

PQM Endorsement: From Criteria to Consensus

Anna Michie, E&M Deputy Task Lead
Brenna Rabel, CBE Technical Director
Kirsten Barrett, E&M Technical SME

April 23, 2026

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.

Meet the E&M Presenters



Anna Michie | E&M Deputy Task Lead



- Provides strategic and technical support on E&M processes and activities
- 10+ years' quality experience

Brenna Rabel | CBE Technical Director



- Facilitates collaboration across CBE activities to ensure consistency and excellence
- 10+ years' quality experience

Kirsten Barrett | E&M Technical SME



- Provides technical expertise and support on E&M processes and activities
- 10+ years' quality experience

Meeting Purpose and Agenda



Purpose

To clarify the E&M process and criteria for committee members and measure developers/stewards.

Agenda

- Overview of the PQM Endorsement Framework
- What is a Staff Preliminary Assessment?
- Holistic Review in Practice
- Key Takeaways
- Q&A and Feedback

Overview of the PQM Endorsement Framework



The Partnership for Quality Measurement (PQM)

Powered by Battelle



Battelle is a consensus-based entity (CBE) funded through the *CMS National Consensus Development and Strategic Planning for Health Care Quality Measurement Contract* to oversee the endorsement and maintenance of clinical quality measures.



Battelle convenes PQM committee members to evaluate quality measures submitted for endorsement or up for routine maintenance. Committee members answer the question: ***is the measure safe and effective for general use and unlikely to result in negative unintended consequences?***

The quality measure endorsement and review processes should be reliable, transparent, attainable, balanced, and, most of all, meaningful.

Safety and Effectiveness



- The purpose of the CBE E&M process is to endorse measures that are **safe and effective**—meeting standards for importance, scientific acceptability, feasibility, and usability—through a transparent, consensus-based review.



By “safe and effective” we mean that:



The measure is consistent with current professional knowledge;



The measure’s use is likely to improve desired health outcomes (net benefit); and



The measure’s use is not likely to increase the risk of unintended or adverse health outcomes.

Measure Review from All Angles

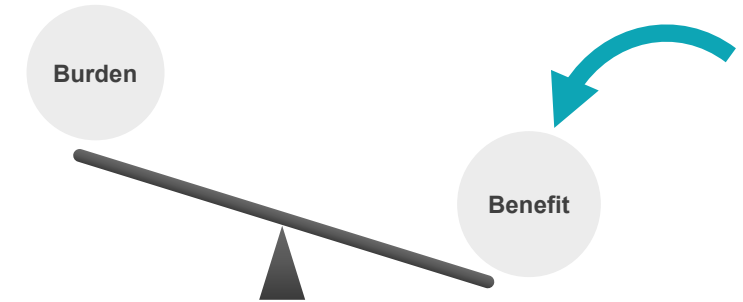


- Endorsement is a holistic determination of safety and effectiveness.
- Committees weigh the totality of evidence and context to decide whether using a measure is likely to improve care without creating undue burden or harm and whether the measure should be endorsed and maintained for accountability.
- Endorsement decisions consider measure performance across evidence, context, and implementation, rather than isolated criteria.
 - There are no must-pass criteria.
- Committees reach endorsement decisions via a transparent, consensus-based process.

Holistic Review is a Risk-Benefit Assessment



Focus quality measurement where there is **the most benefit** for health care system change



		RISK →	
		Risk of measurement	
Impact of measurement		Low uncertainty (Mechanisms are systemic and persistent; evidence is mature)	High uncertainty (Mechanisms are not systemic and persistent; evidence is not mature)
IMPACT ↓	Low (few persons and entities) (Magnitude of improvement to benchmark is low; magnitude of mechanism effect is low)	Do not measure (accept the risk of low quality)	Quality improvement (transfer the risk of low quality)
	High (many persons and entities) (Magnitude of improvement to the benchmark is high; magnitude of mechanism effect is high)	Mitigation or monitoring (control the risk of low quality)	Quality measurement (avoid the risk of low quality)

Holistic Review is a Risk-Benefit Assessment (*cont., 1*)



Measure developers and/or measure stewards make certain explicit or *implicit* assertions or claims about the potential benefits and risks/harms associated with measure use (net benefit).

In general, there are three top-level claims related to measure properties necessary for a measure to yield positive net benefit to persons and entities:

Decision Relevance

Would a person or entity realistically make decisions based on the measure because the measure focus is associated with a material outcome?

