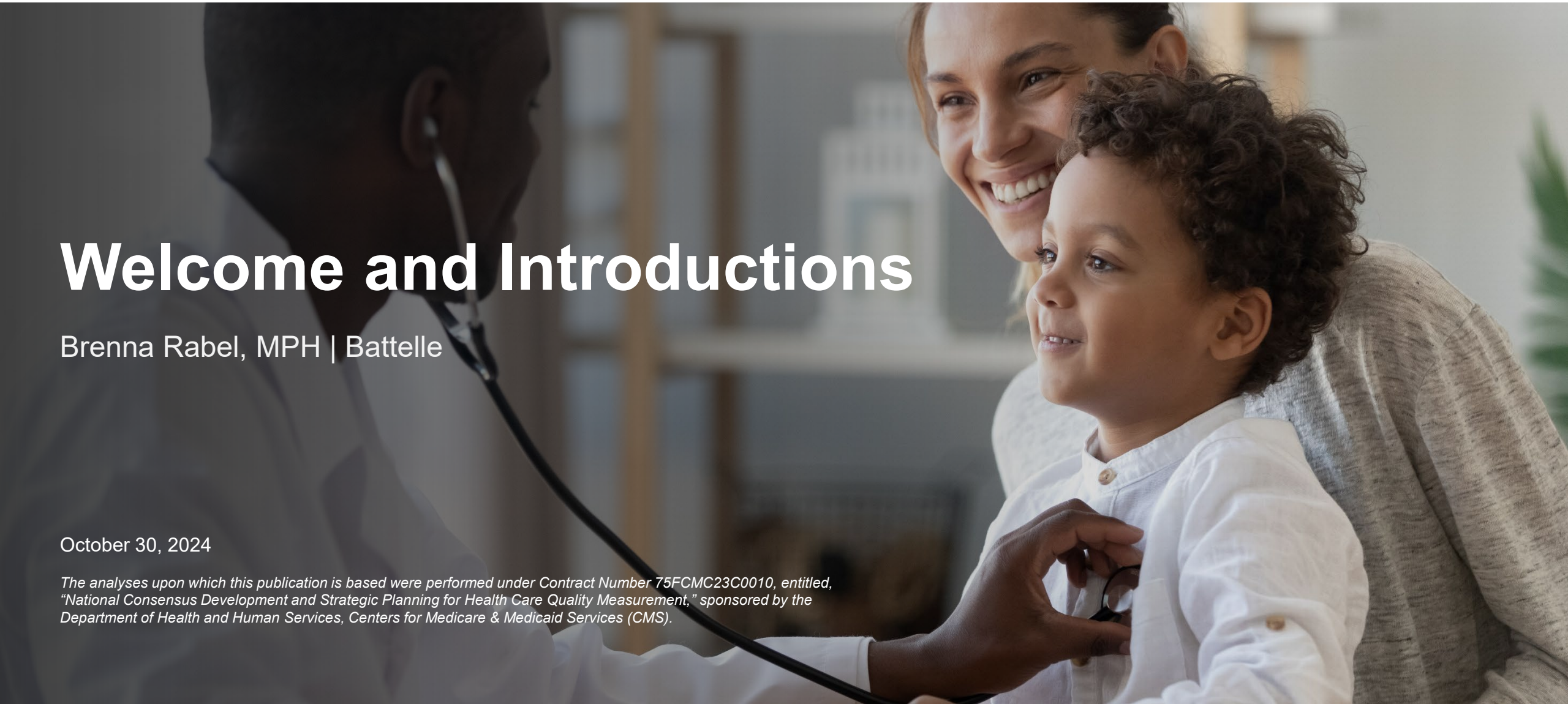
Partnership for Quality Measurement
Powered by Battelle



Welcome and Introductions

Brenna Rabel, MPH | Battelle

October 30, 2024

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

Partnership for Quality Measurement
Powered by Battelle

Welcome

Workshop Objectives

The purpose of today's virtual measure developer workshop is to:

- Describe Battelle's comprehensive strategy for measure endorsement, including sharing best practices for measure submissions.
- Review new guidance for developing robust logic models and elements of the equity domain, including interactive breakout sessions for discussion and shared learning.
- Explore each domain of the PQM Evaluation Rubric, discuss common pitfalls and mitigation strategies, and clarify requirements for new and maintenance measures.
- Discuss policies on instrument-based measures and offer tips for developers to navigate expectations during the endorsement cycle.
- Share and discuss key insights and reflections from the participants.

Agenda

- Morning Sessions
  - 10:15 AM: Comprehensive Evaluation of Measures: A Holistic Approach
  - 10:45 AM: What Good Looks Like
  - 11:15 AM: Importance and Logic Model Guidance (*includes breakout rooms!*)
- Lunch (12:15-12:45 PM)
- Afternoon Sessions
  - 12:45 PM: Feasibility
  - 1:15 PM: Scientific Acceptability: Reliability, Validity, and Risk Adjustment
  - 2:00 PM: Equity Guidance (*includes breakout rooms!*)
  - 3:15 PM: Instrument-Based Measure Guidance
  - 3:45 PM: Use and Usability
  - 4:15 PM: Developer Engagement: What to Expect During the Endorsement Cycle
- Day's End Review: Insights and Reflections, Questions, and Adjourn

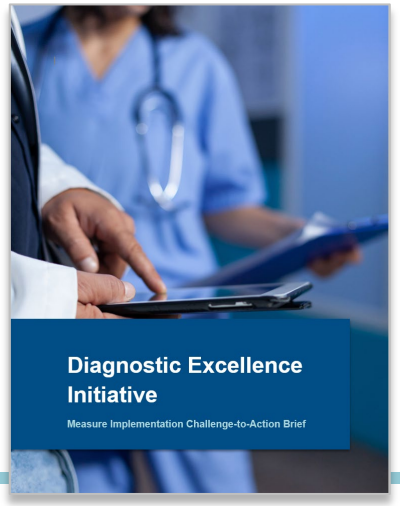
Introduction to Battelle and the Partnership for Quality Measurement (PQM)

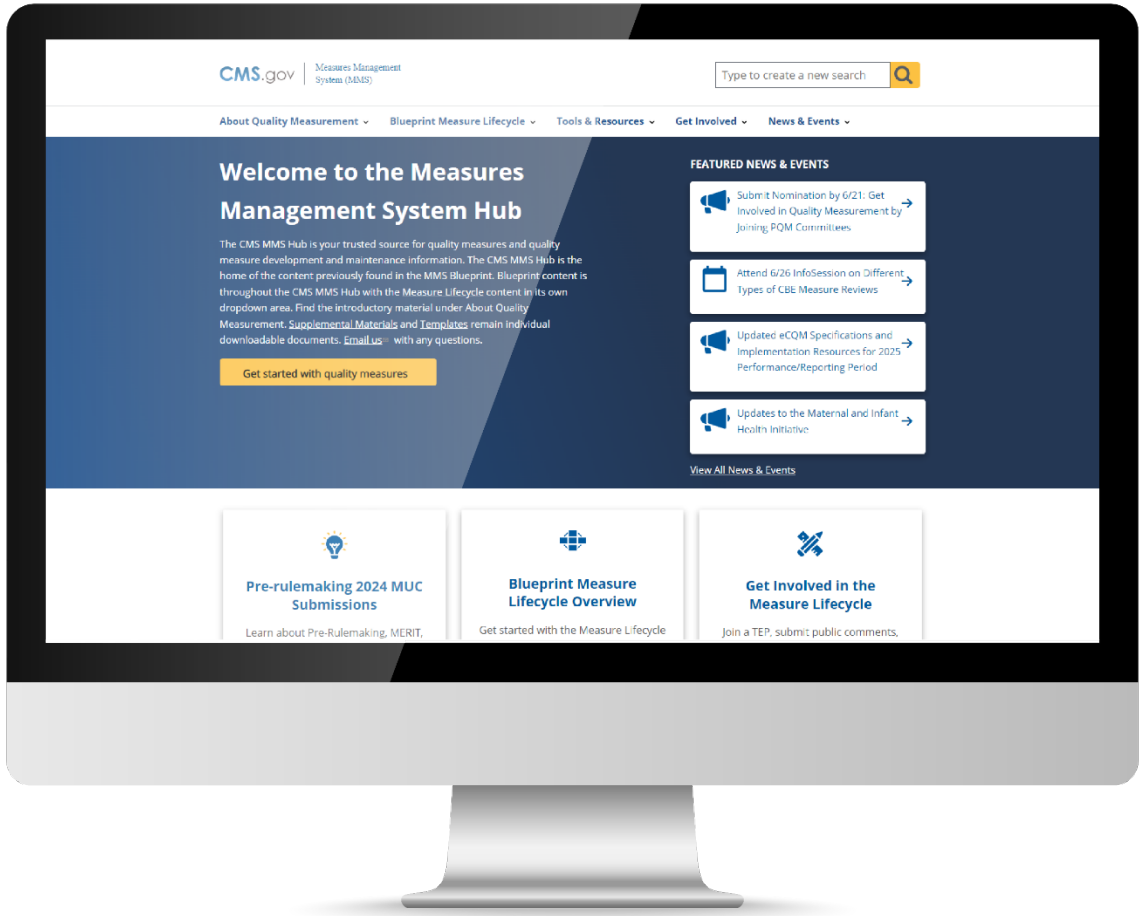
Battelle & Health Care Quality

Battelle is the world's largest independent nonprofit applied science and technology organization.

Over 20 years of contributions and leadership in the science of health care quality measurement including:

- Centers for Medicare & Medicaid Services (CMS) Measures Management System (MMS)
- CMS Blueprint
- Agency for Healthcare Research and Quality (AHRQ) Quality Indicators
- Gordon and Betty Moore Foundation





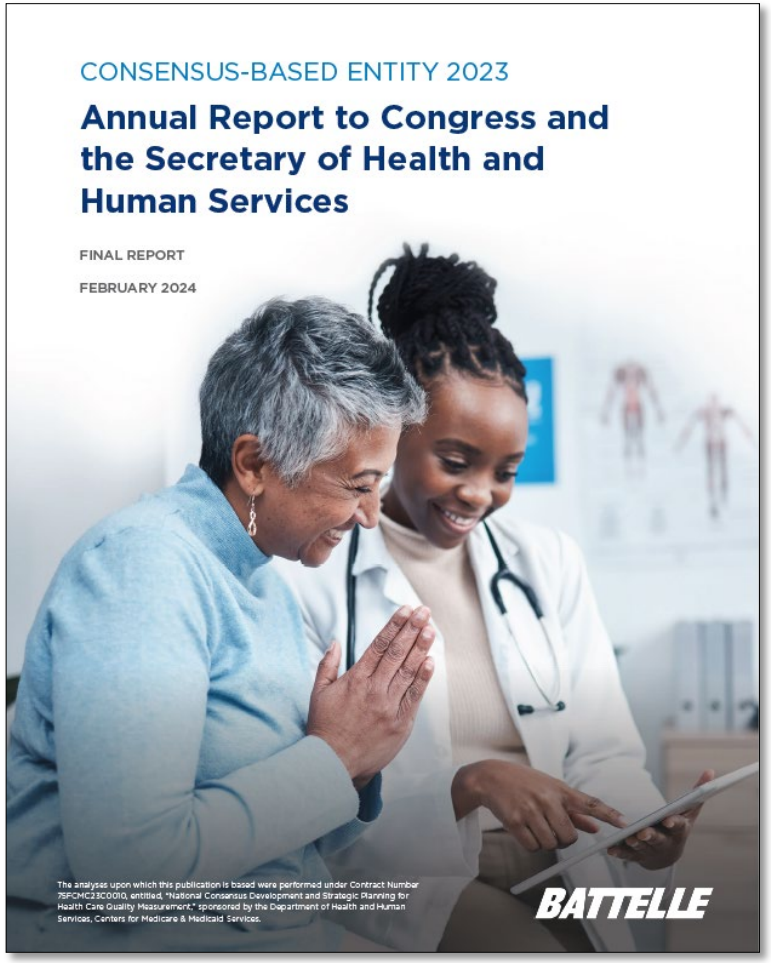
Battelle as a Consensus-Based Entity

CMS-certified consensus-based entity (CBE)

Awarded CMS National Consensus Development and Strategic Planning for Health Care Quality Measurement contract in 2023

Partnership for Quality Measurement

- **Who we are:** Partnership of members across the health care and quality landscape interested in promoting meaningful quality measurement.
- **Vision:** The quality measure endorsement process should be reliable, transparent, attainable, equitable, and, most of all, meaningful.
- **Approach:** Ensure informed and thoughtful endorsement reviews of qualified measures by conducting a consensus-based process involving a variety of experts—clinicians, patients, measure experts, and health information technology specialists.



