Partnership for Quality Measurement
Powered by Battelle

National Consensus Development and Strategic Planning for Health Care Quality Measurement

Endorsement and Maintenance (E&M) Guidebook

October 2025

Prepared by:
Battelle
505 King Avenue, Columbus, Ohio 43201

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. *Restricted:* Use, duplication, or disclosure is subject to the restrictions as stated in Contract Number 75FCMC23C0010 between the Government and Battelle.

Table of Contents

List of Tables

List of Figures

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) 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. Additionally, it serves as a resource for E&M committee members, detailing committee term limits and expectations.

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 annually 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. Our objective is to use 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, Battelle has been a leader in the science of health care quality measurement and improvement. We are highly experienced in conducting 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. This program was 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 Sections 1890 and 1890A).

To facilitate the execution of CBE tasks, we formed the Partnership for Quality Measurement (PQM). PQM is comprised of all interested parties, 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)
- 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 committed to continuously evaluating the E&M process to meet the evolving demands of the health care landscape and incorporate feedback from interested parties. As a result, both the process and the E&M Guidebook will evolve over time. We update the E&M Guidebook annually to ensure it remains a current reference for measure developers, stewards, and E&M committee members navigating the E&M process.

Any major changes to the E&M process, policies, or evaluation criteria undergo a formal public comment period before implementation. We also provide additional educational resources (e.g., webinars, informational guides) on our PQM website as needed. Importantly, we do not apply major changes to the E&M process or measure evaluation criteria to measures currently undergoing the E&M process.

Following the Fall 2024 endorsement meetings, we identified opportunities to refine the E&M process by collecting feedback from committee members and measure developers/stewards who participated in the cycle. Feedback highlighted concerns about the length of measure review meetings—including the Public Comment Listening Session and Recommendation

Group meetings—and challenges related to the Consensus Not Reached endorsement outcome.

To address these concerns, we proposed refinements and conducted a public comment period from March 31-April 20, 2025, to gather feedback on the proposed changes to the voting procedure. The comments and our responses are available on the PQM website. These refinements took effect during the Spring 2025 cycle and have been incorporated throughout this version of the E&M Guidebook.

Battelle also made additional refinements to the E&M process that impact measure submission requirements and evaluation. These refinements were made available for public comment from June 9 through June 29, 2025. All comments received are available on the PQM website.

Changes to the Guidebook

The 2025 edition of this guidebook includes refinements to the E&M processes:

These changes took effect during the Spring 2025 cycle:

- Addressed meeting fatigue by scheduling meetings with five or more measures over 2 days instead of 1.
- Adjusted the consensus threshold to 70% in cases where there are fewer than 20 voting members, while retaining the 75% consensus threshold for groups with 20 or more voting members to maintain process integrity.
- Implemented a reconsideration process for maintenance measures that fail to reach the 75% consensus threshold but receive between 60% and 74% of votes to retain endorsement.
- Required committee members voting for "Do Not Endorse/Remove Endorsement" to provide reasons for their vote as part of refined voting approach.

Changes taking effect in the Spring 2026 cycle:

- Closing Care Gaps domain remains optional for Spring 2026 and Fall 2026.
- Adding measure-specific guidance for pediatric clinical quality measures, cost measures, entity-level reliability and validity testing, and electronic clinical quality measures.
- Establishing percentages for the number of entities needed to meet established reliability estimates for accountable entity-level testing.
- Adding expectations for conducting correlation analyses for accountable entity-level validity testing.
- Adding measure use requirement to include a description of the attributes of an accountability application in which this measure would support.

 | *The analyses upon which this publication (or document) 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. Restricted:* Use, duplication, or disclosure is subject to the restrictions as stated in Contract Number 75FCMC23C0010 between the Government and Battelle.

All measure submission requirements and PQM Measure Evaluation Rubric changes will take effect beginning with the Spring 2026 cycle. For the Spring 2025 and Fall 2025 cycle, the submission requirements and PQM Measure Evaluation Rubric in version 2.1 of the E&M Guidebook will be applied.

Battelle's Portfolio of CBE Measures

We organize measures for E&M by five topical areas. Measure developers/stewards submit measures, which are grouped by similar topics. An evaluation committee then oversees the portfolio of measures by topic (Table 1).

Table 1. E&M Projects

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 • CB3 #2372 Breast Cancer Screening • CBE #3620 Adult Immunization Status
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

 | *The analyses upon which this publication (or document) 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. Restricted:* Use, duplication, or disclosure is subject to the restrictions as stated in Contract Number 75FCMC23C0010 between the Government and Battelle.

Project Title	Areas Covered	Example Measures
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

Additional information for the five projects can be found on E&M Projects page 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 in the searchable repository database. The database is updated regularly as new and maintenance measures are submitted to Battelle for PQM endorsement review.

To use the measure submission function of STAR, measure developers/stewards must first create an account by going to the Measure Submission 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*). The Full Measure Submission deadline is 1 month after the Intent to Submit 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.

 | *The analyses upon which this publication (or document) 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. Restricted:* Use, duplication, or disclosure is subject to the restrictions as stated in Contract Number 75FCMC23C0010 between the Government and Battelle.

Endorsement and Review Process

Overview of the Endorsement Process

The E&M process ensures measures submitted for endorsement are evidence based, increasing the likelihood that they are scientifically sound, and both safe and effective. This means 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 E&M process enables endorsement decision-making in as few as 6 months (from the Full Measure Submission deadline until the end of the project [i.e., through the end of the appeal proceedings]) (Figure 1). Under this E&M process, measures reach their endpoint when the E&M committees render final endorsement decisions (Table 3).

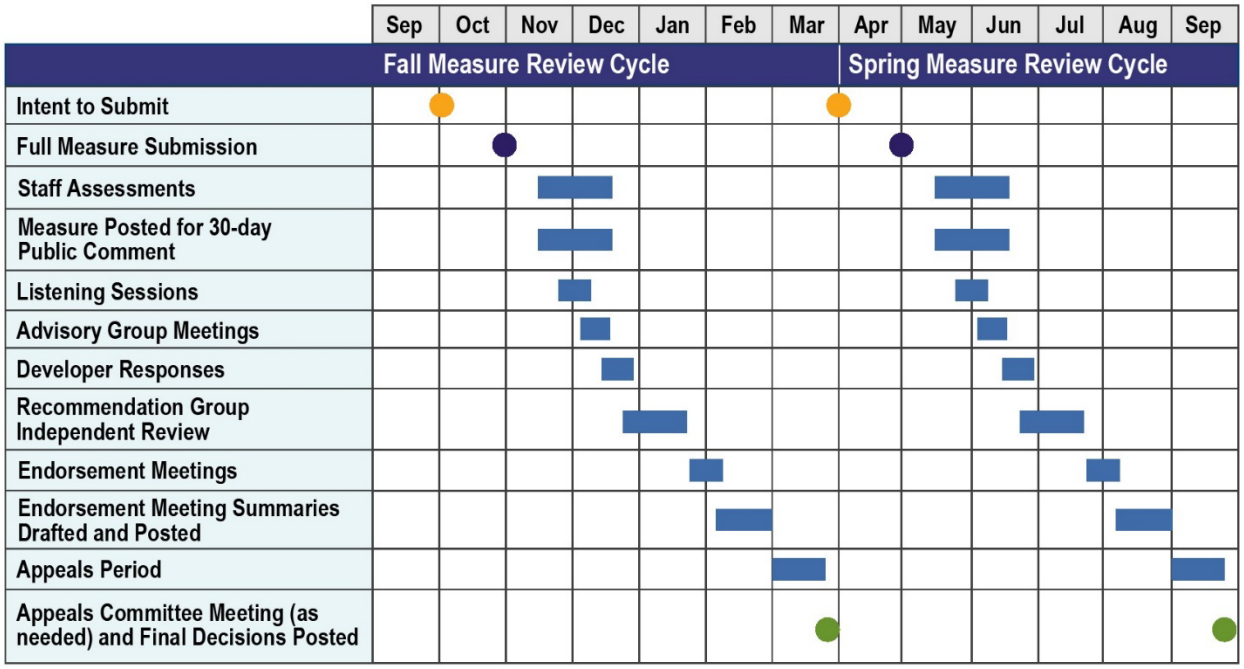


Figure 1. A 6-Month Endorsement Review Process

Table 3. Endorsement Decision Outcomes

Decision Outcome	Description	Maintenance Expectations
Endorsed	**Applies to new and maintenance measures.** The E&M committee endorses a measure if 75% or more of its active voting members agree. If the active voting body has fewer than 20 members,	Measures undergo maintenance of endorsement reviews every 5 years with a status report review at 3 years (*see Evaluations for*

 | *The analyses upon which this publication (or document) 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. Restricted:* Use, duplication, or disclosure is subject to the restrictions as stated in Contract Number 75FCMC23C0010 between the Government and Battelle.

Decision Outcome	Description	Maintenance Expectations
	the consensus required for endorsement is reduced to 70%.	*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 (*see Endorsed with Conditions for more details*). If these conditions are not addressed, the developer/steward should provide a rationale for consideration by the E&M committee review. If the active voting body has fewer than 20 members, the required consensus for endorsement is reduced to 70%.	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 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. If the active voting body has fewer than 20 members, the required consensus for not endorsing the measure is reduced to 70%.	None.
Endorsement Removed	**Applies to maintenance measures only.** Either: • The E&M committee agrees by 75% or greater to remove endorsement of the measure. If the active voting body has fewer than 20 members, the required consensus for removing endorsement is reduced to 70%; 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	None.

Decision Outcome	Description	Maintenance Expectations
	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).	

‡*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).*

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 intended to mitigate residual uncertainties/risks (e.g., barriers to expanded implementation, 1/3 of accountable entities having low reliability estimates) to the implementation or ongoing use of the measure once high uncertainties/risks (e.g., inadequate logic model or face validity, threats to validity) have been effectively addressed. Residual uncertainties/risks are aspects that are not deemed to make the measure unsafe or ineffective, so the measure is therefore endorsable. Instead, they represent areas for potential refinement and improvement in future endorsement reviews, i.e., conditions. These conditions require actions/activities that the developer/steward should undertake to address the residual uncertainties/risks prior to the next maintenance cycle. Conditions are also not what is already required of the measure during maintenance review. For example, if a new measure does not have empirical performance gap testing, this is a requirement for maintenance review and, therefore, should not be a condition for endorsement.

Conditions for endorsement may result in a measure falling into a 3- or 5-year maintenance cycle. This is determined based on the priority of the condition and what is feasible and appropriate for the developer/steward to execute by maintenance endorsement review. Table 4 lists common conditions that may be applied to a measure (i.e., within scope for residual uncertainty/risk), in accordance with the respective domains of the PQM Measure Evaluation Rubric. If a measure receives an "Endorsed with Conditions" designation during its maintenance endorsement review, the E&M committee considers whether the prior condition(s) have been addressed. If not, the committee may choose to remove endorsement unless it agrees with any rationale provided by the developer/steward.

To enhance the efficiency of meetings, E&M committee members are neither asked nor required to propose conditions for endorsement directly. Instead, they should focus on discussing the measure(s) along with the various inputs such as public comments, developer responses, other committee member viewpoints, etc. Committee members are encouraged to express their support or non-support for the measure along with the rationale for their stance.

E&M staff are responsible for listening to these discussions. Before the endorsement voting takes place during the Recommendation Group meeting, we present any proposed conditions to the E&M committee co-chairs and the committee. If the co-chairs agree with the proposed conditions based on the discussions, these conditions are included in the endorsement vote. If the co-chairs do not agree, the co-chairs may propose within-scope modifications (i.e.,

 The analyses upon which this publication (or document) 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. Restricted: Use, duplication, or disclosure is subject to the restrictions as stated in Contract Number 75FCMC23C0010 between the Government and Battelle.

the co-chairs do not agree, the co-chairs may propose within-scope modifications (i.e., modifying conditions to address residual uncertainty/risk, not high areas of uncertainty/risk) (Table 4).

When voting, a Recommendation Group member can choose to:

- **Vote to endorse the measure** if they determine it is safe and effective, and conditions are not needed.
- **Vote to endorse the measure with proposed conditions** if they find the measure safe and effective but believe conditions would further refine the measure based on discussions during the endorsement meeting.
- **Vote not to endorse or to remove endorsement** if they assess that the measure is not safe and effective and specify the high-level uncertainties or risks that have not been adequately addressed.

Table 4. Common, Within-Scope Conditions for Endorsement

PQM Rubric Domain Element*	Example Issue	Risk(s)	Condition	Condition Timeline
Importance				
Performance Gap	Maintenance measure has a narrow performance gap, which may be due to limited data/testing within a population that may not be fully representative.	A narrow gap based on limited data may obscure true variation in performance, making it difficult to justify quality improvement efforts or resource investment.	*[Maintenance measure]* Expand performance gap testing to a larger, more generalizable population.	5 years
Meaningfulness to Target Population	Limited evidence from literature, focus groups, expert panels, etc., to show that the target population (e.g., patients) values the measured outcome, process, or structure and finds it meaningful for improving health and health care.	Without patient input, the measure may lack relevance or meaning to those it affects, reducing adoption by patients, providers, and stakeholders.	Conduct additional evaluation/assessment (e.g., focus groups, surveys) of meaningfulness to the patient community (e.g., patients, caregivers, advocates).	3 years
Feasibility				
Workflow Challenges	New measure has several critical data elements that are available in structured fields but	Clinical staff may resist or inconsistently adopt structured documentation	Solicit feedback from relevant entities to better understand workflow challenges (e.g., implementing,	3 years

 The analyses upon which this publication (or document) 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. Restricted: Use, duplication, or disclosure is subject to the restrictions as stated in Contract Number 75FCMC23C0010 between the Government and Battelle.