Scientific Appropriateness

Should a person or entity use this measure because there are known, effective ways to improve performance?

- Is the mechanism understood and causally credible?

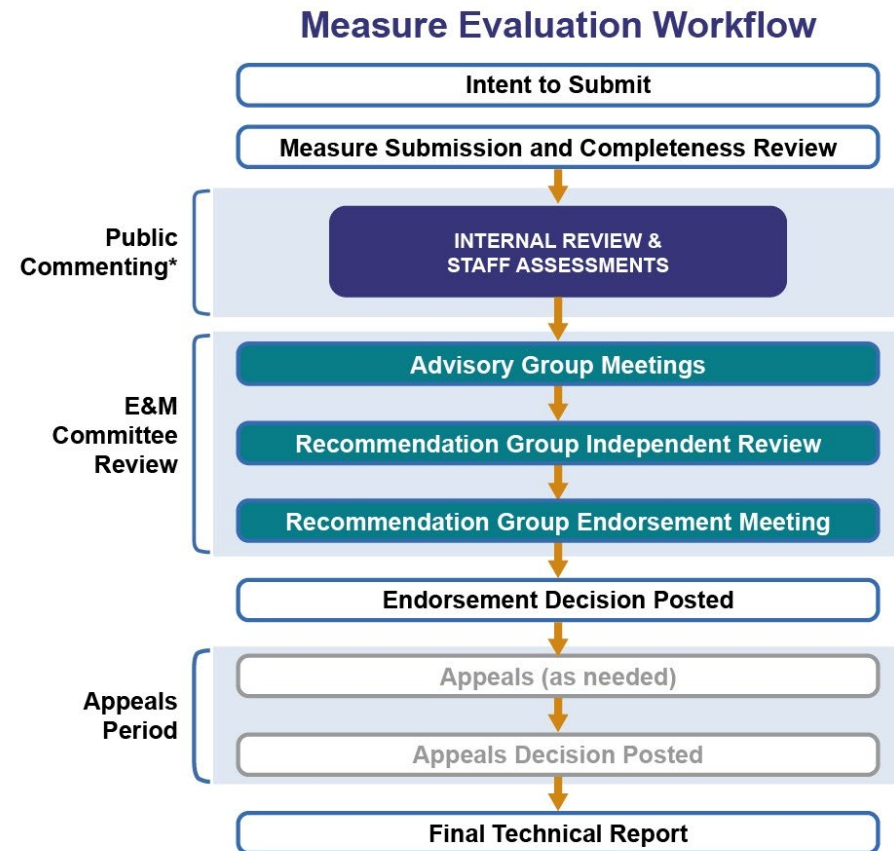
Practical Feasibility and Use

Could a person or entity implement and use this measure given real-world constraints such as burden, data availability, and workflow?

Understanding the Measure Evaluation Workflow



1. Steward/developer submits measure.
2. Battelle posts measure for public comment.
3. Battelle staff conducts internal review (i.e., staff assessment).
 - The Advisory Group and Recommendation Group use these assessments as supplemental inputs.