Battelle Project Team

- Nicole Brennan, MPH, DrPH, Executive Director
- Brenna Rabel, MPH, Deputy Director
- Jeff Geppert, Measure Science Team Lead
- Quintella Bester, PMP, Senior Program Manager
- Matthew Pickering, PharmD, Endorsement and Maintenance (E&M) Task Lead
- Anna Michie, MHS, PMP, E&M Deputy Task Lead
- Beth Jackson, PhD, MA, Social Scientist IV
- Adrienne Cocci, MPH, Social Scientist III
- Stephanie Peak, PhD, Social Scientist III
- Isaac Sakyi, MSGH, Social Scientist III
- Jessica Lemus, MA, Social Scientist II
- Olivia Giles, MPH, Social Scientist I
- Elena Hughes, MS, Social Scientist I
- Sarah Rahman, Social Scientist I

Battelle Project Support Team

Communications

- JJ Knight, MS
- Chauntel Richardson, MPH, CHES

Technical Editing

- Brittany Stojsavljevic
- Catherine McBride, MS

Graphics

- Sarah Kaukeinen

508 Compliance

- Lali Gentry, MPH
- Kelsey Conner

Web Team

- Kim O'Brien, MS, PMP
- Ian Warmbrodt, MBA
- Maureen Hammer, PhD, MS
- Margaret Jokoh

Institute for Healthcare Improvement (IHI)

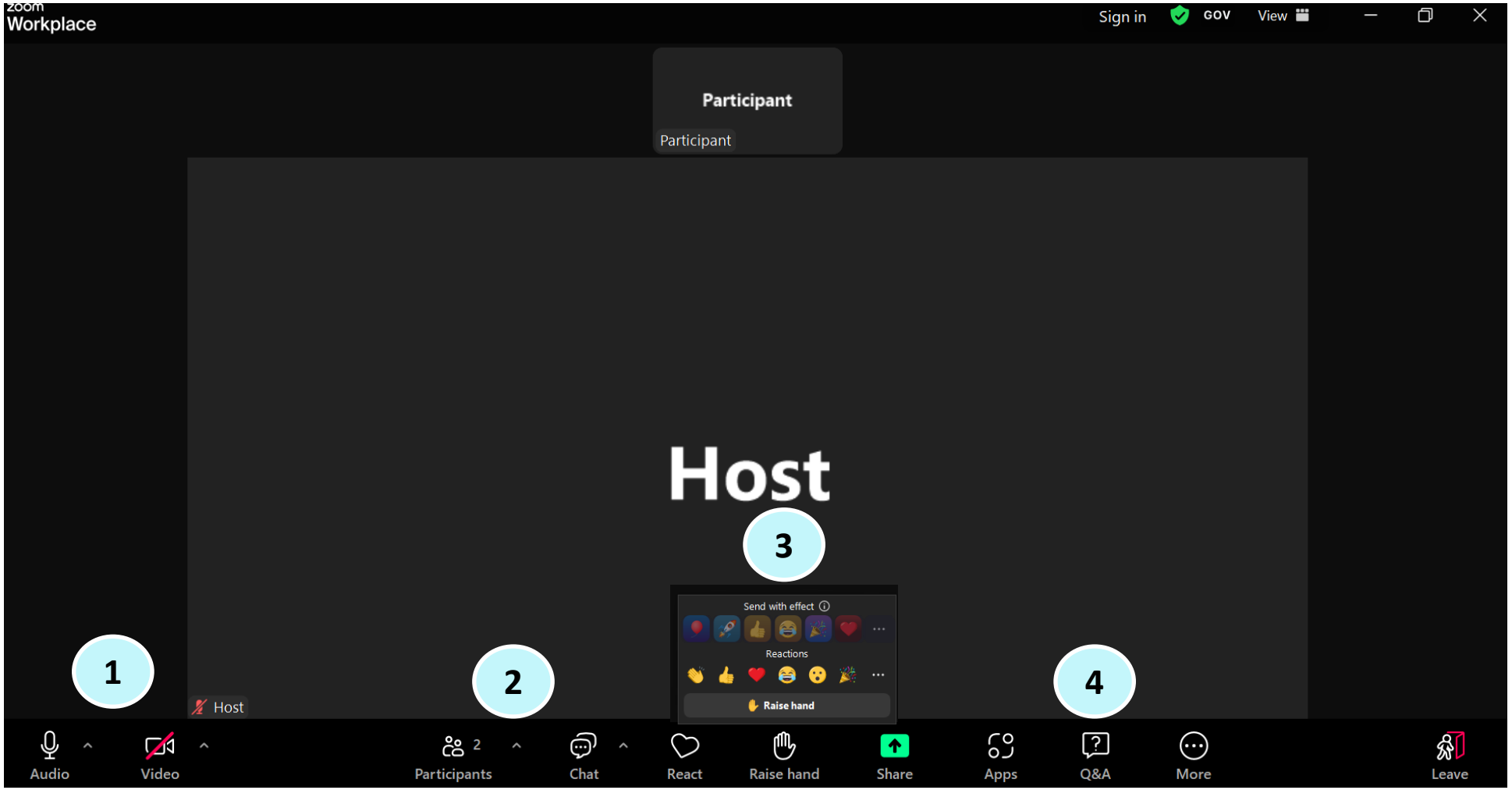
- An independent nonprofit and leading innovator in health and health care improvement worldwide
- Project team and faculty support work across CBE contract
- For E&M, supported guidance for developing logic models and equity

Institute *for* Healthcare Improvement

Housekeeping Reminders

- Each session will have dedicated time for Q&A.
  - Please include questions in the Q&A box, and Battelle staff will triage at the end of each session.
  - There is also time at the end of the day for additional Q&A.
- Presenters will provide any additional instructions for breakout sessions.
  - The system will allow you to mute/unmute yourself and turn your video on/off.
- If you are experiencing technical issues, please contact the project team via chat on the virtual platform or at PQMsupport@battelle.org.

Using the Zoom Platform



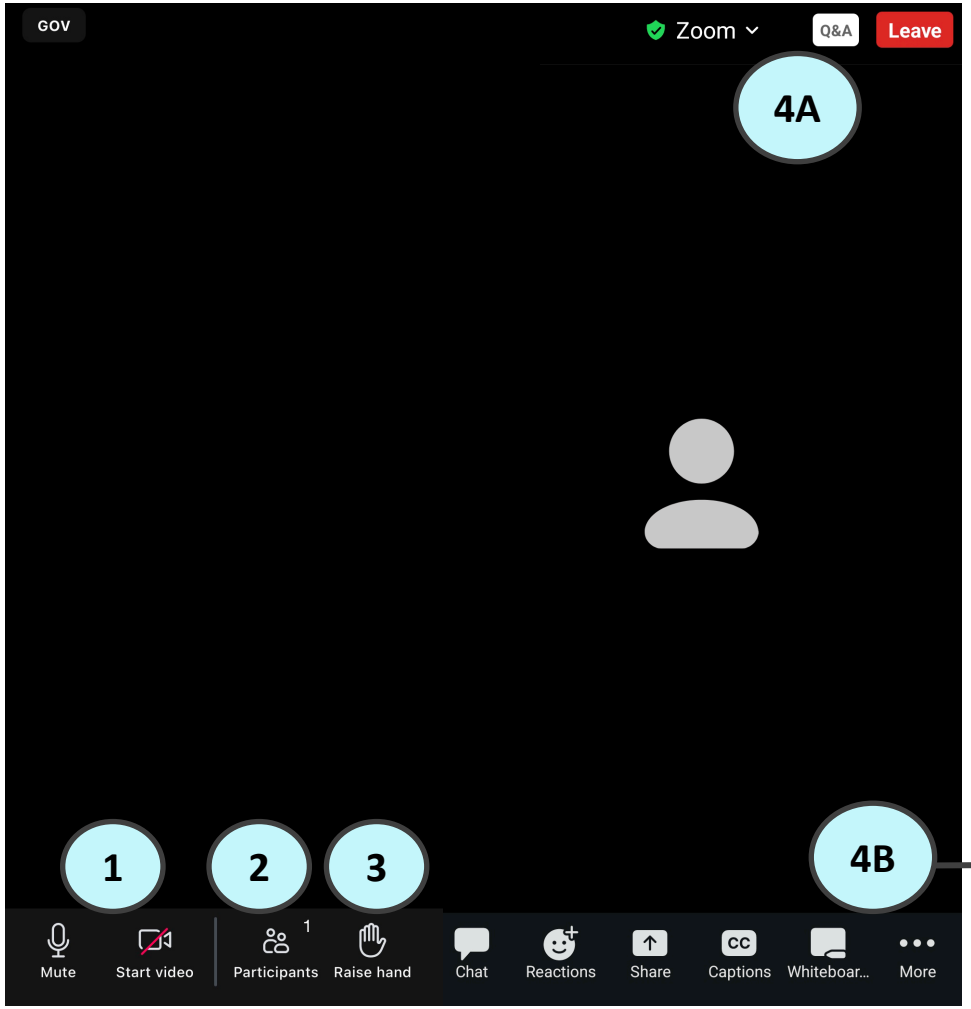
1. Click the lower part of your screen to mute/unmute or start or pause video.

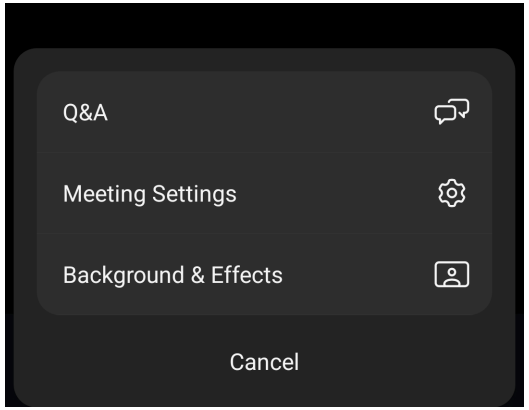
2. Click on the participant or chat button to access the full participant list or the chat box.

3. To raise your hand, select the raise hand button under the react tab.

4. To ask a question, use the Q&A button.

Using the Zoom Platform (Mobile View)



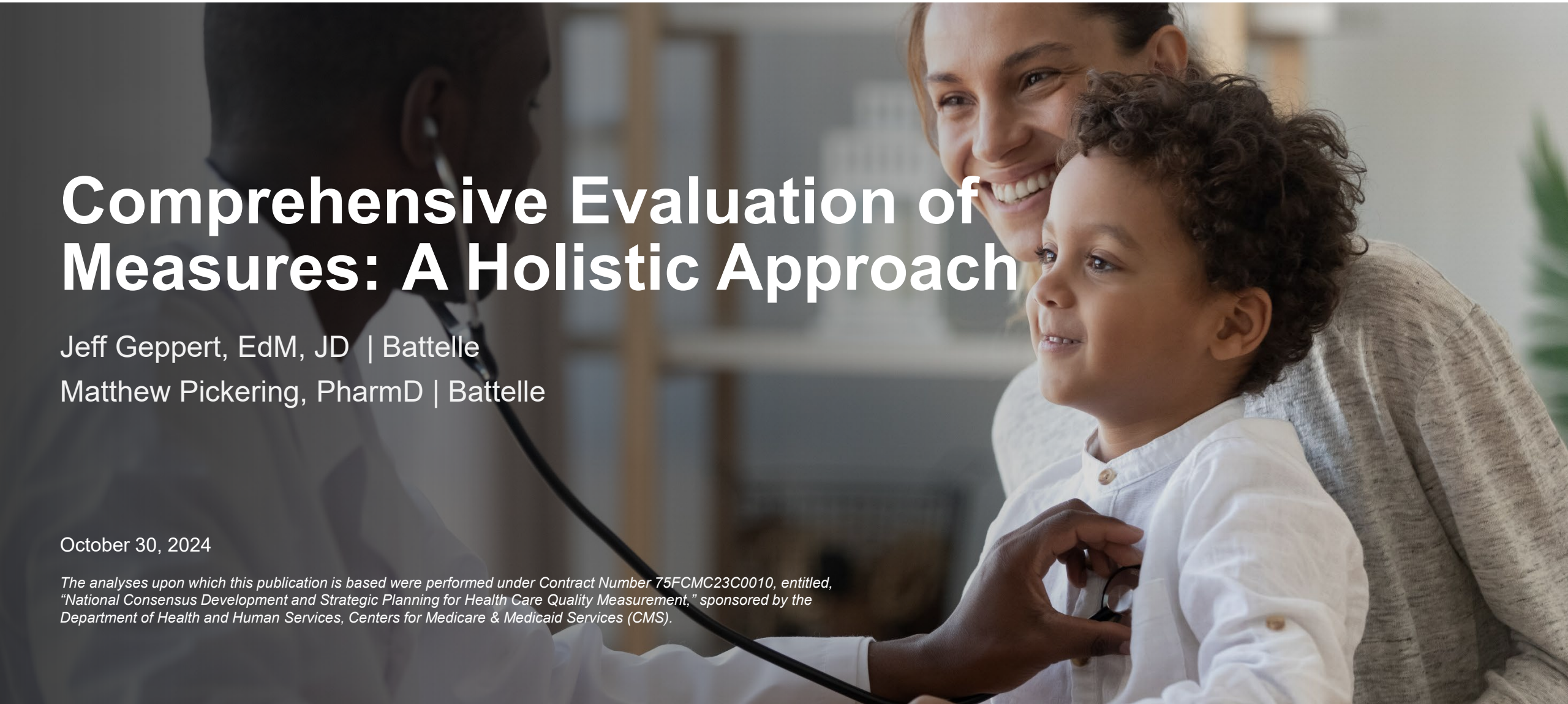


1. Click the lower part of your screen to mute/unmute or start or pause video.
2. Click on the participant button to access the full participant list.
3. To raise your hand, select the raised hand function under the reactions tab.
4. To ask a question, use the Q&A button at the top right of your screen (4A). If you do not see this, you can select the “more” icon (4B).