PQM Rubric Domain Element*	Example Issue	Risk(s)	Condition	Condition Timeline
	which clinical staff are not using consistently.	practices. Quality scores may not reflect actual clinical performance due to missing or unstructured data.	collecting, and/or reporting measure's data elements) and mitigation approaches.	
Electronic Health Record (EHR) Integration	Concerns about clinical burden due to a survey for an instrument-derived measure being outside of the EHR.	Clinicians may resist using the survey, citing time constraints or misalignment with clinical workflows.	Demonstrate progression on EHR integration of the survey.	5 years
Scientific Acceptability				
Reliability	Developer/steward performed accountable entity-level reliability. Most entities have reliability estimates that exceed the accepted threshold of 0.6 but around 30% of facilities are below the threshold.	Measure results may not be reliable for 30% of reporting entities, leading to potential misclassification or unfair performance comparisons.	**Accountable entity-level testing:** Consider mitigation strategies to improve measure's reliability, such as increasing the case volume, including more than 1 year of data, etc.	3 years
Validity	Developer/steward performed accountable entity-level validity but relied on data from large health systems, limiting insight into under-resourced or rural areas.	Relying on data from large health systems may not provide insight into the experiences of under-resourced or rural areas, leaving a gap in understanding and applicability of the measure across different settings.	**Accountable entity-level testing:** Expand mechanistic evaluations as more data become available post-implementation and/or expand mechanistic evaluations to entities that have resource challenges identified.	5 years
Use and Usability				
Expanded Feedback	Measure is currently used, but feedback processes may require enhancements to effectively guide potential measure adjustments.	Lack of a robust feedback process from a variety of relevant interested parties, especially under-resourced and rural areas, does not provide a comprehensive understanding of	**Measure in use:** Refine feedback approach to include additional relevant stakeholders	3 years

PQM Rubric Domain Element*	Example Issue	Risk(s)	Condition	Condition Timeline
		measure usability across different contexts.		
Actionability	Measure improvement results have decreased or plateaued year-over-year.	Implementers of the measure may highlight gaps in how entities can act upon measure results, necessitating refinement.	**Measure in use:** Assess identified implementation challenges/burden and/or understand why improvement over time has decreased or plateaued and develop appropriate mitigations (e.g., implementation guidance).	3 years
Unintended Consequences	Interested parties have voiced concern with respect to a specific unintended consequence due to the measure's use.	The concern indicates that the measure's use may result in unintended negative consequences that the developer/steward has not fully identified or mitigated. This risk highlights the need for ongoing evaluation and responsiveness to feedback.	Assess for specific unintended consequences identified by reviewing literature and engaging relevant entities and/or patients.	5 years

**The Closing Care Gaps domain will remain optional for the Spring 2026 and Fall 2026 cycles. Therefore, the Recommendation Groups cannot place conditions on measures related to the Closing Care Gaps 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.

[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

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.

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 interested-party engagement and promotes consistent application of evaluation criteria (Figure 3).

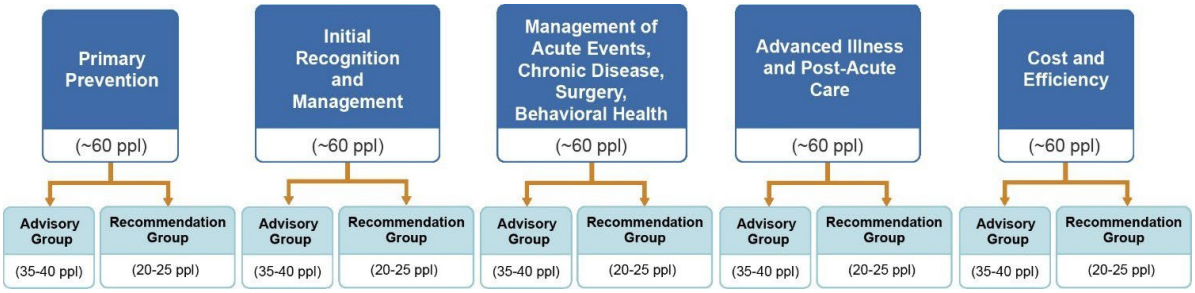


Figure 2. Recommendation and Advisory Group Structure

Advisory (Delphi) Group: The Advisory Group consists of 35-40 people. Members in this group review measures and convene 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.

Each E&M project committee has two co-chairs, who participate in the Recommendation Group's endorsement discussions, attend the Advisory Group meetings, and present a summary of the Advisory Group's deliberations at the Recommendation Group meeting. 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 Recommendation

 The analyses upon which this publication (or document) 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. Restricted: Use, duplication, or disclosure is subject to the restrictions as stated in Contract Number 75FCMC23C0010 between the Government and Battelle.

Group considers the Advisory Group's concerns and perspectives during the endorsement meeting. In addition, the co-chairs' responsibilities are to:

- Co-facilitate endorsement meetings, along with E&M staff.
- Work with E&M staff to achieve the goals of the project.
- Assist E&M 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 5). 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*).

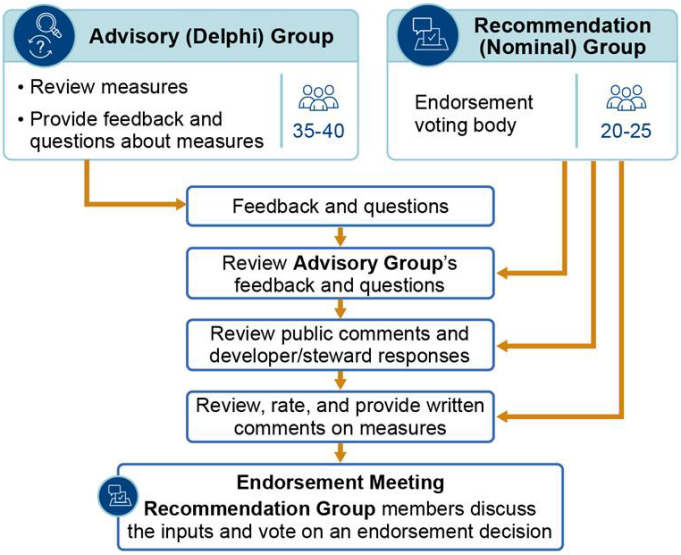


Figure 3. Advisory and Recommendation Group Roles

Individuals with specialized expertise may be added to Recommendation Group membership on an as-needed basis; we determine what specialized expertise may be necessary 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 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, Battelle and our partner, the Institute for Healthcare Improvement (IHI), recruit SMEs from our combined networks; we vet the selection 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 to invite SMEs from their technical expert panels to participate and answer committee questions during endorsement meetings.

 The analyses upon which this publication (or document) 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. Restricted: Use, duplication, or disclosure is subject to the restrictions as stated in Contract Number 75FCMC23C0010 between the Government and Battelle.

Table 5. 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
Population health experts	4	4
Researchers (e.g., health services and alternative payment models)	6	2
Other interested parties (representatives of EHR vendors, provider and facility associations, and experts in areas such as quality improvement/implementation science, care coordination, patient safety, behavioral health, and national policymakers)	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 (Figure 4). 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 assigns 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.

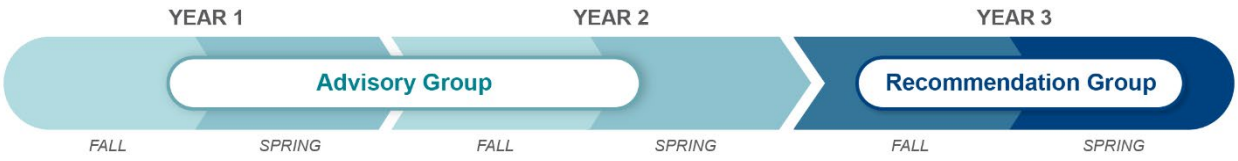


Figure 4. Advisory and Recommendation Group Assignment

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 an internal review and adjustment of committee membership, a call for nominations, and targeted outreach. Beginning in late spring, we publish a call for nominations on the PQM website and send an announcement 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 (membership is free), and they must complete a nomination form and a Personal/Organizational Disclosure of Interest (DOI) form (*Appendix B*). Before finalizing the appointments, Battelle publishes a draft roster of nominees for public comment for transparency and to garner 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 5.

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. The E&M project team uses this information to determine whether the individual's committee membership may continue or if they will be granted inactive status for a cycle.

Inactive Status and Early Termination

We understand the 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 contains questions relevant to the specific measure(s) being reviewed. Battelle provides the Measure Disclosure of Interest Form to committees at the start of each cycle. The form contains 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 disclose relevant interests at a public committee meeting in spoken fashion. The disclosure usually occurs at a committee's endorsement meeting.

Battelle staff will lead this disclosure and instruct committee members regarding information that should be disclosed. Following spoken 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 5). 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 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 is made public.

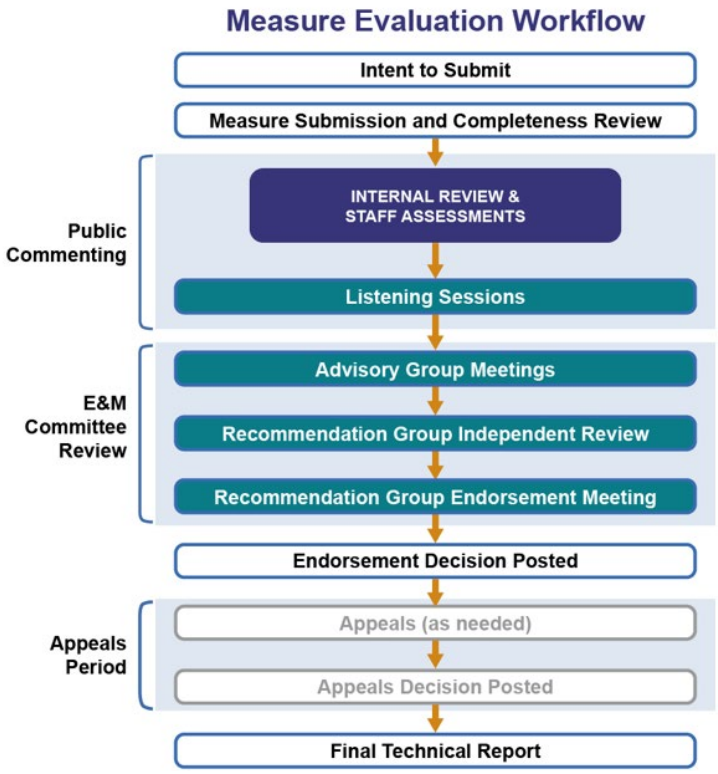
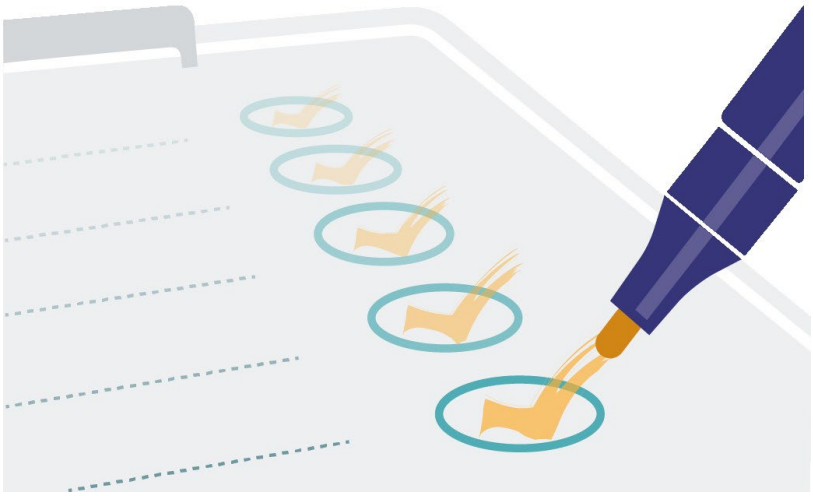


Figure 5. Measure Evaluation Workflow, 6-Month E&M Cycle

 The analyses upon which this publication (or document) 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. Restricted: Use, duplication, or disclosure is subject to the restrictions as stated in Contract Number 75FCMC23C0010 between the Government and Battelle.

Requirements for Measure Consideration

Measures must meet several requirements prior to being considered and evaluated for endorsement. 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.



- The measure developer/stewards signs a Quality Measure Developer and Steward Agreement (QMDSA) form (*Appendix A*), allowing Battelle to publicize the measure, including any proprietary information associated with the measure. If a QMDSA has been previously established, the measure developer/stewards sign an Additional and Maintenance Measures Form (*Appendix A*) listing the new measure(s).
- The measure must include data from any year(s) within 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.

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:

[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).

- Intended measure review cycle
- Measure title
- Measure structure (e.g., single, composite, instrument-derived measure [IDM] set)
- Measure description
- Measure type (e.g., structure, process, outcome)
- Measure specifications (e.g., numerator, denominator, level of analysis, care setting)
- Measure use (e.g. current use/intended use, program details)
- 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 and provide feedback to measure developers/stewards of any issues identified 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. We work iteratively with the developers/stewards to address items discovered during completeness checks (e.g., missing attachments, citations). 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 staff (see *E&M Staff Preliminary Assessment* for more details). Measures that do not pass the completeness check are pulled from consideration.

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 QMDSA or Additional and Maintenance Measures Form is completed and signed (*Appendix A*).

- ☐ Battelle has received complete and adequate responses 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*).
- ☐ For IDM sets, the instrument and each associated IDM has a separate submission.
- ☐ ICD-10 (International Classification of Diseases) codes are used and included, if applicable.

E&M Staff Preliminary Assessment

We review each measure submission using the PQM Measure Evaluation Rubric (*Appendix D*), which includes five domains (Importance, Closing Care Gaps, Feasibility, Scientific Acceptability [i.e., Reliability and Validity], 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 E&M staff preliminary assessments with developers/stewards for a factual review prior to sending them to the Recommendation Group for endorsement consideration. Developers/stewards are asked to conduct a factual review by the requested deadline, which is no more than 5 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, have we accurately summarized the testing results of the measure? 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 E&M 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 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 Session

Two to three weeks prior to the close of the public comment period, we host a Public Comment Listening Session where any interested party can register to attend this virtual session to give a brief spoken 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 the listening session to the respective measure page, and developers/stewards are provided an opportunity to draft written responses via the website. A summary of the public comments is included in endorsement meeting materials, which we make publicly available and share with the Recommendation Group no later than 1 week prior to the endorsement meetings. The Recommendation Group is tasked with reviewing the comments and developer/steward responses for endorsement decision-making.

Advisory Group Meetings

Following the listening session, we convene project-specific 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 in a spoken fashion 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 ask questions and provide feedback regarding the strengths and limitations of the measures during their meeting. 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.