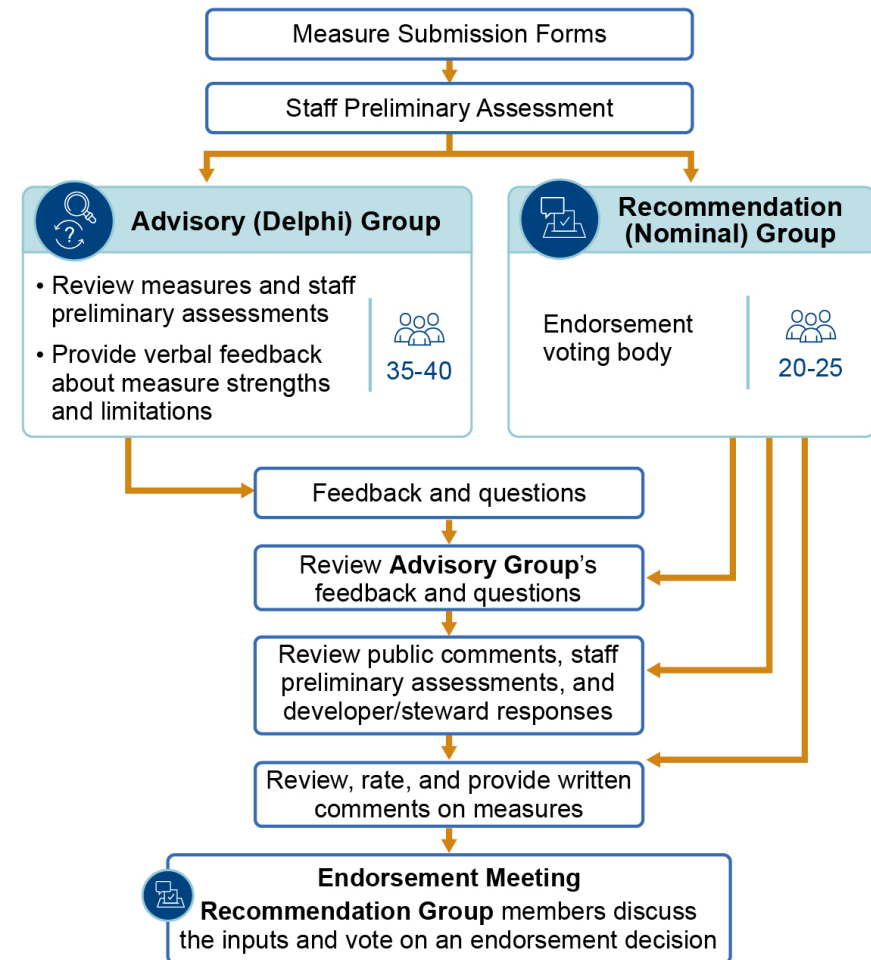


** Members of the public may also provide verbal comments during Advisory Group Meetings.*

Understanding the Measure Evaluation Workflow (cont., 1)

4. Committee review:

- The Advisory Group meets first.
- Battelle staff compile all input from throughout the cycle (i.e., Advisory Group feedback, public comment, staff preliminary assessments, and Recommendation Group members' independent reviews) to support discussion during endorsement meetings.
- The Recommendation Group evaluates all the inputs to inform the final endorsement decision.



What is a Staff Preliminary Assessment?



Purpose of Staff Preliminary Assessments



Staff preliminary assessments are a best practice for providing structure to the holistic evaluations.



Battelle staff create structured preliminary reviews using PQM measure evaluation criteria ([see slide 18](#)) to:

- Identify measure strengths and limitations on which the committee may want to focus their review.
- Assign a preliminary rating (e.g., Met; Not Met but Addressable; or Not Met) for each PQM measure evaluation criterion.



Preliminary ratings serve as **inputs** for committee consideration and discussion, not a final judgment.

- Committee members are expected to review the measure submissions on their own, to understand the evaluation criteria, and to apply the criteria to determine their endorsement decisions.
- Staff preliminary assessments are not a pass/fail decision.
- Staff preliminary assessments are not a substitute for reviewing developer submissions.

Staff Preliminary Assessment Example

CBE #5320 Percentage of Chronic Hyperphosphatemia in Dialysis Patients



Use and Usability Rating: Met

Strengths

- The measure is not currently in use, but the developer described a plan for use in public reporting and payment programs.

The developer provided a summary of how accountable entities can use the measure results to improve performance. Specifically, entities can implement **regular dietary counseling** on a low-phosphorus nutrition plan, prescribe phosphorus binding medications, and **optimize dialysis dosage** to mitigate hyperphosphatemia. These possible actions are reflected in the measure's logic model.

Limitations

- The submission **could be strengthened with a discussion of how the developer concluded that no unintended consequences are anticipated** (e.g., discussions with TEP, examining the cited literature).

Rationale

- This new measure is rated 'Met' for use and usability; there is a **plan for use** in at least one accountability application, and the measure provides actionable information for improvement. The developer reported that no potential unintended consequences were identified.

- The assessments identifies key strengths, including evidence of how the measure could improve performance.
- The assessment also highlights limitations to show areas where additional clarification or justification would strengthen the submission.
- A clear rationale connects the strengths and limitations to the rating of “Met” assigned to the domain by Battelle staff.

How Staff Preliminary Assessments Should Be Used By:



Committee Members

1. Quickly frame independent review.
 - Review staff's rating rationale across criteria (Importance, Scientific Acceptability, Feasibility, and Use/Useability) and complete your own preliminary review.
2. Turn staff flags into targeted questions.
 - Convert each staff-identified issue into specific questions to ask developers in the Advisory/Recommendation Group discussion.
3. Compare with public comments for convergence or divergence.
 - Check whether commenters reinforce or rebut staff's concerns and consider where clarification from the developer would resolve confusion.
4. Focus on decisional, in-scope issues, not background.
 - Use the assessment to understand a handful of in-scope issues that could impact your vote.