Partnership for Quality Measurement

Powered by Battelle

Partnership for Quality Measurement
Powered by Battelle



Comprehensive Evaluation of Measures: A Holistic Approach

Jeff Geppert, EdM, JD | Battelle
Matthew Pickering, PharmD | Battelle

October 30, 2024

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

Meet the Presenters

Jeff Geppert | Sr. Research Leader

- Leads Measurement Science team for endorsement & maintenance (E&M)
- 25+ years' measurement science, health care, and quality experience

Matthew Pickering | E&M Technical Lead

- Oversees E&M processes
- 10+ years' quality experience

Session Objectives

Purpose	Review Battelle's holistic and enhanced approach to measure evaluation under the Partnership for Quality Measurement (PQM).
Agenda	- PQM's approach to measure evaluation, including vision and guiding policies - Q&A

CBE Strategy: Vision

> *When [evidence-based] health and health care policies and programs designed to improve outcomes are not driven by community interests, concerns, **assets**, and needs, these efforts remain disconnected from the people they intend to serve. This disconnect ultimately limits the influence and effectiveness of interventions, policies, and programs.*
>
> *– National Academies of Medicine (NAM), February 14, 2022*

Vision: The vision for the CBE is to realize health care system change through the integration of quality measurement and quality improvement processes, and alignment with the principles of **evidence-based policies and programs** and **meaningful community engagement**. (Transform finance principles)

CBE Strategy: Vision (*cont.*)

CBE Strategy contains three elements:

- a diagnosis of a critical obstacle;
- a guiding policy for overcoming that obstacle; and
- a set of coherent actions for implementing the guiding policy.

Characteristics	Quality Measurement when:	Quality Improvement when:
Change	System change	Behavioral change
Mechanism	Selection and choice	Plan-do-study-act (PDSA)
Focus	Systematic and persistent factors	Context and expertise
Value	Organizational value	User value
Measurement	Clinical quality measures, alternative payment models, clinical decision support, best practice	Key performance indicators, next best action, positive diversions, trajectories, sound practice
Learning	Ability	Effort
High-Reliability Organization	Complicated (solvable, deterministic)	Complex (emergent)
Knowledge	Centralized, general	Decentralized, specific
Computable Health Care	Computable	Non-computable

Integration

CBE Strategy: Critical Obstacle

Metric Myopia — Trading Away Our Clinical Judgment

Reassessing Quality Assessment — The Flawed System for Fixing a Flawed System

Peers, Professionalism, and Improvement — Reframing the Quality Question

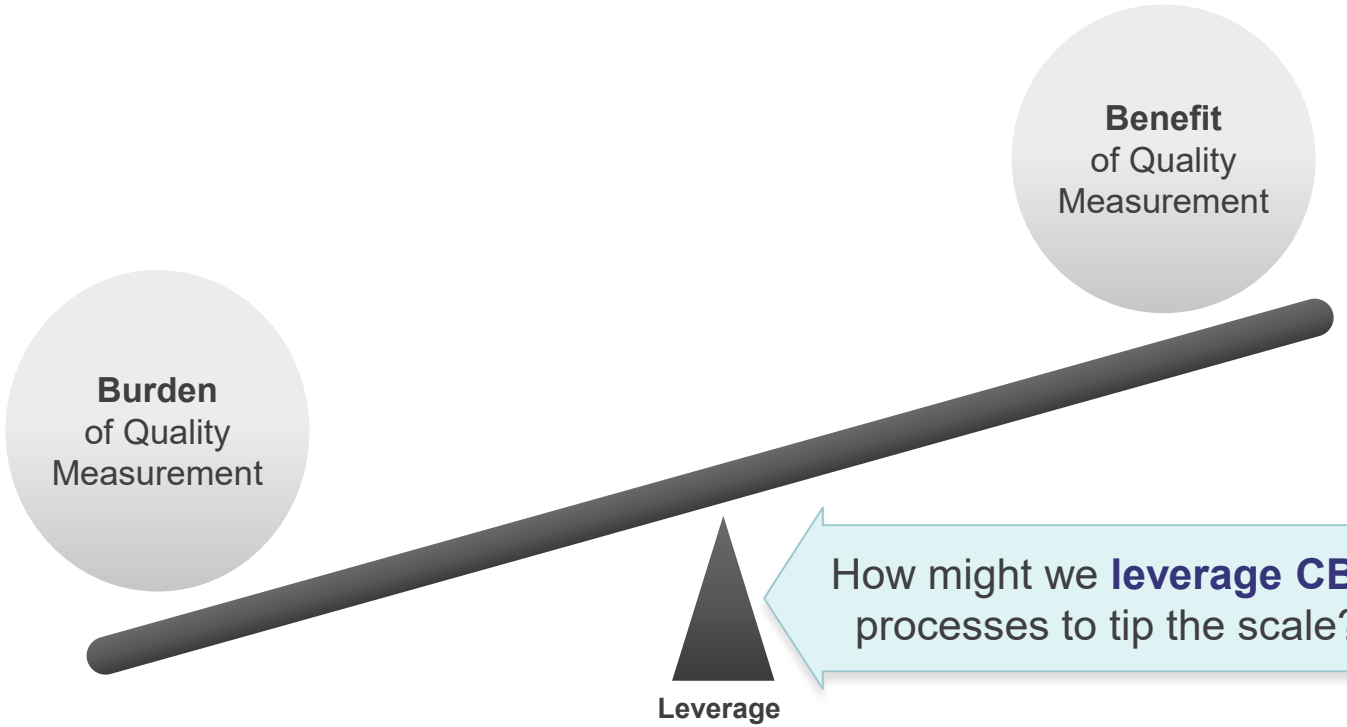
The NEW ENGLAND JOURNAL of MEDICINE

According to 2023 survey data from the Medical Group Management Association (MGMA):

- About three in four respondents said shifting to value-based payments has not improved the quality of care for patients.
- 94% of respondents said it has not reduced practices' regulatory burden, and positive payment adjustments aren't making up for the costs of time and resources spent preparing for and reporting.
- 68% of respondents said paying physicians based on value has not legitimately been successful.

Inside CMS (Vol. 26, No. 49 - December 7, 2023)

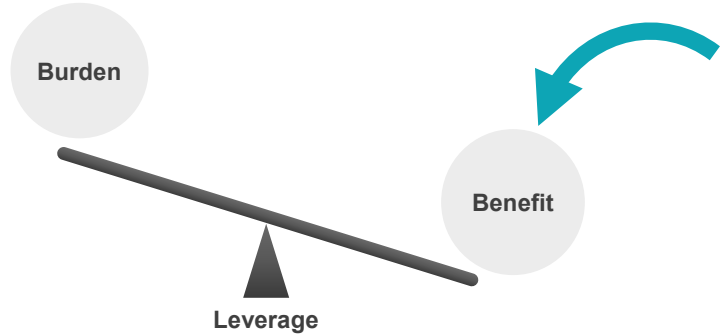
The Perceived Burden of Quality Measurement



CBE Strategy: Critical Obstacle (*cont.*)

Focus quality measurement where there is

the most benefit **for health care system change**



		RISK: Risk of measurement	
IMPACT	**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)
	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)

CBE Strategy: Guiding Policy

1. Generate organizational value for all interested parties

2. Generate trustworthy clinical quality measures (CQM)

a. Make the evidence explicit and explicitly evaluate that evidence.

b. Risks are identified, mitigated, managed, and evaluated to provide confidence in the positive impact of using the measure.

c. Information about steps taken to govern the CQM and address negative impacts and/or reduce bias or harm are documented.

3. Generate consensus

a. Consensus is a participatory process by which a group “thinks and feels” together in route to a decision rule.

b. CQM at various maturity levels may be endorsed given community consensus on trustworthiness.

CBE Strategy: Guiding Policy (*cont.*)

4. Transition from "what *should* you do" to "what *could* you do"

a. Engage the community in creative problem-solving.
b. Encourage full participation.
c. Promote mutual understanding.
d. Foster inclusive solutions.
e. Cultivate shared responsibility:
    1) Honestly acknowledge the challenges being faced.
    2) Provide a method to overcome those challenges.
    3) Focus and coordinate efforts to problem-solve.
    4) Fit-for-purpose: *What can we reasonably do?*
    5) Positive diversions: *What is working and for whom?*

CBE Strategy: Coherent Actions

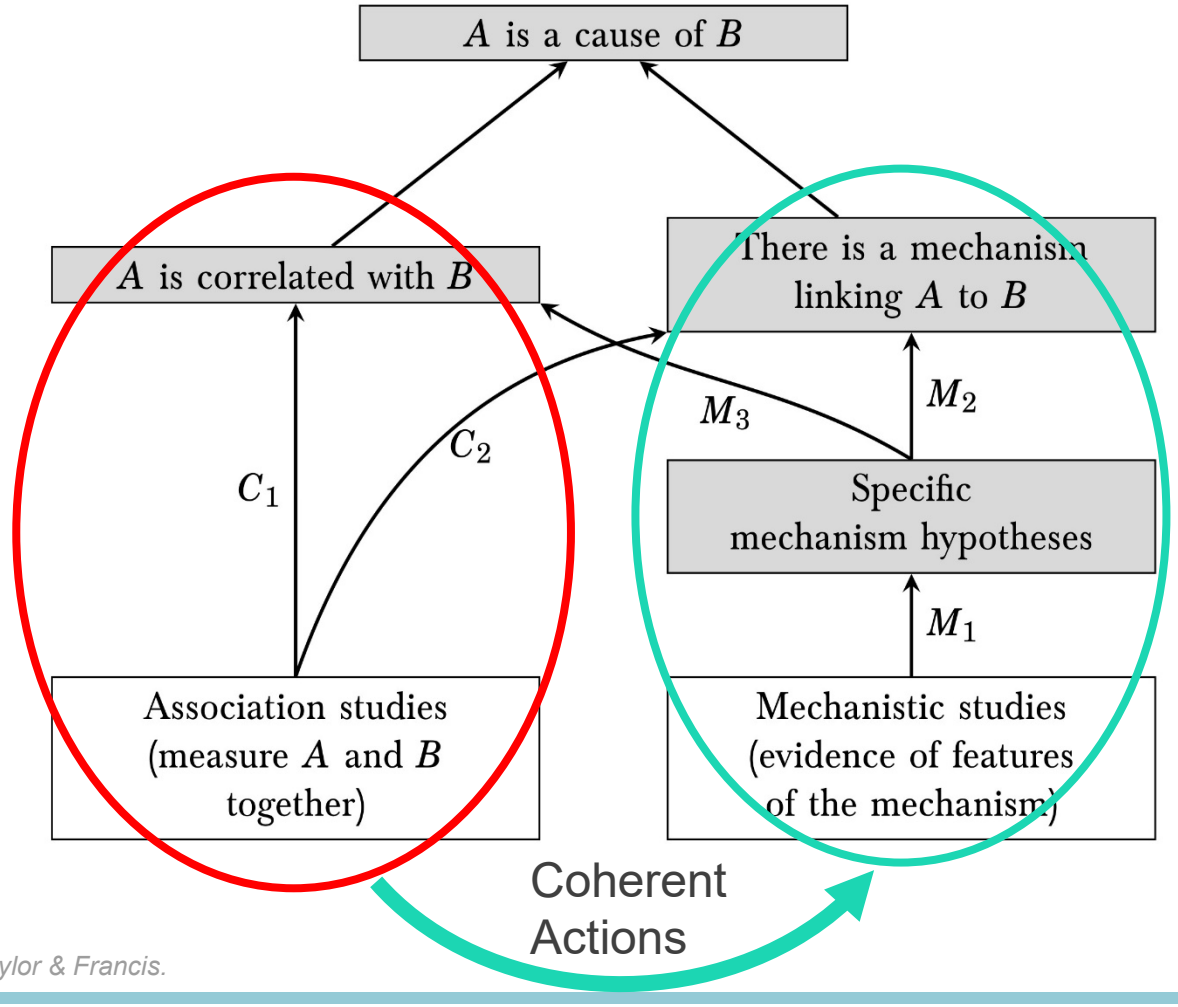
Measure developers and/or measure stewards make certain explicit or *implicit* assertions or claims about the potential benefits and risks/harms associated with measure use (net benefit).