Developers/stewards have 2 business days after the meeting to provide any additional written information, which will be included in the feedback summary. No less than 1 week prior to the endorsement meetings, we share the Advisory Group feedback and questions, along with the

 The analyses upon which this publication (or document) 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. Restricted: Use, duplication, or disclosure is subject to the restrictions as stated in Contract Number 75FCMC23C0010 between the Government and Battelle.

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 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 staff preliminary assessments.

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 virtual and 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 staff and committee co-chairs focus the Recommendation Group discussions on the identified strengths and limitations of the measure(s) under review by prioritizing the findings from the public comments received, the Advisory Group meetings, and the associated developer/steward responses. These facilitated discussions also take into consideration the E&M staff preliminary assessments and results.

After the discussions conclude for a measure, the Battelle facilitator summarizes the Recommendation Group's deliberations before moving to an endorsement vote. Within this summary, the facilitator draws 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." Committee co-chairs and the measure developer/steward are also given an opportunity to share final thoughts.

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

 The analyses upon which this publication (or document) 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. Restricted: Use, duplication, or disclosure is subject to the restrictions as stated in Contract Number 75FCMC23C0010 between the Government and Battelle.

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 for groups with 20 or more voting members. The consensus threshold is lowered to 70% in cases where there are fewer than 20 voting members.[3] Battelle established the consensus threshold 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 assesses 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. 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 6 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 6). 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).

Maintenance measures that fail to reach the 75% consensus threshold but receive between 60% and 74% of votes to retain endorsement (i.e., endorse and/or endorse with conditions) are reconsidered at the end of the endorsement meeting (Scenario 4a or 4b). Battelle staff will lead this process, which includes a focused discussion aimed at resolving areas of disagreement and a subsequent revote to determine whether the consensus threshold is met. If the measure still does not reach the 75% consensus threshold at this stage, the endorsement is removed.

This reconsideration approach only applies to maintenance measures, as the committee has access to a broader base of evidence for maintenance measures—such as information on performance gap, accountable entity-level reliability, and measure use (e.g., measured entity

[3] Adjusting the threshold to 70% upholds the integrity of 0.95000 for the Measure of Consensus in situations with fewer voting members. See Appendix F for more information.

feedback)—which enhances the deliberative process. Consequently, any significant concerns about ongoing risks are expected to be captured in the initial vote. Given the potentially significant implications of losing endorsement for a maintenance measure (e.g., impact on measure credibility, disruption in measure use, loss of interested-party confidence, financial and operational challenges, hindrance to quality improvement efforts), this reconsideration step is critical to ensuring that the committee's final decision reflects a comprehensive and balanced evaluation.

For maintenance measures that fail to reach the 75% consensus threshold and receive less than 60% of votes to retain endorsement, they *are not* reconsidered at the end of the meeting, and their endorsement is removed.

New measures are ineligible for the reconsideration process described above.

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
4a*	**60% to 74%**	0%	26% to 40%	No consensus – Reconsidered at the end of endorsement meeting
4b*	**60% to 74%**		26% to 40%	No consensus – Reconsidered at the end of endorsement meeting

Figure 6. Consensus Voting for Final Endorsement Decisions[4]

**Maintenance measures that fail to reach the 75% consensus threshold but receive between 60% and 74% of votes to retain endorsement (i.e., endorse and/or endorse with conditions) are reconsidered at the end of the endorsement meeting. After further discussion and a revote, if the measure still does not reach the 75% threshold, it loses its endorsement. This reconsideration approach only applies to maintenance measures.*

Voting Confidentiality and Independence

When participating in the voting process within our E&M committee, committee members should understand the expectations regarding confidentiality and independence:

- Each voting member is expected to vote independently, ensuring that their decision reflects

[4] In cases where there are fewer than 20 active voting members, the consensus threshold is lowered to 70%.

 The analyses upon which this publication (or document) 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. Restricted: Use, duplication, or disclosure is subject to the restrictions as stated in Contract Number 75FCMC23C0010 between the Government and Battelle.

their own professional judgment and assessment of the evidence. This independence is crucial for maintaining the integrity and diversity of perspectives in the committee's recommendations.

- While the voting process is not designed to be entirely confidential, members should strive to make their decisions free from external influence. Open and honest discussion is encouraged prior to voting, but the final vote should be based on individual evaluation.
- To align with the American National Standards Institute (ANSI) standards, we collect reasons and suggested resolutions from members voting "Do Not Endorse/Remove Endorsement." This practice supports committee recommendations and ensures that concerns are captured effectively while maintaining confidentiality.
- By adhering to these guidelines, we ensure that our voting process is both robust and representative of the collective expertise within our committee.

Quorum

Effective and organized meeting facilitation is a crucial aspect of a successful consensus-based process as it ensures that discussion remains 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 7).[5]

If less than 60% of the Recommendation Group members are in attendance, then the Recommendation Group does not discuss the measures, and we hold a backup meeting. If meeting quorum is lost during the meeting, the measure evaluation discussion ceases, and we hold an alternative meeting complete the measure review.

[5] For example, if the Recommendation Group has 23 active members, at least 14 members (60% of 23, rounded up) must be present at roll call for the meeting to proceed (meeting quorum).

Figure 7. Quorum Thresholds

Voting quorum is at least 80% of all active Recommendation Group members who have not been recused (*see Conflict of Interest Policy for more details*).[6] If the voting quorum is not met at committee roll call but meeting quorum is achieved, the Recommendation Group proceeds with discussing the measures but final endorsement voting does not occur during the meeting. After the endorsement meeting has ended, the E&M staff shares the meeting recording with Recommendation Group members who were not in attendance and requests they submit their endorsement vote via an offline voting tool. Recommendation Group members 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 all votes are collected, 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 convenes 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

[6] For example, if the Recommendation Group has 23 active members, but three members are recused from voting on a particular measure, voting quorum is based on the remaining 20 members. At least 16 of these 20 members (80%) must participate in the vote for the results to be finalized (voting quorum). Members moved to inactive status before the meeting are not included in the quorum calculation.

decision) the appeal. Consensus is determined to be 70% 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 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

All appeals must also include one or more of the following 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 addition, if a measure's endorsement decision is being appealed, including "Endorsed with Conditions," 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 (*see Appendix G for the Appeal Request Submission Framework*).

Final Technical Report

The E&M staff 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.

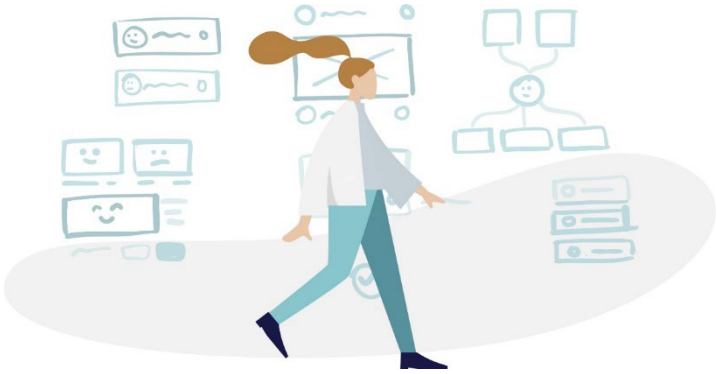
- A summary of the public comments received during the E&M process for the E&M project.
- A summary of any potential high-priority gap areas identified during the E&M project for measure developers to consider for future development.
- A summary of any major concerns or methodological issues raised during performance measure evaluation of the E&M project.

Health Care Quality Landscape

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 non-identical 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. Developers/stewards must explain why existing measures/quality improvement programs are insufficient for addressing the specific health care need.[7]

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 explore each domain of the PQM Evaluation Rubric, discuss common pitfalls and mitigation strategies, and clarify requirements for new and maintenance measures. Additionally, we offer tips for developers to navigate expectations during the endorsement cycle. This enables us to (1) obtain feedback on the recommendations and (2) make developers aware of potential changes to future cycles. This approach 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.

[7] The E&M process does not require independent verification of all existing measures that could be considered related or competing.

Materials from the 2024 measure developer workshop, *What Good Looks Like: A Comprehensive Guide to Successful Measure Endorsement*, can be found on the E&M Resources webpage.

In addition, E&M staff who have measurement expertise can assist developers through the submission and review process. The E&M 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 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.

Guidance for Contacting the E&M staff

To ensure their inquiry is addressed efficiently, measure developers/stewards should follow the guidance below when reaching out to PQMsupport@battelle.org.

Subject Line Categorization

Clearly indicate the nature of the inquiry in the subject line to help us quickly direct the request to the appropriate expert. Use one of the following categories:

- *Introductory Call Request:* For scheduling a call to discuss any aspect of the process.
- *Intent to Submit/Full Measure Submission:* For initial inquiries about the submission process or required forms.
- *PQM Measure Evaluation Rubric:* For questions about rubric domain requirements, interpretation, or expectations.
- *E&M Process and Expectations:* For questions about the evaluation and maintenance process, including deadlines, committee meetings, and general expectations.

Content of the Email

To help us provide the most accurate and timely support, please:

- List all specific questions or areas where assistance may be needed.
- Attach any relevant documents or information that will help us understand the context or the specifics of the inquiry.
- If measure developers/stewards prefer to discuss their questions over a call, indicate this in the email and provide availability for the next 2 weeks so we can schedule a convenient time.

Our team strives to respond to all inquiries within 48 hours. If the request is complex or requires additional consultation, we will inform the requester of the expected timeframe for a detailed response. We encourage developers/stewards to review the available resources on our E&M

Resources webpage before reaching out, as they may find immediate answers to their questions.

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 enters a 5-year maintenance cycle, at which time the developer/steward resubmits the measure to Battelle for PQM endorsement review. Developers/stewards may request an extension of up to 1 year (two consecutive cycles), except if the measure's date of last endorsement has been more than 6 years. Prior to the 5-year maintenance review, at 3 years since the measure's endorsement, developers/stewards provide a status report indicating whether the measure specifications need any changes. The developer/steward may also attest if no changes are required. Once the 3-year status report is submitted, we 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*). Several months before the scheduled maintenance review cycle, Battelle staff contact measure developers/stewards to confirm maintenance assignments.

Annual Updates

Every year 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). **Note:** Expansion of a measure's level of analysis falls under the same CBE ID.
- 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 (e.g., concern with the measure's evidence, updated measure specifications and testing). Any party can request an early maintenance review 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.

We recruit additional SMEs from PQM, as needed, to ensure an appropriate combination of perspectives. 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 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 staff 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 "Measures Table" within the QMDSA must list all the measures (measure number and measure title) being submitted for review.
- For existing measure stewards, only a signed addendum (Additional and Maintenance Measures Form) is needed and will be appended to the existing QMDSA; a new QMDSA is not required.

Only one QMDSA is necessary per measure steward. If the steward is a governmental organization, a QMDSA is not required. Each QMDSA will remain effective for 5 years from the date it is signed. For more information about how to complete the QMDSA, please see the QMDSA Submission Instructions or visit the Measure Submission webpage.

 The analyses upon which this publication (or document) 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. Restricted: Use, duplication, or disclosure is subject to the restrictions as stated in Contract Number 75FCMC23C0010 between the Government and Battelle.

Appendix B: Personal/Organizational Disclosure of Interest Form

1. Your Name:

   Your Organization Affiliation:

   Committee Name:

Describe any personal or organizational relationships subject to disclosure (e.g., disclosures may include relationships with employees of organizations developing or stewarding the measure, or stock options in companies that may benefit from the measures).

2. 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: _____

You and all other persons and organizations must be free of any 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) 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 C: Measure Disclosure of Interest Form

NOTE: *You will be asked to complete this form for each cycle (Spring or Fall).*

1. **Name.** *Required to answer. Single-line text.*
    - Enter your answer

2. **Your Organization / Affiliation.** *Required to answer. Single-line text.*
    - Enter your answer

3. **Do you have any personal or organizational measure conflicts with the measure(s) or measure developer/steward organizations listed below? If you have worked as an employee, collaborator, or consultant of the measure developer/steward listed OR contributed to the development of the measure listed, in any capacity, in the past five (5) years, indicate "Yes" below.** *Required to answer. Single choice.*

    CBE #: Measure Title (Name of Developer/Steward)

    - Yes
    - No
    - Maybe

4. **Please provide a detailed description of your involvement AND/OR any relevant personal or organizational measure conflicts. Please include CBE #, Measure Title, and Name of Measure Developer/Steward.** *Required to answer. Multi-line text.*
    - Enter your answer

5. **Would you or your spouse, domestic partner, or child receive a direct financial benefit from a measure being recommended for endorsement? For example, you own stock in a company that has a financial interest in the measure(s) listed in #3 being endorsed or not endorsed.** *Required to answer. Single choice.*
    - Yes
    - No
    - Maybe

6. **If you selected "Yes" or "Maybe" please provide a detailed description of any direct financial benefit. Please include CBE #, Measure Title, and Name of Measure Developer/Steward.** *Required to answer. Multi-line text.*
    - Enter your answer

7. **In the last 5 years have you received an indirect financial benefit, i.e., not related to the measure(s) under review, of $10,000 or more from a measure developer whose measure is under review? Or have you received 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? 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(s) being reviewed in #3. *Required to answer. Single choice.*

- Yes
- No
- Maybe

8. **If you selected "Yes" or "Maybe" please provide a detailed description of any indirect financial benefit. Please include CBE #, Measure Title, and Name of Measure Developer/Steward.** *Required to answer. Multi-line Text.*
    - Enter your answer

9. **Electronic Certification**
   By entering in today's date below, I certify that I am the individual named on this form and that I have reviewed the Measure Disclosure of Interest Form, and the information given above is true to the best of my knowledge. *Required to answer. Date.*
    - Please input date (m/d/yyyy)

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.

Appendix H provides additional measure-specific guidance for pediatric clinical quality measures, instrument-derived clinical quality measures, eCQMs, entity-level reliability and validity testing, and cost measures.

Importance

Description: Extent to which the measure is important for making significant gains in health care quality or cost where there is variation[8] in or overall less-than-optimal performance.[9] The measure focus is associated with a material outcome.