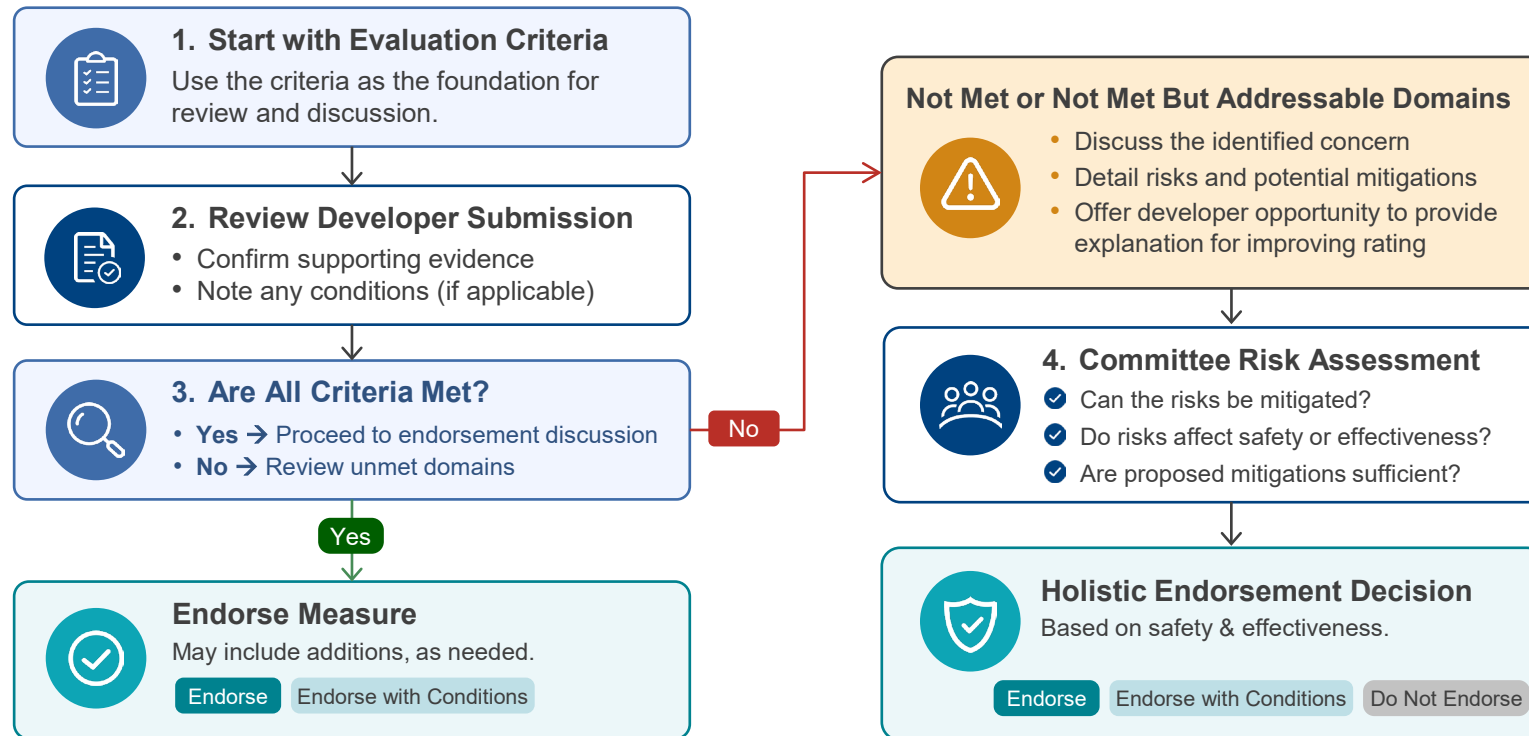
Measure Developers

1. Treat the assessment as a "to do" list to address before committee meetings.
 - Use the staff's domain-by-domain judgements to prioritize clarifications or added analyses and to craft your response for committee review.

Holistic Review in Practice



Measure Evaluation Workflow



Key Principle: Holistic Evaluation for Endorsement or Maintenance

The **committee weighs the totality of evidence** and context to determine whether using a measure is likely to improve care without creating undue burden or harm, and whether **it should be endorsed or maintained for accountability**. There **are no must-pass criteria**, and the decision is made via a **transparent, consensus-based process** with public input.