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:

***Would* claim:**	Person or entity *would* make decisions based on the measure because the measure focus is associated with a material outcome (end/importance).
***Should* claim:**	There are known and effective ways of selection or improvement that the person or entity *should* use (ways/scientific acceptability). ▪ Known: mechanism; effective: causal
***Could* claim:**	Any barriers or facilitators to whether the person or entity *could* use those ways are known and addressed (means/usability).

CBE Strategy: Coherent Actions (Mechanisms)

- **A is a cause of B**
  - A: an intervention (drug, device, procedure, **quality program**, service delivery model, payment model)
  - B: an outcome (mortality, morbidity, harm, functional status, patient experience, workforce burden, **measure focus**)
- **Association claims**
  - A is correlated with B
- **General mechanism claims**
  - A is responsible for B
  - Accounts for the association



Source: Shan, Y., Williamson, J. (2023). Evidential Pluralism in the Social Sciences. United States: Taylor & Francis.

CBE Strategy: Coherent Actions (Mechanisms) (*cont.*)

Inputs (Resources/ Means)	Activities (What the program does- Ways)	Outputs (Direct results of the activities)	Outcomes	Impact (Broad, systemic changes influenced by the program):
• Skilled healthcare professionals (nephrologists, surgeons, nurses). • Training programs for AVF placement and maintenance. • Medical equipment and facilities for surgery and follow-up care. • Patient education materials. • Funding for healthcare initiatives. • Support from healthcare policy and administration. • Access to patient data and healthcare records for monitoring.	• Early screening and identification of patients for AVF. • Preoperative vascular mapping to assess suitability for AVF. • Surgical creation of AVF. • Postoperative monitoring and care for AVF maturation. • Ongoing training and education for healthcare providers. • Patient education and counseling about the benefits and care of AVF. • Policy advocacy for supporting AVF use.	• Number of patients screened for AVF suitability. • Number of AVFs surgically created. • Number of healthcare providers trained in AVF-related procedures. • Educational sessions conducted for patients. • Policy changes or implementations supporting AVF use.	Short-term (Changes resulting from the outputs): • Increased awareness among patients and healthcare providers about the benefits of AVF. • Improved patient selection for AVF placement. • Enhanced skills among healthcare providers for creating and maintaining AVFs. • Improved patient readiness and compliance for AVF surgery. • Policy and systemic changes facilitating increased AVF use. Intermediate term (effects observed as the program matures) • Increased rate of successful AVF placements. • Reduced complications and failures in AVF post-surgery.	• Healthcare Policy and Funding o Program Influence: Advocacy and demonstrated success of the program can lead to changes in healthcare policies, prioritizing funding for AVF procedures and postoperative care. o Systemic Change: Shift in national or regional healthcare funding and policies to support early and efficient access to AVF for eligible patients. • Standardization of Care Practices o Program Influence: Implementation of best practices for AVF creation and maintenance could set a benchmark for care quality. o Systemic Change: Adoption of these standards across healthcare systems, leading to a more uniform approach to hemodialysis vascular access. • Training and Workforce Development o Program Influence: The focus on training and continuous education can highlight the need for specialized skills in nephrology and vascular surgery. o Systemic Change: Changes in medical education and professional development requirements, ensuring a well-trained workforce proficient in AVF management. • Patient Education and Engagement o Program Influence: Comprehensive patient education initiatives can

CBE Strategy: Coherent Actions (Evidence)

Level	Evidence	Source
0	A finding that the outcome is material to front-line clinicians and/or patients (persons) A finding of a consensus quality construct	Literature review or qualitative research (focus group); a structured consensus process (Delphi)
1	A finding of an empirical association between an implicit quality construct and the outcome	Descriptive analytics demonstrating outcome variation (between entities or sub-populations)
2	A finding of an empirical association between an explicit quality construct and the outcome	Diagnostic analytics demonstrating a quality construct-outcome association (entity or sub-populations)
3	A finding of a systematic and/or persistent empirical association between the quality construct and the outcome	Predictive analytics validating a quality construct-outcome causal inference (difference-making)
4	A standardized assessment and management approach quality construct that increases the likelihood of the outcome	Prescriptive analytics, systematic reviews and meta-analysis, clinical decision support, implementation science
5	A mature quality construct through the organized use of experience feedback, and the capture and analysis of context	Effective analytics, human-centered design transforming prescriptive tasks (what should be done) into effective activities (what is actually and effectively done)

Coherent Actions

Pay for Transformation

- Def. **transform finance** is a set of principles and practices based on the belief that finance is a tool that can *create* positive (or negative) outcomes

Transform Finance Principles	Pay for Transformation
Those affected have the chance to design outcomes, govern processes, and share in ownership.	Quality programs should leverage meaningful ***community*** engagement • Accountability for quality and utilization at the community Strengthened partnerships and alliances / Improved health and health care programs and policies Expanded knowledge / Thriving communities
Investors add more value than what they extract as returns.	Quality programs should ***generate*** value by: • Identifying and mitigating barriers to effort for low performing entities • Creating and disseminating resources that emulate the ability of high performing entities
Risks and returns are fairly allocated among stakeholders.	Quality programs should **establish** value through a progressive understanding of how entities transform (generative causation) • Purpose: what works for whom under what circumstances • Transform: structure + agency • Static: evidence-based practice + implementation fidelity • Dynamic: sound practice + “positive diversion”

Source: http://www.transformfinance.org

Incorporating Equity and Justice

Inequality

Refers to a difference in effectiveness across populations.

Inequalities have many potential drivers and do not necessarily indicate a need for policy intervention (e.g., related to personal choice, cultural factors, disposable income).

Inequity

Refers to a difference that persists even after controlling for these drivers/factors.

Specifically, there may be lower access in underserved or under-resourced communities (e.g., rural communities, lower-income households).

Injustice

Refers to cases where policies or other persistent or systematic factors exacerbate, or perpetuate, existing inequities.

Measurement Science Focus Areas

Measure developers make certain explicit or implicit claims about the net benefit associated with intended use.

Target Population

Context

Mechanism Complex

Measure Focus

Harm

Net Benefit

Material Outcome

Usability ("Could")

Scientific Acceptability ("Should")

Importance ("Would")

Measurement Science Focus Areas (*cont.*)

Meaningfulness Claims	Measurement Science Focus
Concept of Interest	
Importance	Assessing “would” claims; association between the measure focus and meaningful outcome
Conformance	Expected value of information (EVI) (trade-offs)
Feasibility	Systems Engineering Initiative for Patient Safety (SEIPS 30)
Context of Use	
Importance	**Performance deciles; achievable benchmark; intervention effectiveness**
Reliability	**Reliability deciles; mitigating low reliability**
Validity	**Assessing “should” claims; logic models/mechanism maps**
Usability	Assessing “could” claims; Human-Centered Design; EPIS and other implementation science frameworks

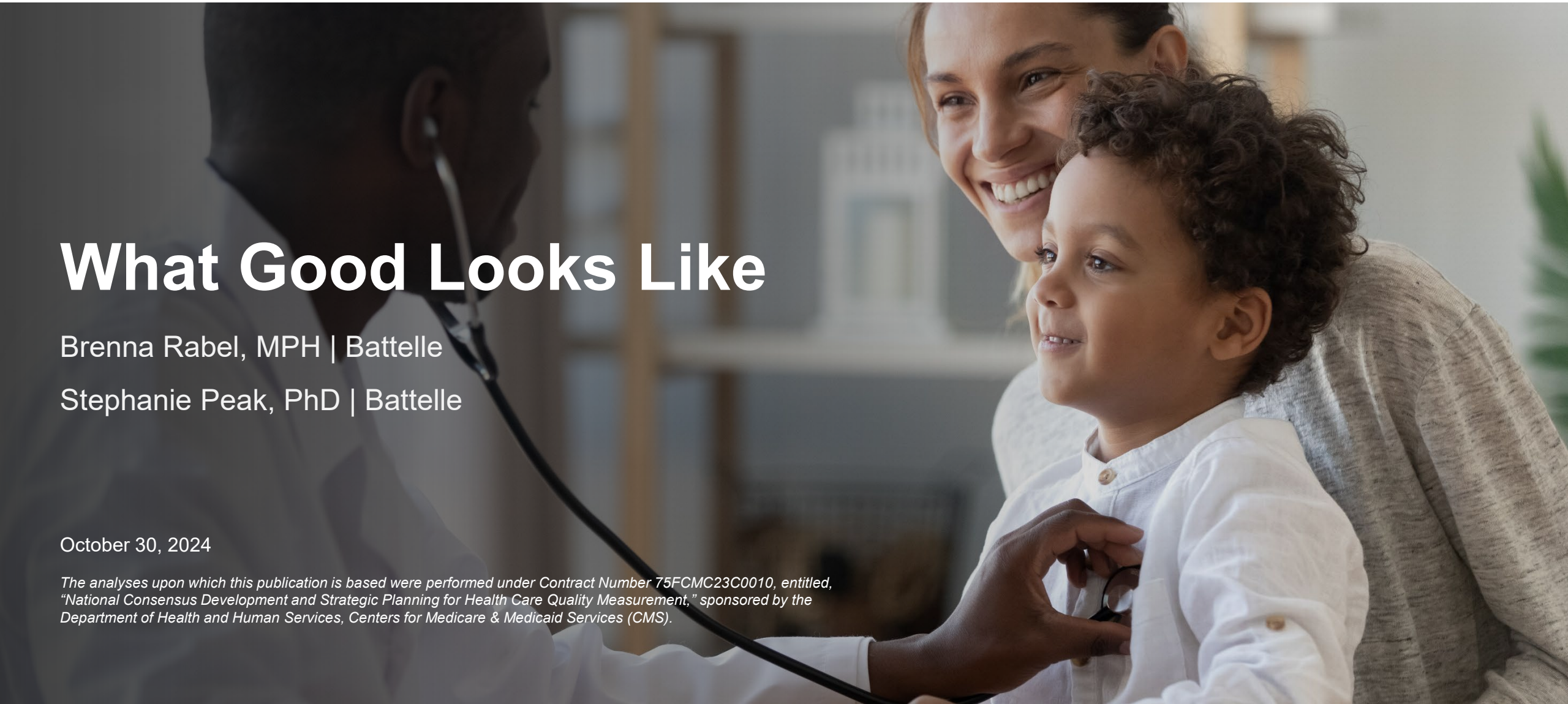
Bold: What we are working on now; **Bolded Green:** is priority

Questions & Answers

Partnership for Quality Measurement

Powered by Battelle

Partnership for Quality Measurement
Powered by Battelle



What Good Looks Like

Brenna Rabel, MPH | Battelle

Stephanie Peak, PhD | Battelle

October 30, 2024

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

Meet the Presenters

Brenna Rabel | Technical Director

- Facilitates collaboration across Consensus-Based Entity (CBE) activities to ensure consistency and excellence
- 10+ years' health care, public health, and quality experience

Stephanie Peak | E&M Support

- Provides technical support on Endorsement and Maintenance (E&M) process and activities
- 18+ years research experience, public health emphasis

Session Objectives

Purpose	To highlight best practices for a comprehensive application submission
Agenda	- Overview of “What Good Looks Like” resource - Best practices and common pitfalls and mitigations for measure submitters - Quality Measure Developer and Steward Agreement (QMDSA) Form, new/upcoming submission requirements, and opportunities for support

What Good Looks Like: Resource

- Covers Intent to Submit (ITS) and Full Measure Submission (FMS) questions
- Includes real-life examples for various measure types (e.g., process vs. outcome measures)
- Explains why each response is effective/appropriate
- Shares quick tips throughout the document that:
  - Emphasize areas of significance
  - Highlight insights to guide developers toward best practices
  - Provide actionable insights to help avoid common pitfalls
- Provides reminders throughout the document on:
  - Consistency across fields, updating information, or requesting assistance, etc.

A.5 Measure Testing (reliability and validity)

Check the boxes to attest to which testing (person/encounter-level or accountable entity-level) for reliability and validity will be available and submitted for each level of analysis by the FMS deadline of the intended review cycle. ***Note:*** *For initial endorsement, you must provide a rationale if empirical person or encounter-level will not be presented in the FMS. For maintenance endorsement, you must provide a rationale if measured/accountable entity testing will not be presented in the FMS.*

A.5a Empirical person- or encounter-level[1] *
Will empirical person- or encounter-level evidence, testing, methodology, and results be presented for this endorsement?

☒ No ☐ Yes

A.5b Empirical accountable entity-level *
Will empirical accountable entity-level evidence, testing, methodology, and results be presented for this endorsement?