Importance Items
Attach a logic model[10] depicting the relationship between structures (i.e., inputs) and processes (i.e., actions) and the desired outcome(s).
Summarize evidence of measure importance from the literature linking the structure/process/intermediate outcome to the desired outcome(s).
[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 at the specified level of analysis.

[8] Variation indicates differences in performance across different providers, populations, or settings. Variation can highlight differences in care delivery and identify areas needing targeted improvements.

[9] Refers to performance levels that do not meet established benchmarks or standards, indicating room for improvement. It is identified when the performance does not achieve desired outcomes or fails to deliver the expected quality of care.

[10] Battelle's Logic Model Guidance for the (E&M) of Clinical Quality Measures explains the purpose of a logic model in the context of quality measurement, describes the elements of a logic model, and highlights common approaches, pitfalls, and mitigations for developing a visually effective logic model.

Importance Items
[For initial endorsement] 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;[11] 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[12] 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 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

[For initial endorsement]

- There is no description of why existing measures/quality improvement programs are insufficient for addressing this health care need; OR
- Proposed measure has the same measure focus and target population as existing measure(s) and offers no advantage in terms of addressing gaps in care, feasibility, potential use, or scientific acceptability; OR
- 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:

[11] 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. The developer/steward should also conduct an evaluation of the quality of the evidence. Often clinical practices guidelines conduct systematic reviews. If a literature review is not possible, the committee would consider a rationale as to why.

[12] 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.

- Criterion is not met (see above), but the reviewer can identify changes to specifications or submission 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 (e.g., 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.

[For initial endorsement]

- Description of existing measures or programs justifies[13] the proposed measure's focus among the proposed measure's target population and/or the proposed measure in comparison 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.

Closing Care Gaps

Description: Extent to which the measure can distinguish differences in care for certain patient subpopulations, which can be used to close gaps in care across those identified subpopulations.

Closing Care Gaps*
[For initial endorsement] Identify and describe from existing literature, internal analyses, etc. known and existing differences in health care and health outcomes across patient subpopulations (e.g., demographics, geography, insurance type) related to the measure focus area. Discuss how the findings relate to the measure and how the measure may contribute to efforts to improve health care delivery and outcomes for these identified groups. *For example, given a measure that is designed to assess the timeliness of follow-up care after hospital discharge, recent literature may show that barriers such as geographic location significantly impact follow-up care. These findings are applicable to the measure because they provide insights into factors that can influence the timeliness of follow-up care. Accountable entities may use this measure to*

[13] Measure developers/stewards must document why the proposed measure is justifiable in comparison 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/payers, 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).

Closing Care Gaps*
identify which subpopulations are most affected and implement targeted strategies to address these barriers.
[For maintenance] If data are available, provide a description of your methodology and approach to empirically test differences in performance scores across identified groups (e.g., demographic, geographic variables).
[For maintenance] If data are available, provide the results and an interpretation of those results, including explanations for any variations in performance scores across different groups. Discuss how these results relate to existing evidence, any limitations found in the results, and the potential impact of these variations on the identified groups.
[For maintenance] If information is available, discuss how accountable entities have utilized the measure to close gaps in health care delivery and outcomes for the identified groups.

**Closing Care Gaps is optional for the Spring 2026 cycle and Fall 2026 cycles.*

Not Met:

- No attempt to identify or describe any known differences in health care and health outcomes relevant to the measure focus area; OR
- No explanation of how the measure may contribute to addressing imbalances in health care delivery and outcomes.

[For maintenance]

- If empirical testing conducted,
  - No description of methodology or approach for empirical testing of variations in performance across different identified groups when variables are available; OR
  - Methodology or approach used for empirical testing is insufficient or not appropriate for testing variations in performance across different identified groups; OR
  - No results or interpretation provided on performance differences across identified groups when testing was conducted; OR
- No discussion or evidence on how accountable entities have used the measure to close gaps in health care delivery and outcomes for the identified groups, when such information is available.

Not Met but Addressable:

- Criterion is not met (see above) but the reviewer can identify changes to the specifications or submission that would strengthen the measure's ability to close gaps in care such that the criterion could be met

Met:

- Reviewer determines sufficient assessment of closing care gaps was conducted (e.g., comprehensive evidence of known variations, appropriate methodology provided, performance score differences tested across identified groups, and interpretation of results);

AND

- Discussion of how accountable entities have used the measure to close gaps in health care delivery and outcomes for the identified groups, when such information is available.

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 structured or unstructured fields within electronic health records or other electronic sources—including the extent of missing data, potential inaccuracies, and audit capabilities—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 because 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 because 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 any costs or burden of data collection, data entry, data validation, and analysis, including the impact on clinician workflow (e.g., modifications), diagnostic thought processes, and patient–physician interaction. Discuss barriers encountered or that could be encountered in implementing the measure specifications, data abstraction, measure calculation, or performance reporting.
Explain the implementer's 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.
Describe how the feasibility assessment informed the final measure, indicating any decisions made to adjust the measure in response to data availability.

Feasibility
If the measure includes proprietary information, 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:

- Reviewer determines that a comprehensive feasibility assessment was not conducted. This includes:
  - [For initial endorsement]
    - No evaluation of whether required data elements are routinely generated during care, available electronically, or specified in structure or unstructured fields; OR
    - There is no credible plan to implement electronic data collection within 1 year if necessary; OR
    - No discussion of missing data, potential inaccuracies, audit capabilities, or the data collection strategy cannot be implemented.
  - [For maintenance] – if changes to measure specifications have occurred
    - No evaluation of how specification changes impact feasibility—such as whether required data elements are routinely generated during care, available electronically (in structured or unstructured fields) and the extent of missing data, potential inaccuracies, and audit capabilities); OR
    - 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.
  - [For maintenance] – if no changes to measure specifications have occurred
    - 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.
- Costs and burdens of data handling, including impacts on clinician workflow and patient interactions, are not well-documented; OR
- Barriers to measure implementation and data security, particularly concerning patient confidentiality, are not thoroughly discussed; OR
- If the measure includes propriety information, there is no description of fees, licensing, or other requirements for using the measure as specified.

Not Met but Addressable:

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

Met:

- Reviewer determines that a comprehensive feasibility assessment was conducted. This includes:
  - [For initial endorsement]

  - Evaluation of the generation and use of data elements during care, their electronic availability, and specification of data fields is complete; AND
  - A credible plan is established to implement electronic data collection within 1 year, ensuring necessary adjustments; AND
  - The assessment discusses the extent of missing data, potential inaccuracies, audit capabilities, and the measure's impact on data structure for maintenance.
- [For maintenance] – if changes to measure specifications have occurred
  - Evaluation of how specification changes impact feasibility—such as whether required data elements are routinely generated during care, available electronically (in structured or unstructured fields) and the extent of missing data, potential inaccuracies, and audit capabilities); AND
  - 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.
- [For maintenance] – if no changes to measure specifications have occurred
  - 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.

- Costs and burdens of data handling, including impacts on clinician workflow and patient interactions, are well-documented; AND
- Barriers to measure implementation and data security, particularly concerning patient confidentiality, are thoroughly addressed; AND
- If the measure includes proprietary information, descriptions of fees, licensing, or other requirements for using the measure as specified are provided.

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 any year(s) within 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.

Scientific Acceptability
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 (*see Appendix H4 for more guidance on entity-level reliability testing)*.
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).
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*

Scientific Acceptability
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 (*see Appendix H5 for more guidance on entity-level validity testing)*.
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 adjustment model ☐ Statistical case-mix adjustment model ☐ Stratification by risk factor category ☐ 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.
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

Scientific Acceptability
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[14]
  - 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]
    - No person- or encounter-level reliability testing or existing evidence provided of all critical data elements (numerator, denominator, exclusions).
- For Accountable Entity-Level Reliability[14,14,15]
  - Signal-to-noise < 0.6 for 70% of more of the accountable entities; OR
  - Split-half reliability (ICC) < 0.6 for 70% or more of the accountable entities; OR
  - At least one accountable entity has a reliability estimate < 0.4 and no justification or descriptive statistics are provided (see Appendix H4); OR
  - [For maintenance]
    - No accountable entity-level reliability testing provided.

Validity

- For Person- or Encounter-Level Validity

[14] Reliability estimate expectations informed by the Scientific Methods Panel-recommended reliability thresholds.

[15] 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.

- Reviewer determines the methodology to assess validity is inadequate/inappropriate;[16] 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 (e.g., unclear reasoning for conducting a correlation analysis with another quality indicator, absence of a hypothesis for correlations, or lack of supportive evidence from mechanistic studies to justify correlation results); 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.[17]
  - [For maintenance]
    - Face validity is the only type of accountable entity-level validity presented.

Risk Adjustment

- For intermediate outcome, cost/resource use, outcome measures, and patient-reported outcome and patient patient-reported experience performance measures (PRO-PM and PRE-PM, respectively), no description or rationale provided with respect to adjusting or not adjusting for differences in patient characteristics across measured entities; OR
- If adjusting for differences in patient characteristics, no conceptual model provided; OR the conceptual model provided includes patient characteristics that do not influence the measured outcome; OR are not present at the start of care; OR the model includes factors that are associated with differences or inequities in care without sufficient rationale; OR
- *[For maintenance]*
  - Analysis does not demonstrate:
    - Variation in prevalence of identified patient characteristics across measure entities; OR
    - Contribution to unique variation in the outcome; OR
    - Impact of risk or case-mix adjustment for providers at high or low extremes of risk or

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

[17] 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 the logic model, including the inputs and actions identified can plausibly be implemented by accountable entities to improve the measure focus and by doing so, the subsequent outcomes identified would occur. Developers should consider a group of at least 12 relevant experts and disclose their degree of consensus and must provide/discuss any areas of disagreement.

case-mix; OR
  - If adjustment is provided, 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[14,14]
  - Internal consistency $\geq$ 0.7; OR
  - Inter-rater agreement $\geq$ 0.4; OR
  - Test-retest reliability (ICC or Pearson correlation) $\geq$ 0.5; OR
  - Linear relationship $\geq$ 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[14,15,14]
  - Signal-to-noise $\geq$ 0.6 for 70% or more of the accountable entities; OR
  - Split-half reliability (ICC) $\geq$ 0.6 for 70% or more of the accountable entities; AND
  - 100% of the accountable entities have reliability estimates of $\geq$ 0.4; OR
  - Less than 100% of accountable entities have reliability estimates of 0.4, but justification and descriptive statistics are provided (see Appendix H4); AND
  - [For maintenance]
    - Accountable entity-level reliability testing provided.

Validity

- For Person- or Encounter-Level Validity

- Reviewer determines methodology employed 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 is adequate and the analytic approach presented is appropriate and thorough (e.g., clear reasoning for conducting a correlation analysis with another quality indicator, clear hypothesis for correlations, and supportive evidence from mechanistic studies to justify correlation results); 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.**

Risk Adjustment

- For intermediate outcome, cost/resource use, outcome measures, and patient-reported outcome and patient patient-reported experience performance measures (PRO-PM and PRE-PM, respectively), a description or rationale is provided with respect to adjusting or not adjusting for differences in patient characteristics across measured entities; AND
- If adjusting for differences in patient characteristics, a conceptual model is provided, which includes:
  - Patient characteristics that influence the measured outcome; AND are present at the start of care; AND the model does not include factors that are associated with differences or inequities in care unless justification is provided; AND
- Analysis demonstrates:
  - Variation in prevalence of patient characteristics across measured entities; AND
  - Contribution to unique variation in the outcome; AND
  - Impact of risk or case-mix adjustment for providers at high or low extremes of risk or case-mix; AND
  - If adjustment conducted, results demonstrate acceptable model performance.

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
Describe the specific attributes of an accountability program and application in which the measure is best suited. This includes a consideration of: • Target populations (e.g., Medicare beneficiaries, beneficiaries enrolled in an accountable care organization [ACO], pediatric patients) • Accountable entities (e.g., hospital, clinician group/practice) • Care settings (e.g., inpatient setting) • Whether risk adjustment of social risk factors is performed at the measure level or intended at the payment level. These attributes provide context for how the measure results can be used to inform judgments and decisions about performance, such as for public reporting, accreditation, performance-based payment, or network inclusion/exclusion. Consideration of risk adjustment or stratification approaches, data sources, and intended use further support the measure's application in accountability programs.
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)
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.
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

Use and Usability
and care setting.

Not Met:

- There is no discussion of the specific attributes of an accountability application in which the measure supports; OR

[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[18] 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:

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

Met:

- There is discussion of the specific attributes of an accountability program in which the measure supports; AND

[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

[18] 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).

- 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), requires, by law, that 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. Along with utilizing Section508.gov as the 508-accessibility guidance, the following steps should be considered 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. 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 web users with visual impairments 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	1.000
40	0.125	0.125	0.750	0.994
20	0.000	0.250	0.750	1.000
20	0.125	0.125	0.750	0.952
40	0.125	0.750	0.125	0.997
20	0.150	0.750	0.100	0.971
40	0.250	0.000	0.750	0.945
20	0.250	0.000	0.750	0.951
40	0.500	0.000	0.500	0.818
20	0.500	0.000	0.500	0.807
15	0.267	0.000	0.733	0.949
15	0.733	0.000	0.267	0.949
15	0.134	0.733	0.133	0.949

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

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

*When the number of respondents is 20 or more, at 0.950 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). When the number of respondents is less than 20, at 0.950 for the Measure of Consensus, at least 70% of respondents (truncating to the 10th decimal) are in agreement (yellow shading).

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: Appeal Request Submission Framework

1. **Appellant Information**

- Name
- Email address
- Phone number

2. **Measure Information**

- CBE #
- Measure name
- Endorsement decision outcome (e.g., Endorsed, Endorsed with Conditions, Not Endorsed, Endorsement Removed)

3. **Grounds for Appeal**

- Specify the criteria under which the appeal is submitted. Appeals may be based on:
  - **Omitted Evidence:** 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.
  - **Misapplication of Evaluation Criteria:** The CBE's measure evaluation criteria were not applied appropriately (including where the committee may have given excessive weight to factors above and beyond the criteria). The appellant must specify the evaluation criterion that they believe was misapplied and why.
  - **Procedural Error:** 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.

4. **Detailed Rationale**

- Provide a detailed explanation of why the appeal should be considered and why the endorsement decision should be overturned.
- Include any supportive evidence demonstrating omitted evidence, a procedural error, and/or misapplication of the specific endorsement criterion.

