclusion in or exclusion from the risk model and/or stratification, as applicable. Clearly indicate the risk factors included in the final risk model and/or used in the final stratification approach.

4.4.4a *[If risk factors are addressed by any method (4.4.1)]* **Attach Risk Adjustment**

Modeling and/or Stratification Specifications *

Provide detailed risk adjustment model and/or stratification specifications, including the method(s), risk factor data sources, and equations, as applicable Please list all risk factors in your conceptual model, clearly indicating which factors were available/tested and which (if any) were retained in final model and/or stratification plan. Also include the data source, code with descriptor, and coefficient for each risk factor in the final risk adjustment model or stratification plan, as appropriate.
One file only; 256 MB limit; Allowed types: .xls; .xlsx; .csv

4.4.5 *[If 4.4.1 includes "Statistical risk adjustment model with risk factors"]* **Calibration and Discrimination** *

Describe the approach and results of calibration and discrimination testing. Describe any over- or under-prediction of the model for important subgroups.

4.4.5a *[If 4.4.1 includes "Statistical risk adjustment model with risk factors"]* **Attach Calibration and Discrimination Testing Results** *

Attach results of calibration and discrimination testing.
One file only; 256 MB limit; Allowed types: .pdf; .zip

4.4.6. *[If risk factors are addressed by any method (4.4.1)]* **Interpretation of Risk Factor Findings** *

Provide your interpretation of the results, in terms of demonstrating adequacy of controlling for differences in patient characteristics (i.e., case mix). Clearly describe the rationale for why each risk factor tested WAS or WAS NOT included in the final model. Describe what the results mean, including what is normally expected in relation to the test conducted.

4.4.7 *[If risk factors are addressed by any method (4.4.1)]* **Final Approach to Address Risk Factors** *

*After testing, what methods or approaches were ultimately used to control for the effects of risk factors? (**Note:** the final approach should be supported by the testing and the rationale provided in 4.4.2-4.4.6). Choose all that apply.*

☐ Statistical risk adjustment model with risk factors
☐ Stratification by risk factor category
☐ Other

4.4.1a Describe other method(s) used

☐ No risk adjustment or stratification.

Partnership for Quality Measurement
Powered by Battelle

Section 5. Equity

5.1 Contributions Towards Advancing Health Equity (*optional*).
Describe how this measure contributes to efforts to advance health equity Provide a description of your methodology and approach to empirical testing of differences in performance scores across multiple socio-contextual variables (e.g., race, ethnicity, urbanicity/rurality, socio-economic status, gender, gender identity, sexual orientation, age). Provide an interpretation of the results, including interpretation of any identified differences and consideration of negative impact or unintended consequences on subgroups.

Partnership for Quality Measurement
Powered by Battelle

Section 6. Use & Usability

6.1 Use

6.1.1. Current Status *

Is this new or maintenance measure currently in use?

☐ No ☐ Yes

6.1.2 *[If initial endorsement]* **Current or Planned Use(s)** *

Choose all that apply

☐ Public Reporting
☐ Public Health/Disease Surveillance
☐ Payment Program
☐ Regulatory and Accreditation Programs
☐ Professional Certification or Recognition Program
☐ Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
☐ Quality Improvement (Internal to the specific organization)
☐ Other

6.1.2a Please specify other current or planned use

6.1.3 *[If maintenance review]* **Current Use(s)** *

Choose all that apply

☐ Public Reporting
☐ Public Health/Disease Surveillance
☐ Payment Program
☐ Regulatory and Accreditation Programs
☐ Professional Certification or Recognition Program
☐ Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
☐ Quality Improvement (Internal to the specific organization)
☐ Other

6.1.3a Please specify other use *

☐ Not in use

6.1.3b Provide more information as to why the measure is not in use and whether there is a near-term (within one year) plan for its use within an accountability application[1] *

[1] Accountability applications are uses of measure performance results about identifiable, accountable entities to make judgments and decisions because of performance. This can be as confidential reporting, reward, recognition, punishment, payment, or selection (e.g., public reporting, accreditation, performance-based payment, network inclusion/exclusion).

6.1.4 *[If Current Status = Yes (6.1.1)]* **Program Details** *

Please provide the following information describing the program(s) in which the measure is currently used:

Name of the program and sponsor

URL of the program

Purpose of the program

Geographic area and percentage of accountable entities and patients included

Applicable level of analysis and care setting

[To add details for another program, click "Add Measure Submission Program" button; To remove a program record entered in error, click "Remove Program" at the top right of the appropriate program details section]

6.2 Usability

6.2.1 Actions of Measured Entities to Improve Performance *

What are the actions measured entities must take to improve performance on this measure? How difficult are those actions to achieve and how can measured entities overcome those difficulties?

6.2.2 *[If maintenance review OR Current Status = Yes (6.1.1)]* **Feedback on Measure Performance** *

Summarize the feedback on measure performance and implementation from the measured entities and others. Describe how you obtained feedback.

6.2.3 *[If maintenance review OR Current Status = Yes (6.1.1)]* **Consideration of Measure Feedback** *

Describe how you considered the feedback when developing or revising the measure specifications or implementation, including whether you modified the measure and why or why not.

6.2.4 *[If maintenance review OR Current Status = Yes (6.1.1)]* **Progress on Improvement** *

Discuss any progress on improvement (trends in performance results, including performance across sub-populations if available, number and percentage of people receiving high-quality health care, geographic area, number and percentage of accountable entities and patients included). If use of the measure demonstrated no improvement, provide an explanation.

6.2.5 *[If maintenance review OR Current Status = Yes (6.1.1)]* **Unexpected Findings** *

Explain any unexpected findings (positive or negative) during implementation of this measure, including unintended impacts on patients.