☐ No ☒ Yes

☒ **A.6 Address health equity (optional)**
I will describe how this measure contributes to efforts to address inequities in health care. This is an optional criterion for FMS.

☒ **A.7 Measure's use or intended use ***
I will provide the measure's use or intended use and actions measured entities must take to improve performance on this measure. For a maintenance measure, I will provide a summary of any progress improvement.

Quick Tip

For initial endorsement, person- or encounter-level empirical testing is required, or existing evidence (e.g., prior research, literature) must be presented to support testing of all critical data elements (numerator, denominator, exclusions).

Because this is a maintenance measure, accountable entity-level empirical testing is required and the developer selects "yes" in question A.5b below.

Reminder

Equity will be a required domain beginning with the Spring 2025 cycle.

Best Practices: Style and Presentation

- Consider who will be reviewing these materials
  - Committee members with a range of backgrounds and experiences, including individuals with little to no measure science expertise
- Use plain language and ensure accessibility
  - Ask yourself: Will patient participants understand what you intend to convey?
  - Provide testing results in plain language to help readers understand the significance of the results. For example:
    - "A minimum signal-to-noise reliability score of 0.658 means that 65.8% of the variation in the measure scores among the measured entities is due to true differences in performance. The higher the percentage the more reliable the measure."
    - "The numerator Positive Predictive Value (PPV) was 98.77, indicating that 98.77% of the falls with injury reported by the measure were confirmed to be valid cases through clinical review of the patients' medical records."

Best Practices: Style and Presentation (*cont.*)

- Narrate the journey of measure development and testing:
  - Explain decision-making and rationale
    - What problem does the measure aim to address?
    - Why is a certain method/approach appropriate for the measure?
    - Include rationale for specification decisions (e.g., exclusions)
  - Provide insights from technical expert panels (TEPs) and feedback from patients
    - How did their input shape the measure?
    - How does the measure align with real-world patient needs?
  - Share any hypotheses that were not confirmed but could provide valuable context for reviewers
    - E.g., how many exclusions did you consider and test before landing on the ones in the final measure?

Pitfalls and Mitigations

Pitfall: Measure rationale

- **Mitigation:** The rationale should explain why a measured entity should report on the measure
  - NOT why the measure was developed
- **Mitigation:** Explain the benefits in quality and/or costs associated with the measure focus
  - Leverage the logic model to explain how reporting on the measure can lead to the desired outcomes, along with any other perceived impacts (such as reimbursement or financial incentives)
- **Mitigation:** Include references to guidelines/models aligned with the measure focus
  - Include information on how the measure aligns with existing guidelines or best practices

Pitfalls and Mitigations (*cont. 2*)

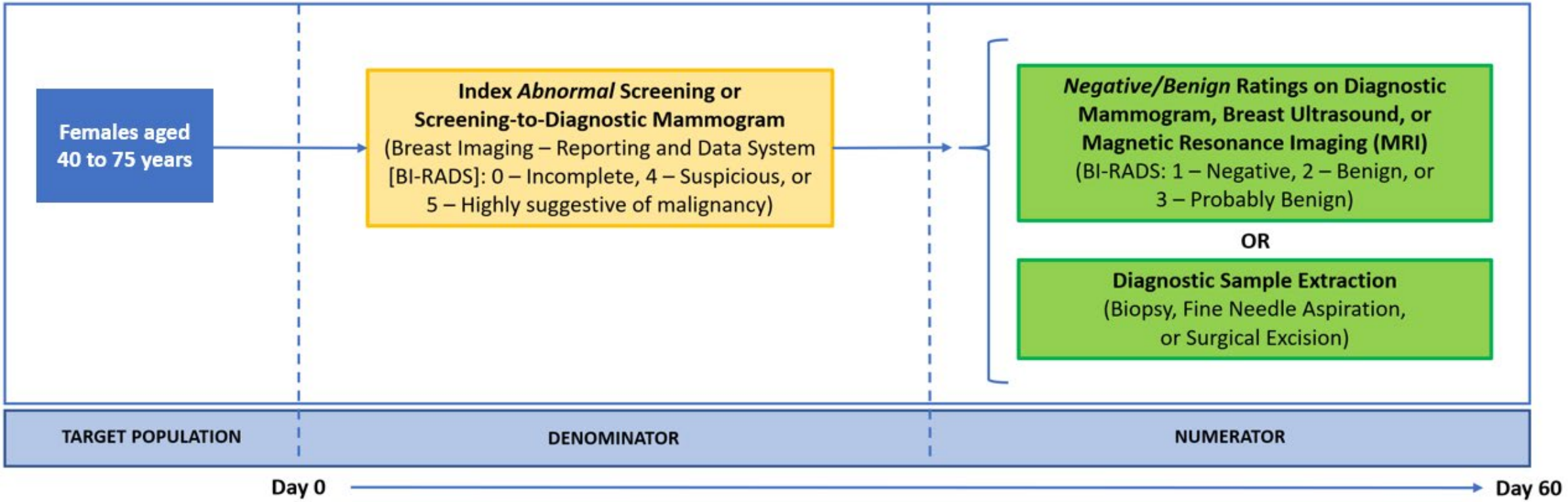
Pitfall: Incomplete measure details/failure to explain how to calculate the measure

- **Mitigation:** Clearly define numerator and denominator criteria
  - Reference specific code/value sets where applicable
  - Define measure elements and any relevant timing (e.g., timeframes for follow-up)
  - Define setting information (e.g., office visit vs. telehealth)
- **Mitigation:** Clearly describe exclusions and exceptions
  - Be sure they are described in plain language and reference where the user can find more detail in the value sets
- **Mitigation:** Include a diagram
  - Show how to calculate the measure in a step-wise fashion and supported with narrative

Specifications Diagram and Rationale – Example 1

Electronic Clinical Quality Measure (eCQM) Specifications Diagram:

Timely Follow-up on Abnormal Screening Mammograms for Breast Cancer Detection



Specifications Diagram and Rationale – Example 1 (*cont.*)

Electronic Clinical Quality Measure (eCQM) Rationale:

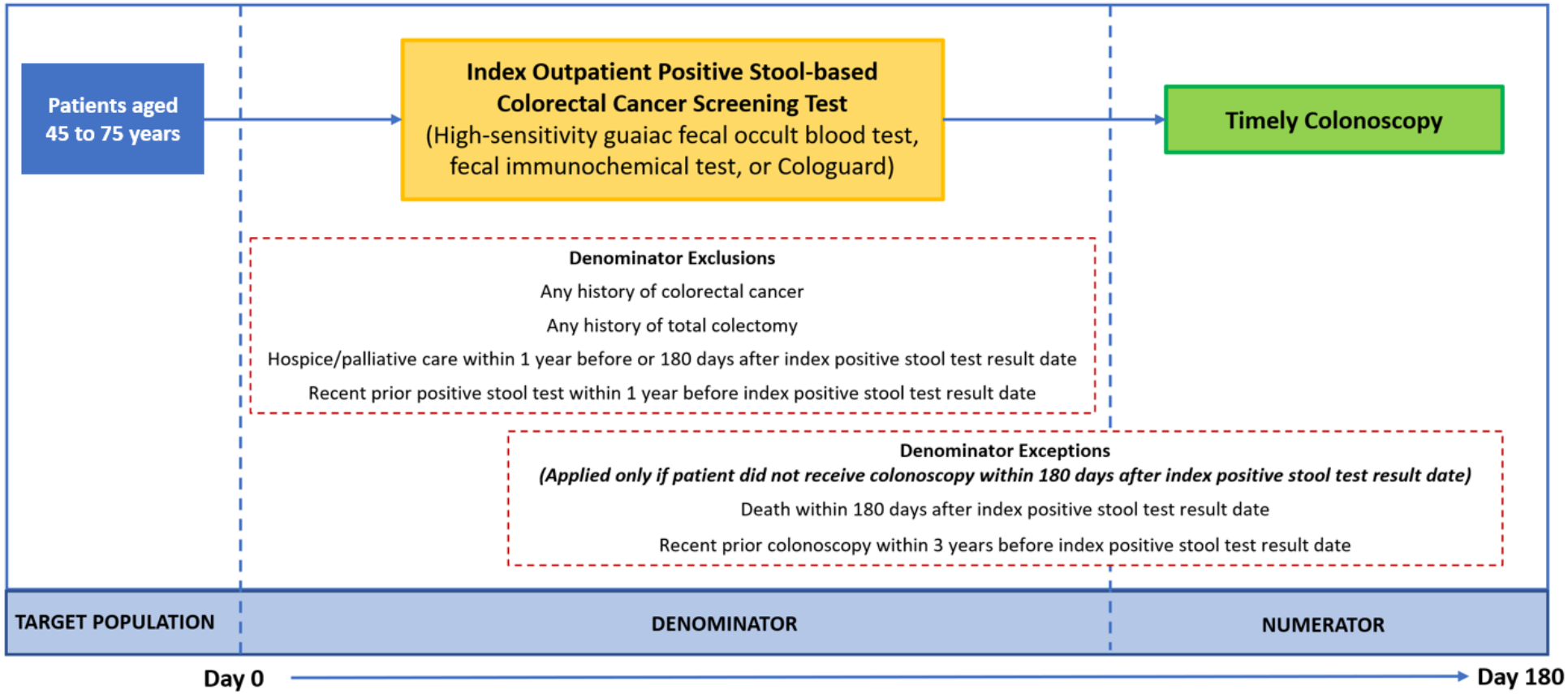
Timely Follow-up on Abnormal Screening Mammograms for Breast Cancer Detection

Criteria	Data Elements	Rationale
Denominator Inclusions	**Females aged 40-75 years**	Current age range for recommended universal breast cancer screening
	Screening mammograms	Recommended for universal breast cancer screening
	Abnormal result of BI-RADS 0, 4, 5	Requires immediate follow-up diagnostic evaluation
Numerator Inclusions	***Negative/Benign* Diagnostic Imaging with BI-RADS 1, 2, or 3:** Diagnostic mammogram, breast ultrasound, or breast MRI	Diagnostic resolution achieved
	Diagnostic Sample Extraction: Biopsy, fine needle aspiration, surgical excision	Diagnostic resolution achieved
	60-day follow-up period	Recommended by the National Breast and Cervical Cancer Early Detection Program

Specifications Diagram and Rationale – Example 2

Electronic Clinical Quality Measure (eCQM) Specifications Diagram:

Timely Follow-up on Positive Stool-based Tests for Colorectal Cancer Detection



Specifications Diagram and Rationale – Example 2 (*continued*)

Electronic Clinical Quality Measure (eCQM) Specifications Diagram:
Timely Follow-up on Positive Stool-based Tests for Colorectal Cancer Detection

Criteria	Data Elements	Rationale
Denominator Inclusions	**Patients aged 45-75 years**	Current age range for recommended universal colorectal cancer screening
	Stool-based tests: High-sensitivity Guaiac, Fecal Immunochemical Test, Cologuard	Stool-based tests recommended for colorectal cancer screening
	Positive result	Requires short-term follow-up diagnostic evaluation
Numerator Inclusions	**Diagnostic evaluation:** Colonoscopy	Gold standard for colorectal cancer diagnosis
	180-day follow-up period	Minimum delay with significant impact on patient outcomes from the published literature
Denominator Exclusions	**Index tests conducted in inpatient or emergency department settings**	Not indicated for colorectal cancer screening
	Recent positive stool-based test (<1 year before index test result)	May be undergoing repeat stool-based testing (not an index test)
	History of total colectomy	Not eligible for colonoscopy
	History of colorectal cancer	Under surveillance for colorectal cancer, not screening
	Receipt of hospice/palliative care (-1 year to +180 days after index test result)	Goals of care focused on comfort measures

Pitfalls and Mitigations (*cont. 3*)

Pitfall: Data sources

- **Mitigation:** Identify and describe all data sources used to calculate and test the measure
  - Clearly distinguish between sources used for testing, risk adjustment, and/or risk stratification
  - Descriptions should detail timeframes, collection databases (e.g., Integrated Data Repository), data source reliability and validity, and any feasibility challenges
  - As needed, select "Other Data Source" in Section 1.20 and describe use in Section 1.25 of the Full Measure Submission Form
- **Mitigation:** Align data sources in Section 1.25 with testing data sources in Section 1.20
  - If there are discrepancies, be sure to explain them in free text

1.20 Testing Data Sources *

Select the data sources for which you have tested and specified the measure. Choose all that apply.