5. **Supporting Documentation**

- Attach any relevant documentation, references, or materials that support your appeal.

Appendix H: Measure-Specific Guidance

H1. CBE Policy on Pediatric Clinical Quality Measures

Overview

Due to small sample sizes or other limitations to data availability, CQMs for pediatric populations sometimes lack supporting data to substantiate importance, validity, and usability claims. Under certain circumstances outlined in this policy, measure developers may submit data on a comparable CQM for an adult population to substantiate those claims.

Policy for Extrapolating Adult Clinical Quality Measures to Pediatric Populations

This policy outlines the criteria and considerations for extrapolating evidence from adult CQMs to pediatric CQMs. Extrapolation is justified when the target population, mechanism complex, and material outcomes demonstrate sufficient similarity between adult and pediatric populations. The policy is structured to support three key claims: importance, validity, and usability.

1. Importance Claim

A pediatric CQM may be supported by evidence based on an adult CQM when the measure focus (e.g., structure, process, or intermediate outcome) is demonstrably associated with a pediatric material health outcome. The reasonableness of extrapolation is assessed by:

- Evaluating whether the condition progression and treatment goals align between adults and pediatric populations.
- Identifying any pediatric-specific considerations, such as developmental differences—including growth physiology and intellectual and emotional development—that may alter clinical significance.
- Ensuring that extrapolated measures will provide actionable insights for pediatric health care decision-making.

2. Validity Claim

Validity is established through two sub-claims:

a. **Association Claim**

Evidence supporting extrapolation must demonstrate that responses to the adult CQM correlate with the intended pediatric outcome. This requires:

- Reviewing clinical trial data, observational studies, and real-world evidence comparing condition progression and the mechanism complex between adult and pediatric patients.
- Assessing the relationship between intervention exposure and expected outcomes, if applicable, to confirm similar response patterns across populations.
- Assessing biomarker validity if surrogate endpoints are used.

b. **Mechanism Claim**

Explicit mechanisms must be articulated to justify extrapolation. These include:

- Target population similarity: The extent to which the target population manifests similarly across age groups.
- Intervention dynamics: The extent to which the intervention is processed, delivered, and acted upon in comparable ways across populations.
- Response to treatment: Evaluating whether the mechanism complex effectiveness observed in adults is expected in pediatric populations.
- If substantial differences exist in any of these areas, additional pediatric data collection may be required before extrapolation is deemed valid.

3. Usability Claim

A pediatric CQM extrapolated from an adult measure must be practically applicable, considering:

- The feasibility of using adult-derived endpoints in pediatric populations (e.g., modifying functional tests for younger children).
- The availability of pediatric-relevant data sources to support the measure's implementation.
- CBE and review committee acceptance of the extrapolation methodology.

Policy Implementation Considerations

- **Target Population:** This guidance applies to target populations of CQMs that solely focus on pediatric populations or where the population spans both pediatric and adult populations. For the latter, the same extrapolation guidance applies, and the developer would provide evidence for the adult population and extrapolated evidence for the pediatric population.
- **CBE Engagement:** Measure developers/stewards should engage in early discussions with the CBE to ensure that extrapolation assumptions align with this policy.
- **Modeling and Simulation:** When direct pediatric evidence is limited, model-based approaches (e.g., care process modeling) can help validate extrapolation assumptions.
- **Safety Considerations:** Safety extrapolation must be rigorously justified, particularly for measures involving therapeutic interventions.

Conclusion

Extrapolation of adult CQMs to pediatric CQMs is appropriate when the target population, measure focus, context, mechanism complex, and material outcomes demonstrate sufficient similarity. A structured assessment of importance, validity, and usability ensures that pediatric measures are both scientifically robust and practically implementable, minimizing the need for pediatric data collection while maintaining high standards of clinical quality measure evaluation.

References

E11A Pediatric Extrapolation: Guidance for Industry. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER) (December 2024). https://www.fda.gov/media/161190/download

Disclaimer: Generative artificial intelligence was used in the initial drafting of the CBE Policy on Pediatric Clinical Quality Measures document. Actual humans with relevant expertise subsequently validated and revised the content.

H2. CBE Requirements for Electronic Clinical Quality Measures

Overview

Electronic clinical quality measures (eCQMs) use electronic standards to reduce the burden of manual data abstraction and reporting for measured entities. However, significant variability within and across EHR vendor products—such as custom implementations versus out-of-the-box configurations—can lead to inconsistencies in how data are captured, which may affect a measure's feasibility and reliability during eCQM testing. Testing should balance the need for representative and rigorous evaluation with practical considerations such as implementation burden, cost, and alignment with current health care quality priorities and market dynamics.

This guidance is intended to supplement the overarching E&M policies, processes, and measure evaluation criteria by outlining specific requirements that apply exclusively to eCQMs. The requirements are based on the CBE's experience in reviewing eCQMs and discussions with experts on the CBE Scientific Methods Panel.

General Requirements

The following clarifications are specific to eCQMs:

- A new eCQM version of an existing endorsed [non-eCQM] measure is ***not*** automatically considered to be endorsed. Given the distinct characteristics of an eCQM, the new measure must be separately 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) or Fast Healthcare Interoperability Resources (FHIR), and Clinical Quality Language (CQL).
- 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 should first check for an existing one that aligns, to avoid unnecessary duplication. 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.

- **Note:** To support a high-quality eCQM submission for endorsement, developers/stewards must evaluate the accuracy of value sets used (i.e., value sets align with the intent of the measure and specifications). Developers are encouraged to attest and/or describe the approach used to confirm the accuracy of the value sets (e.g., SME review).

PQM Measure Evaluation Rubric Guidance

To establish that eCQMs are safe and effective, E&M focuses on several key areas, particularly: Feasibility, Importance, and Scientific Acceptability (i.e., Reliability and Validity). Developers/stewards must justify how the test sites are distinct from one another in a meaningful way to demonstrate broader eCQM implementation.

1. Feasibility and Person- or Encounter-Level (i.e. Data Element) Reliability and Validity

- For initial endorsement, developers/stewards must conduct feasibility and person- or encounter-level testing **within at least one EHR vendor.** To facilitate broader implementation of the eCQM, developers/stewards should consider testing across a number of vendors and test sites they deem appropriate. For example, testing in **three or more** sites of different ownership may provide valuable insights into the measure's feasibility, applicability, and effectiveness. Developers should also provide a rationale to support the number of test sites and/or vendors chosen for testing.
- Submission requirements for eCQMs include a feasibility assessment using the eCQM Feasibility Scorecard. This assessment identifies potential feasibility issues related to the capture of required data elements, including:
  - Data availability: Are the data readily available in a structured format, i.e., stored in fixed fields within the electronic health record (EHR)?
  - Data accuracy: What is the accuracy of the data element in EHRs under normal operating conditions? Are the data source and the individual recording the data clearly specified?
  - Data standards: Is the data element coded using a nationally accepted terminology standard (e.g., Systematized Nomenclature of Medicine—Clinical Terms [SNOMED-CT], Current Procedural Terminology [CPT])?
  - Data workflow: Are the data captured as part of routine clinical care? What impact does this have on user workflow?
- Empirical demonstration of person- or encounter-level (i.e., data element) reliability is required for any unstructured data fields, and person- or encounter-level (i.e., data element) validation is required for all eCQMs. Testing must include all critical data elements (not just agreement of one final overall computation for all patients). At a minimum, the numerator, denominator, and exclusions (or exceptions) must be assessed and reported separately. If person- or encounter-level testing is not possible, justification is required and must be accepted by the E&M committee. Reliability is assumed for electronic data from structured fields.
- 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 medical record (not just the fields used to obtain the electronic data). Developers/stewards should use statistical

 The analyses upon which this publication (or document) 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. Restricted: Use, duplication, or disclosure is subject to the restrictions as stated in Contract Number 75FCMC23C0010 between the Government and Battelle.

analyses, such as sensitivity and specificity, positive predictive value, and negative predictive value.

2. **Performance Gap (Importance) and Accountable Entity-Level Reliability and Validity**

- For maintenance evaluations, performance gap and accountable entity-level reliability and validity must include data from **a minimum of five sites within at least two EHR vendors**. A minimum of five enables the fitting of a quadratic curve and provides some information as to the distribution of results, under strong assumptions. However, developers/stewards should consider testing across a number of vendors and test sites they deem appropriate. Measure developers/stewards are encouraged to prioritize site and patient heterogeneity during testing, considering factors such as variation in resources and implementation approaches (e.g., custom vs. out-of-the-box configurations, comprehensive vs. limited EHRs), geographic region, urbanicity, academic affiliation, and safety net status.
- Performance gap data (Question 2.4 and Table 1 within the Full Measure Submission Form) are optional for initial endorsement but required for maintenance evaluations.
- For maintenance measures, face validity alone is insufficient.
- When using data aggregators or intermediaries as test sites, developers should provide justification that site-level variability and raw data fidelity are preserved, or supplement with direct site-level testing (i.e., not through an aggregator) as needed.

Other Considerations

Recommendations to CMS or other Payers

- Consider funding a centralized testing laboratory that includes a cross-section of organizations—diverse by geography, EHR vendor, academic affiliation, safety net status, and care setting (e.g., primary vs. specialty care).

References

NSF Consulting. (2015, April 13). The Rule of Five: A quick and easy way to reduce uncertainty in business. NSF Consulting. https://nsfconsulting.com.au/rule-of-five-reduce-uncertainty/

Waters Corporation. (2024, March 16). Calculating amount/concentration – Calibration curve fit types: Quadratic fit (Tip #280). Waters. https://support.waters.com/KB_Inf/Empower_Tips_of_the_Week/WKB235238_Calculating_Amount_Concentration_-_Calibration_Curve_Fit_Types_Quadratic_Fit

H3. Instrument-Derived Measure Set Submission Framework

CBE Policy on Instrument-Derived Clinical Quality Measures

The following is the policy of the CBE with respect to instrument-derived 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-derived clinical quality measures and other clinical quality measures. Specifically, all measures are evaluated based on person- or encounter-level (i.e., data element) reliability and validity as well as accountability entity-level reliability and validity.
- For person- or encounter-level (i.e., data element) reliability and validity, measure developers/stewards may cite existing literature to substantiate those properties.
- Measure developers/stewards should use valid and reliable instruments within the CQM. Developers and stewards can justify this by either conducting the appropriate validity and reliability tests of the instrument and attesting that the instrument or survey was developed using a best practice protocol (e.g., Holmbeck, 2009). If the instrument has previously been tested for reliability and validity, developers/stewards may cite this evidence.
- 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, see *Overview of Instrument-Derived Measure Set Submission Framework* below.
- E&M 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.

Overview of Instrument-Derived Measure Set Submission Framework

Beginning with the Spring 2025 E&M cycle, measure developers will submit their instrument-derived measures (e.g., measures derived from Consumer Assessment of Healthcare Providers and Systems [CAHPS] surveys, the Hospice Outcomes and Patient Evaluation [HOPE], or End-Stage Renal Disease [ESRD] Patient Life Goals Survey [PaLS]) using a new submission framework. Unlike the prior framework, which evaluated and endorsed a set of measures from a single instrument as one unit, the new framework evaluates and endorses each derived measure individually.[19]

The requirements and endorsement criteria for instrument-derived measures are the same as other clinical quality measures. The CBE does not endorse instruments or surveys;

[19] Individual instrument or survey items that are combined into a composite and aggregated to the accountable entity level would be reviewed and endorsed as a single measure. Developers are encouraged to review what is feasible and appropriate for their measure(s).

however, developers are encouraged to attest that the instrument or survey was developed using a best-practice protocol.[20] Benefits of this new framework include:

- More granular and appropriate measure evaluations and endorsement decisions.
- Actionable feedback and improvement opportunities at the derived measure level.
- Alignment with other initiatives by CMS, such as the CMS Measures Inventory Tool (CMIT) and the Measures Under Consideration (MUC) Entry/Review Information Tool (MERIT).

This overview introduces the new framework for instrument and instrument-derived measure (IDMs) set submissions and is intended to assist developers in planning for these submissions. For more information on E&M submission processes and to download submission templates, visit the E&M Measure Submission page.

H4. CBE Guidance on Entity-Level Reliability

The following clarifications are specific to entity-level reliability.

Mitigation of Low Reliability

Low reliability does not necessarily mean a measure is not useful; its utility depends on available alternatives and the consequences of its use. When the measure focus is closer to the continuum of "should not happen," (e.g., never events) less emphasis may be placed on reliability. Reasoning should rule out other explanations for the association between the response to the quality program and the measure focus. This means that even if a measure has low reliability, it can still be useful if the reasoning behind its use effectively rules out alternative explanations for the observed outcomes. Essentially, the measure's utility is not solely dependent on its reliability but also on the strength of the reasoning that supports its use in the given context.

Efforts to mitigate harm due to low reliability must be balanced against other considerations, such as reduced validity or quality program participation. For example, the measure developer/steward may consider increasing the minimum sample size (as a larger sample size may provide more accurate and reliable estimates) or extending the period of performance (as this may allow for a more comprehensive assessment of performance over time.

Actual harm depends on the decision context, including features of the quality program such as thresholds or benchmarks, low-volume payment adjustments, and voluntary reporting. Additionally, understanding the features of entities and persons with low reliability scores and high reliability scores can help identify the factors that contribute to reliability issues and informs strategies to mitigate them (Table H4-1).

[20] For example, see 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.