PQM Evaluation Criteria



1



Importance - Extent to which the measure is evidence based AND is important for making significant gains in health care quality or cost where there is variation in or overall less-than-optimal performance.

2



Closing Care Gaps (optional) - 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.

3



Feasibility - 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.

4



Scientific Acceptability [i.e., Reliability and Validity] - Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented.

5



Use and Usability - 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.

Reviewing the Submission



- Review the entire submission; anchor feedback to evidence in submission materials.
- Weigh strengths against limitations.
- Avoid personal preference as sole justification; reviews should be grounded in the PQM Evaluation Criteria.
- Provide clear rationale when marking “Not Met” for domains; articulate threats to safety and effectiveness and note any potential conditions.

Committee Risk Assessment



- If criteria are not met, review unmet domains.
 - Discuss any identified concerns.
 - Offer the developer the opportunity to provide an explanation for improving the rating.
- Consider mitigations or conditions.
 - Can the risks be mitigated?
 - Does risk impact the safety and effectiveness? Or is the risk minor and could be monitored or further explored in 3-5 years?
- Vote on the overall value and risk-benefit balance of a measure.
 - Measures should be safe for patients and effective for the health care system.



Final Endorsement Decision



Endorse

Substantial net benefit.

Endorse with Conditions

A net benefit exists; however, specific mitigations are needed to enhance the benefit to a substantial level.

Do Not Endorse/Remove Endorsement

No net benefit, or net benefit is low and mitigation will not improve it.

Example # 1: How Criteria Interact

Importance + Feasibility Tradeoff



CBE #4715: Cardiovascular Disease (CVD) Risk Assessment Measure

- Measure title: “Proportion of pregnant/postpartum patients assessed for CVD risk using a standardized tool.”
- **Importance:** Met
 - Clear logic model linking standardized risk assessment to improved detection and outcomes.
 - Addresses a critical safety issue, as delays in recognition and diagnosis often drive CVD-related maternal deaths.
 - Supported by a comprehensive literature review and patient input.
- **Feasibility:** Not Met but Addressable
 - Implementation (manual or electronic) may introduce burden and cost.
 - Electronic health record (EHR) integration requires IT resources; manual implementation requires additional staff for data entry and maintenance.
 - Unclear which data elements are readily available versus those needing incorporation.

Committee Decision:

- Recognized the measure addresses meaningful clinical needs but introduces potential burden, particularly costs related to risk assessment and follow-up testing.
- The developer noted a cost analysis of follow-up testing is currently underway.
- Decision: *Endorsed with Conditions* to balance value with feasibility.
 - Condition: Within 3 years, the developer should conduct outreach and research with smaller and rural entities to assess implementation burden and cost.

Example # 2: How Criteria Interact

Scientific Acceptability with Mitigation Pathways



CBE #1463: Standardized Hospitalization Ratio for Dialysis Facilities

- The ratio of observed to expected hospitalizations among Medicare end-stage renal disease (ESRD) dialysis patients at a facility, adjusted for patient characteristics and national rates.
- **Importance: Met**
 - Clear logic model linking facility activities to desired outcomes.
 - Addresses a high-burden issue, with dialysis patients averaging 9.4 hospital days/year.
 - Supported by a comprehensive literature review, empirical studies, and patient input.
- **Scientific Acceptability (Reliability): Not Met**
 - Fewer than 40% of accountable entities met expected threshold of 0.6
 - Reliability concerns largely driven by small facility sizes and limited sample sizes.
 - Developer identified program-level mitigations (e.g., measure composites, small facility adjusters) to improve reliability and reduce unintended payment impacts.

Committee Decision:

- Recognized the measure addresses an important area for monitoring and improvement.
- Viewed reliability limitations as addressable with appropriate mitigation strategies.
- The developer confirmed the mitigation strategies should be applied in accountability programs.
- Decision: *Endorsed*

Key Takeaways



Key Takeaways



Preliminary staff assessments are a structured input, not a final decision.

Committees assess domains using a holistic review that weighs the totality of evidence.

Criteria application must be evidence based and transparent.

Clear communication improves fairness and efficiency.

Q&A and Feedback



Thank You!

Have questions? Contact us at
PQMsupport@battelle.org





Partnership for
Quality Measurement
Powered by Battelle