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

1.20a Specify other data source *

1.25 Data Sources *

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

Pitfalls and Mitigations (*cont. 4*)

Pitfall: Lack of context around results

- **Mitigation:** Provide plain-language description of methods and findings
  - If technical language is unavoidable, add plain-language explanations of challenging concepts
- **Mitigation:** Include reasons for why the analyses used are the appropriate ones
  - How and why did you choose which analyses and tests to run?
- **Mitigation:** Explain what testing results mean
  - What do testing results tell us about the measure?
  - E.g., include more than “this demonstrates high reliability”— explain

Quality Measure Developer and Steward Agreement (QMDSA) Form

- Contract between Battelle and the measure steward
  - Outlines the steward's responsibilities for measure submission, review, and maintenance
  - If proprietary information prevents a steward from providing measure information, please contact PQMSupport@battelle.org before the Intent to Submit deadline to discuss QMDSA options
- Must be submitted **on or before** the project's Intent to Submit deadline
  - Effective for 5 years
  - Only one QMDSA is necessary per measure steward
    - **New Measure Stewards:** Include a completed addendum with the QMDSA listing all measures (measure number and title)
    - **Existing Measure Stewards:** Add new measures by completing an "Additional and Maintenance Measures Form"
  - Government organizations (e.g., CMS, CDC) do not require a QMDSA

Upcoming Requirements

Beginning Spring 2025:

- Measure submissions **must** use data from any year(s) within the last 5 years
  - Ensures relevance and reliability of performance results
  - Impacts both new and maintenance measures
  - This includes data used for:
    - Testing, performance gap and trend analyses, risk adjustment, and stratification
- The Equity domain will be **required** (attend 2:00 PM session on Equity guidance)

Opportunities for Support

- Battelle is available for technical assistance prior to measure submission. We can:
  - Provide support in framing feedback from TEPs and patients
  - Anticipate challenges associated with limited testing data
  - Suggest how to structure a logic model
  - Share considerations for a risk-adjustment conceptual model
  - Give clarifications on measure submission form items and/or elements of the PQM Measure Evaluation Rubric
- Measure developers/stewards should contact PQMsupport@battelle.org to request technical assistance and state what assistance is needed—sharing questions in advanced is preferred

Resources

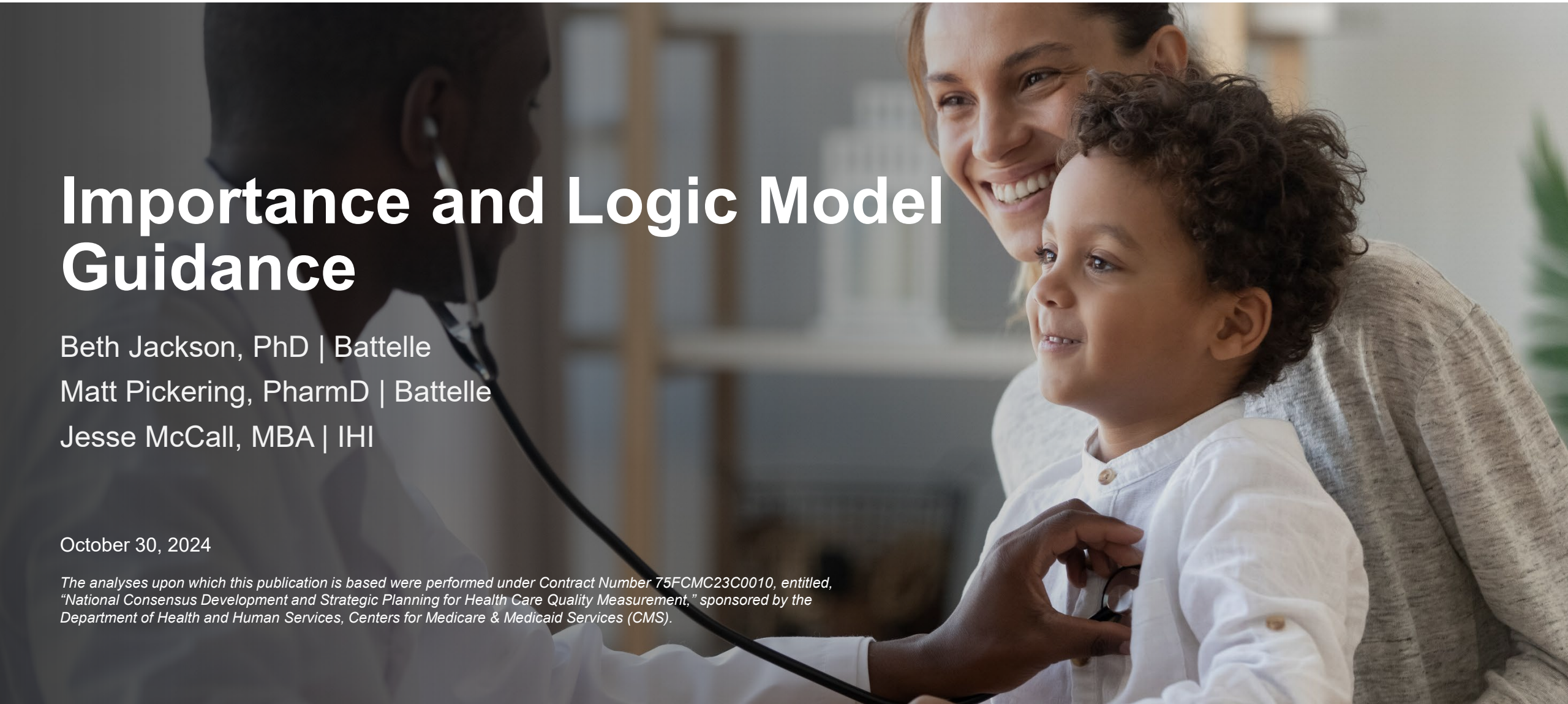
- **What Good Looks Like** – annotated examples of a process and an outcome measure submission highlighting best practices.
- **Intent to Submit and Full Measure Submission Forms** – downloadable submission templates.
- **PQM Measure Evaluation Rubric* (Fall 2024)** – provides measure evaluation criteria for Fall 2024 cycle. An updated PQM Evaluation Rubric effective for the Spring 2025 cycle can be found here.
- **E&M Guidebook** – provides information about the various steps of the E&M process, including each phase of review, possible endorsement decision outcomes, the appeals process, E&M policies and procedures, and the E&M committee structure.
- **QMDSA Form** – form and guidance for measure stewards submitting measures to Battelle for endorsement review.
- **QMDSA Submission Instructions** – provides information about the QMDSA process.
- **Additional and Maintenance Measures Form** – shortened form the measure stewards can use to add additional or maintenance measures to an existing QMDSA.
- **Measure Management System (MMS) Hub/Blueprint** – provides a start-to-finish overview of quality measure development, implementation, and maintenance steps and processes.

Questions & Answers

Partnership for Quality Measurement

Powered by Battelle

Partnership for Quality Measurement
Powered by Battelle



Importance and Logic Model Guidance

Beth Jackson, PhD | Battelle
Matt Pickering, PharmD | Battelle
Jesse McCall, MBA | IHI

October 30, 2024

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

Meet the Presenters

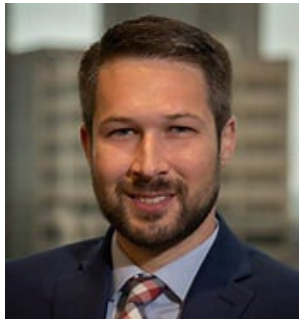
Beth Jackson | Evaluation Lead

- Coordinates endorsement and maintenance (E&M) staff assessments, provides evaluation technical support
- 10+ years' research and evaluation experience

Matthew Pickering | E&M Task Lead

- Oversees E&M processes
- 10+ years' quality experience

Jesse McCall | IHI Senior Director



- Supports logic model guidance development
- 17+ years as a process improvement professional

Session Objectives

Purpose	To review new guidance for developing a robust logic model and review common pitfalls within the importance domain.
Agenda	- Importance Criterion - Logic Model Draft Guidance - Q&A - Breakout Sessions

Importance Criterion

Matt Pickering, E&M Task Lead

Importance Criterion

Importance Definition: The 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.

Assertions:

- There is an association between the measure focus and a material outcome.
- A “material” outcome is one that persons or entities would rely upon in making decisions.
- The benefits of performance improvement to the achievable benchmark of care exceed the burden associated with data collection and reporting.
- A person or entity *would* make decisions based on the measure because the measure focus is associated with a material outcome.

Evaluating Importance

- The measure has a clear business case.
- There is a demonstrated relationship between health care structures and/or processes and the desired outcome(s).

Structure, process, and intermediate outcome measures	*Outcome and patient-reported outcome performance measures (PRO-PMs)*	*Cost/resource measures*
• An association between the measure focus and a material health outcome.	• These measures can effectively capture changes in health status that result from provided care.	• Evidence that changes in the measure can lead to more efficient care without compromising quality.

- A gap in care exists *(maintenance endorsement).*
- The relevant patient finds the measure focus meaningful.

Pitfalls and Mitigations

Pitfall: Meaningfulness to patients is not demonstrated

- **Mitigation: More comprehensive documentation**
  - Expand discussion of how meaningfulness was established, which may include a more thorough discussion of how patients/caregivers evaluated this assertion.
  - Fortify discussion of literature by stating explicitly how cited sources support the meaningfulness assertion.
- **Mitigation: Leverage technical expert panel (TEP) to more fully explore meaningfulness**
  - Consider expanding the TEP's evaluation of meaningfulness beyond yes/no voting on one or two standard items to include a discussion of meaningfulness and report responses to Likert scale items and/or textual analysis.
- **Mitigation: Improve patient participation**
  - Expand patient and caregiver representation on the TEP.

Pitfalls and Mitigations (*cont.*)

Pitfall: Limited data to show a performance gap

- **Mitigation: Fortify testing data**
  - Expand testing data by adding sites or adding years of data to the measurement period.
- **Mitigation: Explore sub-group differences**
  - Evaluate performance gap within sub-groups of race/ethnicity, sex, age, geography, or other risk factors.
- **Mitigation: Expand literature review**
  - Look for additional sources that demonstrate a gap or show disparities in performance.

Pitfalls and Mitigations (*cont. 2*)

Pitfall: Evaluation of evidence is incomplete or missing

- **Mitigation: Expand evidence discussion**
  - The evidence evaluation should include a summary of the quantity, quality, and consistency of the evidence.
  - When evidence is graded, the grade and the scale used should be provided verbatim.
  - Gaps in evidence should be clearly identified and the implications considered.
- **Mitigation: Justify references**
  - Ensure each reference listed is specifically addressed in the discussion and that the text clearly cites each relevant source in the appropriate location(s).

Pitfalls and Mitigations (*cont. 3*)

Pitfall: Logic model is superficial/sparse

- **Mitigation: Define all elements and relationships**
  - The logic model should specify and define the relationships between inputs, activities, and outputs of the quality program, and how these influence the measure.
- **Mitigation: Identify assumptions and external factors**
  - Ensure the logic model identifies underlying assumptions and factors outside the control of the quality program that may influence the measure.
- **Mitigation: Connect with evidence and face validity**
  - Logic model elements and the relationships between those elements should be supported in the review of the evidence.
  - Face validity testing should include evaluation of the logic model to establish buy-in by stakeholders.

Expectations for New vs. Maintenance

New Measures

- If implemented, what is the measure's anticipated impact on important outcomes?
  - What is the business case?

Maintenance Measures

- Provide evidence of performance gap or measurement gap by providing performance scores on the measure as specified (current and over time) at the specified level of analysis.
- Revisit and revise the logic model regularly, and for each maintenance submission.

Logic Model Draft Guidance

Beth Jackson, Evaluation Lead

What is a Logic Model?

"A logic model…is a graphic illustration of how a program or intervention is expected to produce desired outcomes. It shows the relationships among the inputs and resources available to create and deliver an intervention, the activities the intervention offers, and the expected results."[1]

"Logic models illustrate a sequence of cause-and-effect relationships—a systems approach to communicate the path toward a desired result."[2]

"Once a program has been described in terms of the logic model, critical measures of performance can be identified."[3]

Also known as a "theory of change" or "theory of action," a logic model is tested through collecting and evaluating evidence.

In clinical quality measurement, a logic model:

- Describes the context and environment in which a measure exists
- Specifies the causal relationship between inputs, activities, outputs, and outcomes
- Assists stakeholders in understanding the measure logic and how/whether the marshalled evidence supports the theorized causal pathway

1. AHRQ. 2013. *The Logic Model: The Foundation to Implement, Study, and Refine Patient-Centered Medical Home Models.* Publication No. 13-0029-EF.
2. Millar, A., R.S. Simeone, and J.T. Carnevale. 2001. Logic models: a systems tool for performance management. *Evaluation and Program Planning* 24:73-81.
3. McLaughlin, J.A. and G.B. Jordan. 1999. Logic models: a tool for telling your program's performance story. *Evaluation and Planning* 22:65-72.