Table H4-1. Features of Entities and Persons at Low- and High-Reliability Entities

Features	Low-Reliability Entities	High-Reliability Entities
Person-Level Features	*For example:* • *Demographics (age, gender, ethnicity)* • *Health status or conditions (e.g., diabetes, hypertension)*	*For example:* • *Demographics (age, gender, ethnicity)* • *Health status or conditions (e.g., diabetes, hypertension)*
Entity-Level Features	*For example:* • *Performance score distribution of low-reliability entities* • *Type of health care facility (hospital, clinic, long-term care)* • *Size of the entity (number of beds, staff size)*	*For example:* • *Performance score distribution of high-reliability entities* • *Type of health care facility (hospital, clinic, long-term care)* • *Size of the entity (number of beds, staff size)*
Average Person-Level Features by Entity	*For example:* • *Average age of patients served* • *Prevalence of chronic conditions among patients*	*For example:* • *Average age of patients served* • *Prevalence of chronic conditions among patients*
Geographic Features	*For example:* • *Location (urban, rural, suburban)* • *Regional health care access and availability*	*For example:* • *Location (urban, rural, suburban)* • *Regional health care access and availability*

Low reliability might be acceptable in certain contexts, such as:

- When the measure focus has a direct causal association (e.g., complications, health care-associated infections [HAI]).
- When there is no alternative for decision-making. A low reliability measure is often better than no measure at all (e.g., Farmer's Almanac).
- A related high reliability measure can inform choices regarding structure, process, or outcome.
- Trade-offs in harm may favor low reliability over low validity or reduced quality program participation.

Alternatively, some considerations that might ***not*** warrant low reliability include:

- When certain identifiable populations are more likely to receive care from low reliability entities.
- When choices made based on the measure have the potential for material harm to entities or persons.
- When choices made based on the measure results in significant waste, futility, or injustice (low importance, validity, usability).

Accountable Entity-Level Reliability Testing

Table H4-2 reflects the current recommendations from the CBE on calculation of entity-level reliability. Developers and stewards may consider using various methods or approaches. Regardless of the method chosen, they should provide a clear explanation of the method used and guidance on how to interpret the results.

Note: For any accountable entities that have less than 0.6 reliability estimates, developers may consider reporting both entity- and person-level descriptive statistics for those entities. As described above, understanding such characteristics can identify the factors that contribute to reliability issues and help to ensure that measures are applied effectively and that any potential harm due to low reliability is minimized.

Table H4-2. Entity-Level Reliability

Type of Data/Measure	Recommended Method
Binomial (i.e., observed rate)	Signal-to-noise method calculated as suggested by Nieser and Harris (2024). (Note: the N-H method addresses the issue that the Adams [2009] method estimates reliability of 1.0 when the observed rate is 0% or 100%.)
Non-binomial (i.e., risk-adjusted rate, ratio, hierarchical model, composite)	Intraclass correlation coefficient (ICC) followed by applying the Spearman-Brown prophecy formula adjustment and permutation sampling OR Spearman Rank Correlation (SR) using the Spearman-Brown prophecy formula adjustment and permutation sampling. (Note: Both the ICC and the SR, using the Spearman-Brown prophecy formula adjustment, may be used without permutation sampling, but permutation sampling improves the estimates).

References

Adams, J.L. (2009). The Reliability of Provider Profiling: A Tutorial. [online] *Rand.org*. https://www.rand.org/pubs/technical_reports/TR653.html

Nieser, K.J. and Harris, H.S. (2024) Comparing methods for assessing the reliability of health care quality measures. *Statistics in Medicine*: 43(23).

Warrens, M.J., (2017) Transforming intraclass correlation coefficients with the Spearman–Brown formula. *Journal of Clinical Epidemiology*. (85): 14-16.

Kalbfleisch, J.D., He, K., Xia, L. Li, Y. (2018). Does the inter-unit reliability (IUR) measure reliability? *Health Services and Outcomes Research Methodology*. 18:215-225.

Aume, L., White, W., Chen, P., and Geppert, J. (2025). An Assessment of Reliability Estimation Methods for Binomial Health Care Quality Measures (forthcoming)

Spearman–Brown Prophecy Formula: https://methods.sagepub.com/ency/edvol/encyc-of-research-design/chpt/spearman-brown-prophecy-formula

- (k is the ratio of the entity n to the median n across all entities).

H5. CBE Guidance on Entity-Level Validity

The following clarifications are specific to entity-level validity:

Correlation and Causal Explanation:

- **Assessing Correlation:** This involves examining the relationships between different indicators of quality (which can be other endorsed measures) to see if they correlate as expected based on the theoretical constructs they are supposed to measure.
  - **Example:** A developer assesses patient falls measure with another indicator of patient safety, use of patient safety protocols. The developer found a correlation coefficient of -0.25 between use of safety protocols and reduction in patient falls. Although in the expected direction, this correlation is lower than expected based on the theoretical constructs that both measures aim to improve patient safety.

- **Providing Causal Explanation:** If the correlations are not as strong as expected, providing a causal explanation helps to understand whether the measures are valid or if other factors are influencing the results.
  - **Example:** To explain the low correlation, the developer might investigate whether different implementation levels of safety protocols across various departments affect the effectiveness of these protocols in reducing patient falls. They might also consider external factors such as patient demographics or departmental differences that could influence the results.

Evidence from Mechanistic Studies:

- **Literature Review:** This involves reviewing existing research to see if it supports the constructs the measures are intended to assess. This includes mechanistic studies relevant to the measure's focus. While much literature may pertain to related topics, it is essential to justify the external validity of the evidence concerning the specific measure.
  - **Example:** A review of literature on the mechanisms by which safety protocols reduce patient falls might detail how specific interventions, such as the use of bed alarms, non-slip footwear, or enhanced staff training on patient mobility directly contribute to reducing the incidence of falls. These mechanistic insights provide a solid foundation for understanding how and why certain safety protocols are effective, thereby supporting the validity of a measure that assesses the implementation and effectiveness of these protocols. For example, the studies might reveal that bed alarms are significantly less effective in wards where patients have higher mobility or cognitive impairments, explaining the lower overall correlation.

- **Justification of External Validity:** This is crucial when the direct evidence from the literature does not exactly match the constructs of the measures. If the existing literature does not directly address the measure but is deemed relevant, developers must provide a justification for its applicability. This includes detailing how the findings from the literature support the measure's logic model and intended outcomes.
  - **Example:** If the reviewed literature primarily discusses safety protocols in contexts other than patient falls, such as preventing medication errors or reducing infections, mechanistic insights can still be used to justify their applicability to fall prevention. For instance, the fundamental mechanisms of enhancing staff vigilance and adherence to protocols can be similar across different safety contexts. This similarity supports the measure's validity by demonstrating that effective protocol implementation can lead to safety improvements, regardless of the specific adverse event being targeted.

H6. CBE Guidance on Cost Measures

Overview

Cost measures include both episode-based cost measures, which cover a range of procedures, acute inpatient medical conditions, and chronic conditions, and population-based cost measures, which focus more broadly on the entire spectrum of care. Cost measures are sometimes challenging to review due to the complexity of the measure specification, the challenge of attributing cost to clinicians or groups, and the potential relationship between better performance on the cost measure and quality of care.

The intent of this guidance is to provide information to measure developers and E&M committee members about the CBE's approach to reviewing cost measures. The guidance is based on the CBE's experience in reviewing cost measures and discussions with experts in cost measure methodology on the CBE Scientific Methods Panel.

Objective

The objective of the guidance is to focus the review of cost measures on a few material assertions and to standardize to some degree the claims and evidence presented to substantiate measure properties. Our intent is to reduce cost measure evaluation burden and to support a shared understanding among both measure developers and E&M committee members, while at the same time continuing to ensure that cost measures are safe and effective for use in accountability applications.[21]

Guidance

Principles

[21] 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).

The cost measures guidance is informed by the following principles:

1. Cost measures are not a stand-alone quality construct but rather must be interpreted in the context of quality to capture efficiency or value (outcome/cost)
2. While often characterized as resource measures, as constructed cost measures are estimates of what a payer (e.g., health plan, Centers for Medicare & Medicaid Services) would pay, i.e., the cost to the payer. As constructed, the measure focus often uses standardized prices, which are fixed dollar amounts applied uniformly to specific health care services or events, such as diagnoses or procedure codes (e.g., diagnosis-related group [DRG], current procedural terminology [CPT]). This approach allows assessment of relative resource use, independent of the actual cost of service delivery.
3. Accountable entities will vary in the resources available for service delivery per case. Resources available and applied at the accountable entity-level are imperfectly measured by standardized prices, and actual resources and accountable entity costs of furnishing care may be higher or lower than the estimates based on standardized prices.
4. A primary harm of cost measures is the potential to drive down utilization and decrease access to necessary care or stinting. A secondary harm may be discouraging treatment for patients anticipated to be high cost.
5. Primary benefits of cost measures are more efficient care (outcome/cost) and fewer complications and/or avoidable utilization (e.g., emergency department visits, additional procedures, and inpatient admissions or re-admissions).

PQM Measure Evaluation Rubric Guidance

To establish that cost measures are safe and effective, E&M focuses on several key areas, particularly: Importance, Scientific Acceptability (i.e., reliability and validity), Feasibility, and Usability. Table H6-1 provides practical solutions to common challenges encountered within these domains, such as aligning standardized prices with available resources and ensuring necessary care is not compromised. These solutions help ensure that cost measures address health care priorities effectively, supporting their relevance and applicability in real-world scenarios.

Table H6-1. Cost Measure Guidance

Problem	Potential Solution
Standardized prices or dollar-weighted services do not reflect the resources available for service delivery (i.e., actual resource use and costs may differ from estimates based on standardized prices).	Stratify entities by payer mix (e.g. Medicare, Medicaid, commercial, other).
Potential for a decrease in necessary care.	Identify services that may be substituted with lower quality care, and stratify entities based on the propensity to receive those services (post risk adjustment). Consider whether valuable higher cost services (e.g., rehabilitation services) are being used

 The analyses upon which this publication (or document) 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. Restricted: Use, duplication, or disclosure is subject to the restrictions as stated in Contract Number 75FCMC23C0010 between the Government and Battelle.

Problem	Potential Solution
	at appropriate levels for appropriate patients and whether risk adjustment or stratification adequately protects against this. OR Consider excluding services in which there is a high likelihood of stinting on a specific service due to high cost.
Demonstrate more efficient care (outcome/cost), as shown in the green boxes in Table H6-3 (i.e., better quality performance for a given level of cost performance or, equivalently, better cost performance for a given level of quality performance).	Populate the relationship between better performance on the cost measure and better performance on relevant measures of quality (process or outcome) (See Table H6-3).*
Demonstrate fewer complications or avoidable utilization (e.g., emergency department visits, additional procedures, and inpatient admissions or readmissions).	Association and mechanistic studies on better performance on the cost measure and lower rates of complications and avoidable utilization. Focus on care that most clearly reflects complications or avoidable care (e.g., emergency department use, higher than expected numbers of provider/facility visits).

***Note:** Table H6-3 is just one example developers/stewards may use. Alternative formats (e.g., scatterplots), may also be used. Regardless of the approach chosen, developers/stewards should provide a clear explanation of the method used and guidance on how to interpret the results.

The current PQM Measure Evaluation Rubric (*see Appendix D*) provides a structured framework for evaluating cost measures to ensure they meet the necessary criteria for endorsement. Each rubric domain is rated as "Met," "Not Met, but Addressable," or "Not Met," based on the respective evaluation criteria. Table H6-2 illustrates how these rubric elements apply to cost measures, offering practical examples.

Table H6-2. PQM Measure Rubric Domain Cost Examples

Rubric Domain	Applicability to Cost Measures	Example

Importance	• Cost measures should be developed within the framework of a logic model to ensure a structured approach that clearly links them to health care priorities related to efficiency and value (i.e., improving outcomes while reducing cost). Priorities may include:
	  ○ Addressing the leading causes of inefficient care as a driver of morbidity and mortality.
	  ○ Targeting areas of high resource use.
	  ○ Focusing on conditions with high severity.
	  ○ Addressing financial strain and burden experienced by patients due to high out-of-pocket health care costs.
	• To ensure relevance and effectiveness in tackling critical health challenges, there must be a clear, evidence-based description of the resources and actionable steps that an accountable entity can adopt to improve cost measure performance. This ultimately enhances care quality and ensures patient safety.
		• An episode-based cost measure for heart failure management can address these priorities by focusing on a condition that significantly impacts morbidity and mortality rates, involves substantial resource utilization due to frequent hospitalizations and ongoing care needs, and requires intensive management due to its severity.
		• Logic model application:
		  ○ **Inputs:** Identification of heart failure as a high-impact condition requiring significant health care resources and intensive management.
		  ○ **Activities:**
		  ▪ Implementing standardized care pathways to ensure consistent and efficient treatment.
		  ▪ Enhancing patient education to empower self-management and reduce dependency on hospital resources.
		  ▪ Utilizing telemonitoring programs to continuously monitor patient conditions and prevent acute episodes.
		  ▪ Applying data analysis to identify areas for cost reduction and outcome improvement.
		  ○ **Outputs:** Improved care coordination, increased patient engagement in self-care, and enhanced monitoring of patient conditions.
		  ○ **Outcomes:** Reduction in the frequency of hospitalizations, improved management of heart failure symptoms, and decreased health care costs.
		  ○ **Impact:** Overall enhancement in the efficiency of heart failure management across the health care system, leading to better patient outcomes and reduced financial strain on patients and the system.