Using the Guidance, Developers/Stewards Will Be Able to Use a Logic Model to:

- Concisely articulate the goals and anticipated investments, activities, outcomes, and feedback mechanisms associated with a clinical quality measure
- Identify and map available evidence and information needs to support model claims
- Visually communicate the measure logic in a manner accessible to a wide range of audiences
- Engage with stakeholders to establish face validity of the model and buy-in through critical evaluation of the quality improvement goals, strategy, and evidence for achieving those goals

PQM's Guiding Principles for Logic Models

Principle	Description
Clarity and Simplicity	Substantial effort should go into developing the logic model, but the model itself should distill how specific inputs and activities lead to desired outcomes and impacts.
Evidence Based, Measurable	Logic models should be grounded in evidence to support the expected strategies and outcomes and include measurable indicators for each element.
Stakeholder Involvement	Development should involve a diverse range of stakeholders to incorporate multiple perspectives and ensure models are informed by real-world clinical and patient contexts.
Alignment with Goals	Developers and stewards should ensure that all components of the logic model are aligned with the overarching goals and objectives of the quality improvement intent.
Patient-Centeredness, Equity	Logic models should reflect the priorities of patients through a focus on patient outcomes and experiences and ensure that quality improvements benefit all groups.
Flexibility	Logic models should be "living documents" to be adapted as new evidence develops and insights evolve.

Operationalizing Measure Logic Models

In the logic model framework for clinical quality measures

- The *measure focus* is equivalent to a logic model *outcome*.
- The quality construct should be expressed as the model's *inputs*, *activities*, and *outputs.*

Relationships between logic model elements are supported by appropriate evidence

- The *performance score* is the direct measure of the principal outcome of the quality construct.
- The performance score "finding" is sound only if the other logic model elements are grounded in evidence.

Feedback mechanisms enabling continuous improvement are a key component

Logic models should be specific to and plausible for the measure context

Logic Model Template

Inputs (Resources: Means)	**Activities** (What the program does: Ways)	**Outputs** (Direct results of activities)	**Outcomes** (Short-term, intermediate-term, long-term/goals)	**Impacts** (Systemic changes influenced by the quality program)
•	•	•	• ***Includes the measure focus***	•
Feedback Mechanisms (How continuous improvement is achieved)				
•				
Assumptions (Underlying beliefs about the quality program and context)				
•				
External Factors (Conditions outside the quality program's control)				
•				

Logic Model Element: Inputs

Inputs (Resources: Means) EXAMPLES
• Clinical guideline (knowledge) • Assessment instrument (tool) • Electronic health record (technology) • Programmer (staff)

Definition	Guidelines for Developers
The necessary resources and investments for a quality improvement (QI) program to succeed, including knowledge technology, tools, physical materials, and other resources.	Consider the value of existing resources as well as likely needed investments. What is the potential for variation in resources between entities? Are there obstacles such as cost or a consensus on clinical practice?

Logic Model Element: Activities

Activities
(What the program does: Ways)

EXAMPLES

- Training detailing standard of care for diagnosis (training)
- Development of clinical decision support system (CDSS) (tool)
- Creation of patient registry (data resource)

Definition	Guidelines for Developers
Actions of entities that are expected to affect outcomes and that have measurable outputs. Activities may include staff training, development of data resources and tools, or other focus areas.	Consider the evidence for and against specific QI efforts in the context of the clinical activities and practice settings relevant to the measure focus. Which interventions have been successful? Can these activities be feasibly integrated into existing workflows?

Logic Model Element: Outputs

Outputs
(Direct results of activities)

EXAMPLES

- Number of staff trained (reach)
- Improved knowledge of/support for the standard of care (quality)
- Provider utilization of CDSS during patient care (acceptability)

Definition	Guidelines for Developers
The measurable results of program activities, which quantify how well a quality program was implemented. Outputs can include indicators of the implementation's reach, quality, and acceptability/sustainability.	Consider the extent to which processes of the hypothesized QI program have been evaluated. How can the value of a QI program's implementation be determined? Is the QI program sustainable without ongoing investments?

Logic Model Element: Outcomes

Outcomes
(Short-term, intermediate-term, long-term/goals)

EXAMPLES

- Increased utilization of appropriate screening
- Reduced complications
- Enhanced patient self-efficacy
- Improved clinical documentation

Definition	Guidelines for Developers
The measure focus should be specified as one of the outcomes of the logic model. May include direct effects of outputs (short), effects as the program matures (intermediate), and/or effects on the condition the program aims to address (long). Some outcomes may be unintended.	Consider the different outcomes that may result from a QI program and where in this chain of events the measure focus best fits. Is the measure a short-, intermediate-, or long-term outcome of the program? What important outcomes might precede or come after the measure focus?

Logic Model Element: Impacts

Impacts (Systemic changes influenced by the quality program) EXAMPLES
• Reduced disease prevalence and disability • Reduced productive time lost • Enhanced health-related quality of life • Reduced disparities in preventable conditions

Definition	Guidelines for Developers
Broad system-, community-, or environmental-level effects that represent the ultimate change the QI program is intended to catalyze, such as reduced disease burden or a more efficient health care system.	Consider the longer-term, beneficial changes that are expected to follow from improvement in the measure focus and accompanying outcomes. What are the primary economic, social, and public health implications? How do the impacts improve equity and/or reflect patient priorities?

Logic Model Element: Feedback Mechanisms

Definition	Guidelines for Developers
A system or process designed to gather information about the QI program's progress and outcomes, allowing for adjustments and improvements to be made to the program based on that feedback. Some program activities may be part of this mechanism.	Consider how entities and providers learn about their progress and opportunities for improvement, especially after a QI program has been implemented. Is there an infrastructure to collect and communicate these learnings? What other kinds of feedback should be collected and from whom?

Feedback Mechanisms (How continuous improvement is achieved)

EXAMPLES

- Performance dashboard (direct feedback to providers about their improvement)
- Program monitoring data are used to revise training, update CDSS

Logic Model Element: Assumptions

Definition	Guidelines for Developers
These are the conditions that must hold for the logic model to be supported. Assumptions typically are not tested or measured. These can include gaps in the evidence base, expectations regarding availability of key resources, or beliefs about behavior.	Consider plausible factors that can potentially affect the operation of the model but that cannot be observed directly or easily. What is asserted to be true but cannot feasibly be demonstrated? What distinguishes your assumption from another element, such as an external factor?

Assumptions (Underlying beliefs about the quality program and context)

EXAMPLES

- Providers will comply with clinical guidelines.
- Supporting evidence from one practice setting applies to a different practice setting in which it has not been evaluated.

Logic Model Element: External Factors

Definition	Guidelines for Developers
Circumstances or events that are not within the control of the entity implementing a QI program that may affect outcomes. Policy, economic, social, and other factors may serve as barriers, facilitators, or moderators of the effect of a QI program on outcomes.	Consider what an environmental scan reveals about likely external factors that may complicate the operation or effectiveness of the QI program. What is the likelihood and anticipated size of each factor's effect? For an outcome measure, which factors are appropriate for risk adjustment or stratification?

External Factors (Conditions outside the quality program's control)

EXAMPLES

- Government regulations affecting medication prescribing (policy)
- Structural inequities affecting patients' ability to obtain services or adhere to a plan of care (socioeconomic)

Logic Model Example: CBE #4440e

Measure Title: Percent of hospitalized pneumonia patients with chest imaging confirmation

Measure Type: Process

Measure Description: The chest imaging-confirmed measure of pneumonia diagnosis is a process measure of inpatient hospitalizations that identifies the proportion of adult patients hospitalized with a discharge diagnosis of pneumonia and who received systemic or oral antimicrobials at any time during admission who received chest imaging that supported the diagnosis of pneumonia, as recommended by clinical practice guidelines. The measure applies to a target population of adult hospitalized patients.

Logic Model: CBE #4440e Percent of Hospitalized Pneumonia Patients with Chest Imaging Confirmation

Inputs	Activities	Outputs	Outcomes	Impacts

Tools & Tech Investments:
• EHR, Data warehouse, computational resources, data extraction tools
 	
Additional staffing Investments:
• Data analyst; coding specialist
• [Usual Staffing]
 	
[Guidelines/policies were not provided, if applicable]
	Data Processing:
	• Extract, analyze, and report data
	
	[Usual Patient Care Activities]
	
	• *Measure Performance Dashboard*
	• *Clinical Decision Support Tools*
		Diagnostic performance score
		
		Target population for other pneumonia quality measures
			Short-term (Changes resulting from the outputs):
			• ***Measure Focus:*** *Increased use of chest imaging to confirm pneumonia diagnosis*
			• Patient (& Provider): Increased awareness & attention to diagnosis
			• System: Ability to calculate and compare pneumonia incidence & outcomes, ability to track interventions to improve pneumonia care
			
			Intermediate term (Effects observed as the program matures)
			• Provider:
			  ○ Improved workup and documentation
			  ○ *Unintended consequence of measure: overuse of CT and ultrasound*
			• Patient:
			  ○ Reduce inappropriate antibiotic exposure
			  ○ Improve workup for alternative diagnoses
			  ○ Increase self-efficacy (to question a diagnosis)
			
			Long-term/Goals (Changes in the condition the program aimed to address-Ends)
			• System, professional societies, and quality organizations:
			  ○ Standardized population for other pneumonia quality measures
			  ○ Better evidence from population studies
			• More accurate diagnoses
			• Better health outcomes
			• [Reduced antibiotic resistance]
			• Improved transparency & trust
				• *Reduced antibiotic resistance in the community*

Logic Model: CBE #4440e Percent of Hospitalized Pneumonia Patients with Chest Imaging Confirmation, continued

Feedback Mechanisms
[Feedback mechanisms not described in LM.]
Assumptions
• Systems have access to high-quality electronic health record data and computational resources to support the capture and analysis of evidence of pneumonia on chest imaging. • Clinicians accept chest imaging as a standard confirmatory test for pneumonia. • Feedback to accountable entities will reduce misdiagnosis of pneumonia and inappropriate antibiotic use.
External Factors
• Changing technologies and standards for chest imaging and documentation (i.e., emergence of lung ultrasound) may make measurement more technically challenging but also more important • Policies and provider attitudes regarding measures will affect the short, medium and long-term outcomes.

Logic Model Best Practices

- Logic models should be concise and easily understood by external audiences.
- Leverage the logic model to help establish face validity of the measure, through engagement with your TEP and stakeholders to critically evaluate its claims.
- Review and revise the logic model as the evidence base grows, the practice environment shifts, and insights evolve.
- Carefully consider the role of each indicator in the logic model and how this is affected by context. An assumption in one model (e.g., clinicians will comply with clinical guidelines) may be the focus of an activity in another (e.g., training about guidelines).
- Balance comprehensiveness with parsimony—consider the relative importance/value of each indicator given the specific clinical activity and practice setting.

Questions & Answers

Breakout Sessions:
1. Developing Logic Models, or
2. Enhancing Logic Model Guidance

Jesse McCall, IHI Senior Director

Breakouts: Purpose and Logistics

- **Purpose:** To either: 1) develop an example logic model when provided with measure information or 2) discuss opportunities to enhance logic model guidance
- **Length:** 20 minutes
- Participants can choose one of five breakout rooms (see next slide for specifics).
- Each room will have a moderator.
- Feedback will be used to update and finalize logic model guidance by early 2025.
- Key takeaways will be shared at the 4:45 PM “Days End Review” session.

Breakouts: Options and Polling

Select your preferred breakout session topic:

Option 1: Developing Logic Models

- Participants will be provided with measure information and asked to build assigned areas of a logic model.
- Targeted audience: May be particularly beneficial for those with less experience in developing logic models.
  - 1.a – Inputs, Activities, Outputs
  - 1.b – Outcomes and Impact
  - 1.c – Feedback Mechanisms, Assumptions, and External Factors

Option 2: Enhancing Logic Model Guidance

- Participants will be asked to provide their feedback on ways to enhance the logic model guidance for the assigned areas.
- Targeted audience: Individuals with experience in developing logic models who can offer insights into gaps or ambiguities in the guidance.
  - 2.a – Inputs, Activities, Outputs
  - 2.b – Outcomes and Impact

Resources

- **Endorsement and Maintenance Guidebook** – This document contains comprehensive information pertaining to the Endorsement and Maintenance process, including conditions and non-negotiables for each domain, as well as a detailed measure evaluation rubric that outlines a successful submission for each domain.
- **PQM Measure Evaluation Rubric Worksheet** – Reviewers use this document is used to guide their assessments of measures under review for initial endorsement or maintenance. It includes PQM evaluation criteria and key considerations for reviewers' assessment of measures.
- **The Logic Model: The Foundation to Implement, Study, and Refine Patient-Centered Medical Home Models** – This resource from AHRQ provides a toolbox for developing logic models, using a primary care transformation demonstration as the use case.
- **PQM Logic Model Guidance** – Anticipated release in early 2025.