Scientific	• Reliability testing should ensure that the measure	• In the context of a heart failure episode-based care

Rubric Domain	Applicability to Cost Measures	Example
Acceptability – Reliability	produces consistent results. This consistency is crucial for safety, as it guarantees that the measure reliably identifies the intended health care quality aspects without variation that could lead to misinterpretation or errors.	measure, since the measure focus uses standardized prices, the process of generating standardized prices is algorithmic and reliability is generally assumed. However, evidence should be provided to ensure the accuracy and consistency of the underlying data being used within the standardization algorithm. • For accountable entity-level testing, using the Intraclass Correlation Coefficient (ICC) with Spearman-Brown adjustment is recommended to ensure that the measure accurately captures cost variations due to systematic differences between accountable entities rather than random variance. This involves assessing the proportion of variance attributed to actual differences in care delivery versus noise, ensuring that the measure consistently reflects true performance differences across various health care entities. • By evaluating these aspects, the measure can reliably inform health care providers about cost efficiency and guide improvements in heart failure care.
Scientific Acceptability – Validity	• Validity of cost measures refers to an overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the extent to which a measure accurately captures the intended costs that are influenced by the accountable entities attributed to those costs. The measure focus is on standardized prices, ensuring that the measure reflects reimbursements rather than the cost of providing those services. • Validity of cost measures involves ensuring that the measure can be impacted by those accountable entities through the reduction of the billable utilization of services and properly adjusts for external factors not under their influence. This means that there is evidence of accountable entities' ability to impact	• A heart failure episode-based care measure should reflect standardized prices that accountable entities can impact, such as those related to hospital readmissions and medication adherence. • Mechanistic studies showing reductions in heart failure complications or lower rates of health care resource utilization related to heart failure provide supportive evidence of accountable entities' ability to impact the measure. • Risk adjustment should be considered and applied by including factors like patient demographics or clinical history to account for variations in patient populations. Explicit consideration of provider-treatment choices, such as post-hospital institutionalization versus home health care, may also be integrated into the risk adjustment

Rubric Domain	Applicability to Cost Measures	Example
	billable services (e.g., mechanistic studies), which is essential for the measure's effectiveness. This should include an explicit consideration of treatment choices that can influence costs. By incorporating standardized measures like Hierarchical Condition Categories (HCCs), clinical severity measures (as applicable), and appropriate treatment indicators into risk adjustment covariates, there is a better account for bias, ensuring more accurate assessments of accountable entities' performance. • Additionally, establishing acceptable model performance may be achieved by benchmarking risk adjustment models against previous measures with similar populations.	covariates to enhance model performance.
Feasibility	• Feasibility of cost measures involves ensuring that the required data are readily available or can be captured without undue burden, facilitating their implementation for performance measurement. This ensures that the measure can be implemented effectively without imposing excessive burdens on health care providers, ultimately leading to more informed decision-making and improvements in care delivery.	• In a heart failure episode-based care measure, data elements such as hospital readmissions and medication adherence should be routinely generated and used during care delivery. • Data should be accessible through electronic health records (EHRs) or other electronic sources (e.g., claims). If not, a credible path to electronic collection should be specified. • The data collection strategy should be practical, considering factors like data source availability, timing, frequency, and costs associated with proprietary measures.

Usability	• Usability of cost measures focuses on their practical application and impact on health care decision-making. This involves:
	  ○ Tracking progress and identifying areas for improvement.
	  ○ Actively seeking and incorporating feedback from accountable entities and other interested parties to refine the measure and enhance its
		• A heart failure episode-based care measure should demonstrate a clear relationship between better performance on the cost measure and improved quality of care outcomes, such as reduced complications and/or avoidable utilization.
		• Unexpected findings may include unintended consequences, such as a potential decrease in necessary care or more efficient care.

Rubric Domain	Applicability to Cost Measures	Example

	effectiveness.
	• Usability also emphasizes the importance of ongoing improvement in results and the identification of unexpected findings, both positive and negative, to ensure the measure remains relevant and beneficial over time.
	• While a measure may be valid, poorly designed incentives can lead to negative outcomes, such as reducing the quality of care to cut costs. This highlights that unintended consequences are not necessarily a threat to the validity of a measure but should be considered in the measure's use and usability.
	• In terms of the potential decrease in necessary care, this refers to the risk that cost measures might inadvertently lead to reductions in essential health care services as entities aim to lower costs.
	  ○ Addressing this involves pinpointing specific health care services that might be replaced by lower-quality alternatives, which could compromise patient care. Then, after risk adjustment, stratifying entities based on their propensity to provide those services.
	• To effectively analyze and interpret the relationship between cost performance and quality of care, **Table H6-3** provides an example framework for cross-referencing the Importance Table (performance score by decile) at the entity level.*
	  ○ This involves comparing the cost measure with one or more relevant measures of process or outcome. For each dimension (cost and quality), worse performance is categorized as the bottom 30%, neutral performance as the middle 40%, and better performance as the top 30%.	

Rubric Domain	Applicability to Cost Measures	Example
	○ Each cell in **Table H6-3** should contain the mean cost measure, the mean quality measure, and the number of entities and persons.	

***Note:** Table H6-3 is just one example developers/stewards may use. Alternative formats (e.g., scatterplots), may also be used. Regardless of the approach chosen, developers/stewards should provide a clear explanation of the method used and guidance on how to interpret the results.

Table H6-3. Relationship Between Better Performance on the Cost Measure and Quality of Care*

	Quality Performance		
Cost Performance	**Worse**	**Neutral**	**Better**
Better (i.e., Lower Cost)	Somewhat efficient	Efficient care	Efficient care
Neutral	Not efficient care	Somewhat efficient	Efficient care
Worse (i.e., Higher Cost)	Not efficient care	Not efficient care	Somewhat efficient

***Note:** Table H6-3 is just one example developers/stewards may use. Alternative formats (e.g., scatterplots), may also be used. Regardless of the approach chosen, developers/stewards should provide a clear explanation of the method used and guidance on how to interpret the results.

Conclusion

The objective of this guidance is to focus the review of cost measures on a few material assertions, standardizing the claims and evidence presented to substantiate measure properties. By doing so, it aims to streamline the evaluation process, ensuring cost measures are both safe and effective for accountability applications.

Developers also make conformance claims (alignment of measure specification with intent), importance claims (plausible, achievable decrease in costs), reliability and validity claims (a plausible mechanism complex responsible for better performance), and usability claims (facilitators or barriers to implementation of that mechanism complex). These claims, the evidence to support these claims, and the arguments about why the evidence supports these claims are also considered in measure reviews.

By focusing on these key areas, the guidance supports the evaluation of cost measures that enhance health care efficiency and value, ensuring they contribute to improved health care delivery without compromising essential services. This approach fosters a shared understanding among measure developers and E&M committee members, promoting informed decision-making and advancing health care quality and efficiency.

Other Considerations

Recommendations to CMS or other Payers

1. Create an online tool to allow entities to compute standardized prices
2. Price standardization methodology, including codes, be made available in the public domain
3. Develop a Medicare Advantage dollar-weighted services imputation methodology
4. Consider a maximum complexity threshold for the mechanism complex (using tools like the intervention Complexity Assessment Tool for Systematic Reviews [iCAT_SR][22] to

[22] Cadogan CA, Rankin A, Lewin S, Hughes CM. Application of the intervention Complexity Assessment Tool for Systematic Reviews within a Cochrane review: an illustrative case study. HRB Open Res. 2020 Jun 1;3:31. doi: 10.12688/hrbopenres.13044.1. PMID: 32596632; PMCID: PMC7309054

evaluate and set this threshold). The more complex the intervention, the greater the usability challenges, so only develop cost measures below that threshold.

5. Consider an optimal time period based on the care requirements of the condition and to capture long-term adverse effects, specifically focusing on a 1-year timeframe for chronic conditions.

Appendix I: E&M Guidebook Public Comments and Battelle Responses

Overview

Version 3.0 of the E&M Guidebook was posted on the PQM website for public comment on June 10, 2025, for 20 calendar days. During this commenting period, Battelle received comments from nine organizations and three individuals. Public comments focused on the following:

- General E&M processes and policies
- Committee structure and voting outcomes
- Measure-specific guidance - pediatric, cost, and electronic clinical quality measures
- Accountable-entity level testing guidance - reliability and validity
- Closing Care Gaps

All comments received are available on the PQM Website.

Comment Themes

Below is a summary of key themes emphasized by commenters and Battelle's response.

Overall Sentiment

Commenters expressed appreciation for the recent updates to the E&M Guidebook, noting that several revisions incorporated earlier feedback and demonstrated PQM's responsiveness. They highlighted improvements such as increased specificity in reliability rubrics, new appendices on cost measures, and clearer guidance on reliability and validity testing, which were seen as valuable for standardizing and improving the quality of submissions. Commenters also commended Battelle's commitment to gaps in care, as well as enhancements to the voting process that promote transparency and fairness.

Alongside these positive remarks, commenters raised substantial concerns and offered recommendations for further improvement. Key issues included the need for greater transparency in committee processes and decision-making, clearer and more flexible requirements for reliability and validity testing (especially for measures with small denominators or unique populations), and more practical guidance for eCQMs and pediatric measures. There were also calls for better alignment across guidance documents and for ensuring that committee expertise aligns with the complexity of the measures under review.

Overall, while commenters welcomed the Guidebook's progress in responsiveness, clarity, and process improvements, the feedback also reflects a strong desire for continued refinement to ensure the guidance is practical, evidence-based, and transparent.

General E&M Processes and Policies

Several commenters requested more transparency in the development and revision process of the E&M Guidebook by clearly documenting changes, explaining the rationale behind decisions, and

 The analyses upon which this publication (or document) 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. Restricted: Use, duplication, or disclosure is subject to the restrictions as stated in Contract Number 75FCMC23C0010 between the Government and Battelle.

identifying the individuals or groups involved in policy development. Other commenters asked to state all criteria explicitly in the main text rather than in footnotes and to ensure consistency across all related documents (Guidebook, templates, rubrics). Commenters expressed a need for greater clarity regarding the definition of "material change," specifically seeking guidance on when to classify a measure as new versus subject to maintenance review. There were concerns about endorsing measures with significant conditions, particularly in the context of high-stakes accountability, and suggested that measures with major flaws should be revised rather than endorsed with conditions. Additionally, commenters emphasized the importance of considering steward resources and feasibility when establishing timelines for meeting conditions and requested that Battelle present Guidebook updates in a track-changes format to enhance clarity.

Commenters further expressed significant concern about the lack of transparency and openness surrounding the Scientific Methods Panel (SMP) and related decision-making processes. They noted that the SMP is not publicly identified, its members are not listed, and there is no opportunity for public nomination or recruitment. Commenters described the process as "opaque" and not optimal for managing scientific advisory functions. They also criticized the absence of clear documentation on who was consulted, how recommendations were reached, and the rationale behind key decisions. Collectively, these comments call for a more transparent, inclusive, and well-documented process for committee membership and policy development.

We appreciate the feedback provided and are committed to enhancing transparency by documenting changes and providing rationales for decisions. We will ensure that criteria are stated explicitly in the main text with supplemental information. Furthermore, we recognize the importance of balancing priorities and the need for certain changes to require more time for the community to adjust. We are committed to considering steward resources and feasibility when establishing timelines for meeting conditions. We will also seek way to present updates clearly to the future versions of the Guidebook for public comment to improve clarity and facilitate understanding.

Regarding the SMP, we acknowledge the need for greater transparency and will work towards implementing a nomination and selection process similar to other panels. This will include posting the roster and providing more insight into the topics discussed by the SMP.

Committee Structure and Voting Outcomes

Some commenters supported requiring clear rationales for non-endorsement votes and suggested reconsidering maintenance measures that receive substantial support. Others questioned the rationale for lowering the consensus threshold for small committees and recommended expanding the committee size instead. Additionally, commenters recommended ensuring that members of the Cost and Efficiency Committee possess appropriate expertise in cost measures and payment policies.

We appreciate the feedback and are committed to enhancing the transparency and effectiveness of our committee structure and voting processes. To align with the American National Standards Institute (ANSI) standards, we will collect reasons from Recommendation Group members voting "Do Not Endorse/Remove Endorsement." This practice supports ensures that concerns are captured effectively while maintaining confidentiality.

We agree with the suggestion to expand the committee size and will continue to implement mitigation strategies to increase

Recommendation Group numbers to be closer to 25 people. This includes recruiting members from the Advisory Group and moving members to inactive status if they have not been engaged throughout the E&M cycle. With respect to the expertise on the Cost and Efficiency Committee, we appreciate this comment and seek to seat individuals on this committee who possess relevant expertise, including clinicians and patients with lived experience for specific conditions. As with our Guidebook, we post draft committee rosters for public comment. We welcome comments on those rosters based on any gaps identified. Further, we recognize that developers of cost measures convene Technical Expert Panels (TEPs) to help advise on their measures. We encourage developers to nominate those individuals to serve on E&M committees to provide their expertise.

Measure-Specific Guidance - Pediatric, Cost, or Electronic Clinical Quality Measures

Regarding pediatric measures, several commenters strongly opposed extrapolating adult measures to pediatric populations, emphasizing important physiological and clinical differences. They also criticized the use of generative artificial intelligence (AI) in drafting pediatric policy and called for dedicated pediatric-specific development and testing. Commenters appreciated the enhancements made to the guidance related to pediatric quality measures, emphasizing the importance of addressing gaps in care and providing detailed guidance on accountable entity-level reliability and validity testing.

For cost measures, commenters requested clearer and more actionable guidance, including explicit evaluation rubrics and the use of appropriate terminology, such as “performance” instead of “quality.” They highlighted the need to recognize the policy context in which cost measures are used and called for committee members with relevant expertise in cost measurement. Additional feedback suggested identifying and describing known variations in health care and health outcomes related to the measure focus area, including demographic characteristics and groups historically facing barriers to healthcare access.

Regarding eCQMs, commenters opposed the automatic endorsement of new eCQM versions of existing measures. They stressed the need for separate evaluations due to differences in data sources and implementation challenges. Commenters raised concerns about the sufficiency of testing requirements, such as the number of sites or vendors involved, and highlighted the practical challenges developers face, particularly during initial and maintenance endorsement. They recommended greater flexibility in eCQM testing requirements to better align with practical realities while maintaining methodological rigor.

We appreciate the feedback and are committed to refining our guidance to better support measure developers. Regarding pediatric measures, we acknowledge the concerns about extrapolating adult measures to pediatric populations due to physiological and clinical differences. If substantial differences exist in areas such as target population, mechanism complexity, and material outcomes, additional pediatric data collection may be required before extrapolation is deemed valid. While pharmacokinetic modeling focuses on physiological and pharmacokinetic extrapolation, we have shifted to care process modeling. With respect to the use of AI, we recognize the growing role of AI in our field and are committed to using it responsibly. AI supported the drafting of pediatric extrapolation guidance, but we engaged pediatric CQM developers for input and updated our AI disclosure to reflect expert review. We will continue to ensure transparency and qualified expertise in the process.

For cost measures, commenters requested clearer and more actionable guidance, including explicit evaluation rubrics and appropriate terminology, such as "performance" instead of "quality." The current PQM Measure Evaluation Rubric is applicable for evaluating cost measures for endorsement. For the Importance domain, cost measures need a logic model showing the linkage of inputs and actions to achieve the measure focus, supported by evidence and expert input, including from the patient community. We agree that for cost measures, specifically, the guidebook and submission materials should note improvement in performance rather than quality. As we head into the Spring 2026 cycle, the submission materials will be updated accordingly.

Regarding eCQMs, we have corrected the automatic endorsement policy, stating that new eCQM versions of existing measures are not automatically endorsed. To establish that eCQMs are safe and effective, E&M focuses on Feasibility, Importance, and Scientific Acceptability (i.e., Reliability and Validity). Developers/stewards must justify how test sites are distinct to demonstrate broader eCQM implementation. Testing should balance representative and rigorous evaluation with practical considerations such as implementation burden, cost, and alignment with current healthcare quality priorities. For initial endorsement, feasibility and person- or encounter-level testing must be conducted within at least one EHR vendor. Broader implementation should consider testing across multiple vendors and sites. For maintenance evaluations, data from a minimum of five sites within at least two EHR vendors is required, prioritizing site and patient heterogeneity during testing.

Lastly, with respect to the cost-quality trade-off table, this is not required. We provide guidance that offers flexibility, allowing developers to use alternative approaches like scatterplots, with transparency and justification for their methods.

Accountable-Entity Level Testing Guidance - Reliability and Validity

Commenters expressed concern about inconsistencies and a lack of clarity in the reliability criteria, particularly regarding the thresholds for what constitutes an acceptable level of reliability. They noted contradictions between different sections of the Guidebook and questioned the rationale for using specific cutoffs, such as requiring a reliability score of at least 0.6 for 70% of accountable entities. Commenters emphasized that rigid application of these thresholds could exclude valuable measures, particularly those used by smaller entities or in cases where risk adjustment and broader inclusion may lower reliability scores. They urged Battelle to provide more flexibility and a nuanced approach that considers the context and tradeoffs in measure development.

Additionally, commenters highlighted the need for mitigation strategies to improve measure reliability, such as increasing case volume or including more than one year of data. They noted that relying on data from large health systems may not provide insight into the experiences of under-resourced or rural areas, leaving a gap in understanding and applicability of the measure across different settings.

Regarding testing methods, commenters generally supported the adoption of new methodologies, such as the Nieser and Harris approach, but requested more explanation about the reasons for these changes and practical guidance for implementation. They also inquired whether the continued use of established methods, such as the Adams approach, would remain permissible. Commenters highlighted the need for clear instructions on required analyses, especially when person-level data may not be available, and noted that

 The analyses upon which this publication (or document) 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. Restricted: Use, duplication, or disclosure is subject to the restrictions as stated in Contract Number 75FCMC23C0010 between the Government and Battelle.

additional requirements could increase the burden on measure developers.

Regarding validity, commenters highlighted the absence of objective, transparent criteria and called for clearer definitions and processes. They requested explicit guidance on the use of face validity and greater transparency in the process of determining validity. Commenters emphasized the importance of assessing the impact of missing data and measure exclusions as part of the validity testing methodology.

We sincerely thank all commenters for their valuable feedback on the Accountable Entity Level Testing Guidance, particularly regarding reliability and validity. Your insights are crucial in refining our processes and ensuring that our guidance is both practical and effective.

Regarding reliability criteria, we acknowledge the concerns about inconsistencies and the lack of clarity in thresholds. The expected estimates listed were informed by prior discussions at the Scientific Methods Panel (SMP) and previous measure evaluation ratings under the Battelle process. These thresholds have not been treated as absolute requirements. For example, if an accountable entity scores below 0.6, a strict application would mean rating reliability as "Not Met." Instead, we have utilized "Not Met, but Addressable" (NMBA) ratings to identify areas for improvement while acknowledging practical constraints. This approach balances current rulemaking with the understanding that not all entities will reach a reliability of 0.6. Until these criteria are enforced, NMBA ratings will be provided unless there are clear reasons for a "Not Met" rating, such as when 70% of entities fall below the threshold.

We appreciate the comments regarding small denominators and have previously provided various webinars, including our Measure Developer Workshops (MDW), on potential mitigations developers may consider for improving reliability estimates for entities with lower case volumes. Strategies such as conducting descriptive analyses for entities scoring lower than 0.6 can help identify factors contributing to reliability issues and inform mitigation strategies. These strategies have been included in the E&M Guidebook.

Regarding the adoption of new methodologies, such as the Nieser and Harris approach, we clarify that these are not required. Developers may consider using these approaches to show the distribution of reliability estimates at the accountable entity level. We have also removed the statement regarding Inter-Unit Reliability (IUR) from the guidance and encourage developers to add justification and explanation for all methods they use, including IUR.

For validity, we recognize the need for clearer definitions and processes and continually seek to provide clarity and additional guidance. To that regard, we plan to work with the SMP to advise on guidance for face validity testing to support measure developers and endorsement reviews. Assessing the impact of missing data and measure exclusions is an important part of the validity testing methodology and developers should incorporate this within their measure submission.

Finally, for the split-half with Spearman-Brown method, entity-level reliability can be estimated by applying the Spearman-Brown prophecy formula to the Intraclass Correlation Coefficient (ICC) based on the size of each entity. Developers should first conduct the split-half and ICC and then apply the Spearman-Brown formula.

Closing Care Gaps

Commenters expressed strong support for Battelle's commitment to alleviating care

gaps among patient subpopulations. However, they cautioned against making the “Closing Care Gaps” domain a mandatory requirement before the necessary data infrastructure and workflows are in place. They warned that requiring this domain too soon could unintentionally disqualify high-value clinical measures, particularly those that lack comprehensive demographic data due to current limitations in data collection systems. They requested the Closing Care Gaps domain account for available data capture in the context of each specific measure. Commenters recommended that the domain be encouraged rather than required in the near term and that flexibility be maintained to account for the varying data capabilities of different measure types.

We sincerely thank all commenters for their strong support of Battelle’s commitment to alleviating care gaps among patient subpopulations. Your feedback is invaluable as we strive to advance our measure endorsement processes.

Recognizing the constraints expressed by developers, the "Closing Care Gaps" domain will continue to be optional for the Spring 2026 and Fall 2026 cycles. We understand the concerns regarding the proposal to mandate this domain, particularly given the current limitations in data collection systems and infrastructure. We will continue to work closely with the measurement community to develop practical guidance that supports the closing of care gaps without imposing undue burdens on measure developers.

 The analyses upon which this publication (or document) 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. Restricted: Use, duplication, or disclosure is subject to the restrictions as stated in Contract Number 75FCMC23C0010 between the Government and Battelle.

Appendix J: Acronyms

Please note: The following list encompasses acronyms that Battelle commonly encounters and uses in its work as a CBE. Not all the acronyms will appear in this document.

Acronym	Definition
ACA	Affordable Care Act
ACC	American College of Cardiology
ACO	Accountable Care Organization
AGC	After Government Contract
AHIP	Formerly known as American Health Insurance Partnership
AHRQ	Agency for Healthcare Research and Quality
AI Pilot	Artificial Intelligence Pilot
AIPAC	Advanced Illness and Post-Acute Care
AIR	American Institutes for Research
ANOVA	Analysis of Variance
ASCO	American Society of Clinical Oncology
ASCQR	Ambulatory Surgical Center Quality Reporting Program
ASCs	Ambulatory Surgical Centers
C&E	Cost and Efficiency
CAH	Critical Access Hospital
CAHPS	Consumer Assessment of Healthcare Providers and Systems
CBE	Consensus-Based Entity
CBE ID	Consensus-Based Entity Identification
CDC	Centers for Disease Control and Prevention
CDS	Clinical Decision Support
CDSS	Clinical Decision Support System
CIS	Clinical Information Systems
CMIT	CMS Measures Inventory Tool
CMMI	Center for Medicare and Medicaid Innovation
CMS	Centers for Medicare & Medicaid Services
CO	Contracting Officer
COIs	Conflicts of Interest
COR	Contracting Officer's Representative
CPG	Clinical Practice Guidelines
CQL	Clinical Quality Language

 The analyses upon which this publication (or document) 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. Restricted: Use, duplication, or disclosure is subject to the restrictions as stated in Contract Number 75FCMC23C0010 between the Government and Battelle.

Acronym	Definition
CQM	Clinical Quality Measure
CQMC	Core Quality Measures Collaborative
CSAC	Consensus Standards Approval Committee
DEL	CMS Data Element Library
Del.	Deliverable
DOI	Disclosure of Interest
dQMs	Digital Quality Measures
DRC	Direct Reference Code
E&M	Endorsement and Maintenance
EC	Electronic Copy
eCQI	Electronic Clinical Quality Improvement
eCQM	Electronic Clinical Quality Measures
ED	Emergency Department
EHR	Electronic Health Record
EPC	Evidence-Based Practice Center
ESRD QIP	End-Stage Renal Disease Quality Improvement Program
EVI	Expected Value of Information
FAQs	Frequently Asked Questions
FFS	Fee-For-Service
FHIR	Fast Healthcare Interoperability Resources
FMS	Full Measure Submission
FY	Fiscal Year
HACRP	Hospital-Acquired Conditions Reduction Program
HCBS	Home and Community-Based Services
HCD	Human-Centered Design
HEDIS	Healthcare Effectiveness Data and Information Set
HH QRP	Home Health Quality Reporting Program
HH VBP	Home Health Value-Based Purchasing
HHS	Department of Health and Human Services
HIQR	Hospital Inpatient Quality Reporting
HOPD	Hospital Outpatient Department
HOPE	Hospice Outcomes and Patient Evaluation
HOQR	Hospital Outpatient Quality Reporting
HQMF	Health Quality Measurement Format

 The analyses upon which this publication (or document) 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. Restricted: Use, duplication, or disclosure is subject to the restrictions as stated in Contract Number 75FCMC23C0010 between the Government and Battelle.

Acronym	Definition
HQR	Hospice Quality Reporting
HQRP	Hospice Quality Reporting Program
HRRP	Hospital Readmission Reduction Program
HSAG	Health Services Advisory Group
HTML	Hypertext Markup Language
HVBP	Hospital Value-Based Purchasing
IAW	In Accordance With
ICD	International Classification of Diseases (International Statistical Classification of Diseases and Related Health Problems)
IHI	Institute for Healthcare Improvement
IMPACT Act	Improving Medicare Post-Acute Care Transformation Act
IPF	Inpatient Psychiatric Facilities
IPF QRP	Inpatient Psychiatric Facility Quality Reporting Program
IPPS	Inpatient Prospective Payment System
IQR	Inpatient Quality Reporting
IR	Initial Recognition
IRF	Inpatient Rehabilitation Facilities
IRF QRP	Inpatient Rehabilitation Facility Quality Reporting Program
IT	Information Technology
ITS	Intent to Submit
LLMs	Large Language Models
LTACH	Long-Term Acute Care Hospitals
LTCH	Long-Term Care Hospital
LTCH QRP	Long-Term Care Hospital Quality Reporting Program
MA	Medicare Advantage
MACRA	Medicare Access and CHIP Reauthorization Act
MACS	Medicaid: Adult Core Set
MAQIP	Medicare Advantage Quality Improvement Program
MAT	Measure Authoring Tool
MCCS	Medicaid: Child Core Set
MCO	Managed Care Organization
MERIT	Measures Under Consideration Entry/Review Tool
MIPPA	Medicare Improvement for Patients and Providers Act of 2008
MIPS	Merit-based Incentive Payment System

 The analyses upon which this publication (or document) 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. Restricted: Use, duplication, or disclosure is subject to the restrictions as stated in Contract Number 75FCMC23C0010 between the Government and Battelle.

Acronym	Definition
MLTSS	Managed Long-Term Service and Support
MMS	Measures Management System
MS-DOI	Measure-Specific Disclosure of Interest
MSR	Measure Set Review
MSSP	Medicare Shared Savings Program
MUC	Measures Under Consideration
n	Sample Size
NCDC	National Consensus Development and Strategic Planning for Health Care Quality Measurement Contract
NCQA	National Committee for Quality Assurance
NHDNG	Novel Hybrid Delphi and Nominal Groups
NHQI	Nursing Home Quality Initiative
NLP	Natural Language Processing
NQF	National Quality Forum
NQS	CMS National Quality Strategy
NTTAA	National Technology Transfer and Advancement Act
OMB	Office of Management and Budget
OP	Option Period
OY	Option Year
PA	Preliminary Assessment
PAC/LTC	Post-Acute Care/Long-Term Care
PaLS	Patient Life Goals Survey
PAM	Patient Activation Measure
PCHQR	PPS-Exempt Cancer Hospital Quality Reporting
PDF	Portable Document Format
PIE Form	Pre-Meeting Initial Evaluation Form
PL	Project Leader
PM	Project Manager
PMP	Project Management Plan
POC	Point of Contact
PPS	Prospective Payment System
PQA	Pharmacy Quality Alliance
PQM	Partnership for Quality Measurement
PRA	Paperwork Reduction Act

 The analyses upon which this publication (or document) 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. Restricted: Use, duplication, or disclosure is subject to the restrictions as stated in Contract Number 75FCMC23C0010 between the Government and Battelle.

Acronym	Definition
PRMR	Pre-Rulemaking Measure Review
PRO	Patient-Reported Outcome
PROM	Patient-Reported Outcome Measure
PRO-PMs	Patient-Reported Outcome Performance Measures
Q&A	Question & Answer
QC	Quality Control
QCDR	Qualified Clinical Data Registries
QDM	Quality Data Model
QI	Quality Improvement
QMDSA	Quality Measure Developer and Steward Agreement
QPP	Quality Payment Program
REHQR	Rural Emergency Hospital Quality Reporting (Program)
SDOH	Social Determinants of Health
SES	Socioeconomic Status
SLIN	Subline Item Number
SMEs	Subject Matter Experts
SMP	Scientific Measures Panel
SNF	Skilled Nursing Facilities
SNF QRP	Skilled Nursing Facility Quality Reporting Program
SNF VBP	Skilled Nursing Facility Value-Based Purchasing
SOP	Standard Operating Procedure
SOW	Statement of Work
SSA	Social Security Administration
STAR	Submission Tool and Repository
SUD	Substance Use Disorder
TBD	To Be Determined
TEP	Technical Expert Panel
TL	Task Lead
UMLS	Unified Medical Language System
USCDI	United States Core Data for Interoperability
VSAC	Value Set Authority Center
Yale CORE	Yale Center for Outcomes Research and Evaluation


Partnership for
Quality Measurement
Powered by Battelle