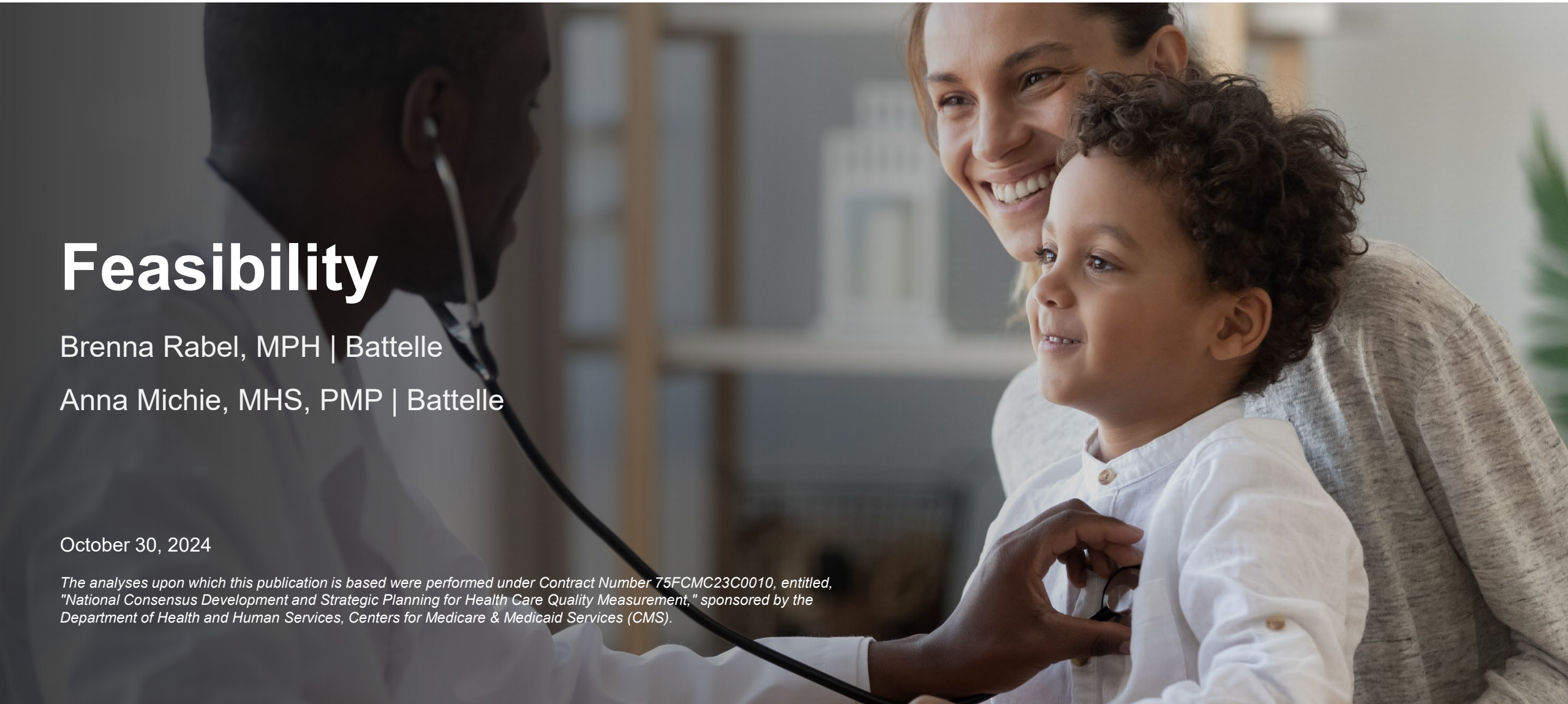
Partnership for Quality Measurement

Powered by Battelle

Partnership for Quality Measurement
Powered by Battelle

Lunch (12:15 – 12:45 PM ET)

Partnership for Quality Measurement
Powered by Battelle



Feasibility

Brenna Rabel, MPH | Battelle

Anna Michie, MHS, PMP | Battelle

October 30, 2024

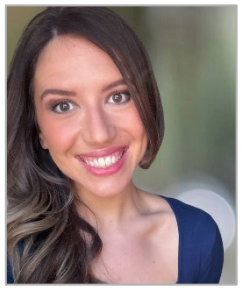
The analyses upon which this publication is based were performed under Contract Number 75FCMC23C0010, entitled, "National Consensus Development and Strategic Planning for Health Care Quality Measurement," sponsored by the Department of Health and Human Services, Centers for Medicare & Medicaid Services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

Meet the Presenters

Brenna Rabel | Technical Director

- Facilitates collaboration across Consensus-Based Entity (CBE) activities to ensure consistency and excellence
- 10+ years health care, public health, and quality experience

Anna Michie | E&M Deputy Task Lead



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

Session Objectives

Purpose	To review common pitfalls and mitigations for the feasibility domain and expectations for new versus maintenance measures.
Agenda	- Review measure feasibility expectations - Explain expectations for new vs. maintenance measures - Describe common pitfalls and discuss mitigation approaches - Q&A

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.
- There is an explicit articulation of the people, processes, and technology required for data collection and reporting.

Feasibility Assessment

- Describe whether all required data elements are: 1) routinely generated and used during care delivery, and 2) available in electronic health records or electronic sources
    - If not, provide a credible near-term path (within 1 year) to electronic collection
- Explain the extent of any missing data, measure susceptibility to inaccuracies, and the ability to audit data to detect problems
- Estimate the costs or burden of data collection, data entry, and analysis including the impact on clinician workflow, diagnostic thought processes, and patient-physician interaction
- Explain any barriers encountered or that could be encountered in implementing the measure specifications, data abstraction, measure calculation, or performance reporting
- Describe the 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
- Identify unintended consequences

Feasibility Scorecard

- Required for electronic clinical quality measures (eCQMs)
- Ensure you are using the approved template
- Testing of eCQMs should be conducted within electronic health record (EHR) systems from at least two EHR vendors
  - Collection and storage of required data elements may be different across EHR systems
  - Beyond this minimum requirement, developers/stewards should test on the number of health systems/facilities they deem appropriate

Feasibility Scorecard: What Good Looks Like

- Describe the EHR systems used for testing and which site used which system

3.1 Feasibility Assessment:

Thirteen hospitals participated in the evaluation of feasibility—four Epic and nine Allscripts users. All hospital sites confirmed that the data elements used in the measure are captured within the EHR in a structured and codified manner either using nationally accepted terminology standards or local system codes that could be easily mapped. However, one Epic hospital did not always use their structured fields to capture a fall that occurred during hospitalization. For this reason, the site opted to not proceed with reliability and validity phases of testing. Of note, three other Epic sites used in all testing phases did not encounter the same workflow challenges. Please see **Table 2 in logic model attachment** for combined feasibility scores for data availability, data accuracy, data standards, and workflow across all 13 hospitals.

	MEASURE INFORMATION		
Measure Title	Hospital Harm – Falls with Injury		
Care Setting	Inpatient/Hospital		
Level of Analysis	Facility		
EHR System #1	Epic	**EHR System #8**	Allscripts
EHR System #2	Epic	**EHR System #9**	Allscripts
EHR System #3	Epic	**EHR System #10**	Allscripts
EHR System #4	Allscripts	**EHR System #11**	Allscripts
EHR System #5	Allscripts	**EHR System #12**	Allscripts
EHR System #6	Allscripts	**EHR System #13**	Epic
EHR System #7	Allscripts		

Feasibility Scorecard: What Good Looks Like (*cont.*)

- List all key data elements used to calculate the measure
- Clearly identify data availability, data accuracy, data standards, and workflow issues, if any

EHR System	Epic				
		DATA AVAILABILITY	DATA ACCURACY	DATA STANDARDS	WORKFLOW
#	Data Element	Is the data readily available in a structured format, i.e., resides in fixed fields in EHR?	What is the accuracy of the data element in EHRs under normal operating conditions? Are the data source and recorder specified?	Is the data element coded using a nationally accepted terminology standard?	Is the data captured during the course of care? And how does it impact workflow for the user?
		Score	Score	Score	Score
1	Encounter, Performed: Emergency Department Visit	1	1	1	1
2	Encounter, Performed: Encounter Inpatient	1	1	1	1
3	Encounter, Performed: Observation Services	1	1	1	1
4	Adverse Event: Inpatient Falls	1	1	1	0
5	Diagnosis: Inpatient Falls	1	1	1	0
6	Medication, Active: Anticoagulants for All Indications	1	1	1	1

Feasibility Scorecard: What Good Looks Like (*cont. 2*)

- Describe a plan to overcome data challenges for any data elements that score 0

DATA ELEMENT FEASIBILITY PLAN			
For data elements that score 0, provide plan for projected use of element.			
Data Element	**How is the data element used in computation of measure - e.g. numerator, denominator?**	**Explain how the data element is feasible within the context of the measure logic?**	**What is the plan for readdressing this data element?**
Adverse Event: Inpatient Falls	Used in numerator	This data element is for structured clinical documentation that a fall occurred during the hospitalization. Hospitals tend to capture detailed information specific to inpatient falls within a risk management module, however hospitals generally capture structured documentation that a fall occurred or that the provider was notified of a fall occurrence. For feasibility, we assessed 13 hospitals to determine the ability to structurally capture such documentation, and all sites have the defined fields within the EHR.	While all 13 hospitals evaluated have the ability to document inpatient fall occurrence within their EHR systems, one hospital had inconsistent clinical documentation workflows and did not always utilize their structured fields. For this reason, the site opted to not proceed with reliability and validity phases of testing. Of note, this was an Epic site, and 3 other Epic sites used in all testing phases did not encounter the same workflow challenges. The test site is aware of their documentation challenges and will work to make improvements to better enable capture going forward.
Diagnosis: Inpatient Falls	Used in numerator and denominator exclusion	This data element is for coded documentation that a fall occurred during the hospitalization (numerator) or that a fall occurred prior to admission (denominator exclusion). Given that falls that result in death and serious injury are included in the list of serious reportable events, hospitals are often diligent about what events occur within the context of the hospitalization and which are atrributed to circumstances outside of the hospitalization. For feasibility, we assessed 13 hospitals to determine ability to capture diagnosis information and their associated present on admission indicators. All hospitals capture diagnosis information and POA indicators.	While all 13 hospitals capture diagnosis information and POA indicators, one site was not capturing fall documentation (whether in hospital or prior to admission) within current documentation workflows. Of note, this was an Epic site, and 3 other Epic sites used in all testing phases did not encounter the same workflow challenges. The test site is aware of their documentation challenges and will work with their clinical documentation specialists to make improvements to enable capture going forward.

Feasibility-Informed Final Measure

- Describe any changes made to the measure specification because of feasibility testing (e.g., adding measure logic to allow for electronic capture of additional or different data elements)

While mechanical ventilation was captured in structured fields at all sites, documentation was not standardized. For example, some information was found in respiratory free text notes and start/end times were not discrete. Recognizing mechanical ventilation may have variability, we also evaluated intubation and extubation documentation for consideration in the measure specification. Though these two elements were more frequently captured, there are opportunities to expand electronic capture. For example, two of 13 sites documented rapid response interventions (including intubation) on paper and scanned into the EHR and others documented intubation/extubation in anesthesia free text notes for certain procedural areas (e.g., gastrointestinal lab, cardiac cath lab).

3.3 Feasibility Informed Final Measure:

Due to variable documentation for mechanical ventilation (as described above), the measure also accommodates the use of intubation and extubation outside of a procedural area to trigger a postoperative respiratory event.

Proprietary Information and Fees, Licensing, or Other Requirements

- Indicate whether your measure or any of its components are proprietary, with or without fees
- 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)
  - Are there costs related to data collection and validation (e.g., vendors)?
  - Can the measure be reproduced and distributed?
  - Are there differences for commercial vs. non-commercial purposes?
  - Are there licensing requirements (especially for surveys and patient-reported outcome tools)?

Expectations for New vs. Maintenance

New Measures

- Describe the extent to which the required data elements:
  1. Are routinely generated and used during care delivery, AND
  2. Are available in electronic health records or other electronic sources or provide a credible near-term path (within 1 year) to electronic collection, AND
  3. Have a data collection strategy that can be implemented.

Maintenance Measures

- If measure specifications **have** changed:
  - Discuss the extent to which those changes impact 1-3 above AND
  - Describe any measure implementation challenges that occurred because of the data elements and provide the mitigation strategy used or a near-term path (within 1 year) to overcome the challenges.
- If measure specifications have **not** changed:
  - Describe any measure implementation challenges that occurred because of the data elements and provide the mitigation strategy used or a near-term path (within 1 year) to overcome the challenges.

Pitfalls and Mitigations

Pitfall: Data elements are available in EHR structured fields but are not being used

- **Mitigation:** Conduct feasibility testing in other sites to determine if the issue is widespread or site specific; describe differences in the feasibility assessment
- **Mitigation:** Discuss with technical expert panel and other SMEs to gain insight on burden and an estimation of the resources/staff time associated with changing workflows
- **Mitigation**: Explore use of natural language processing (NLP)

Pitfalls and Mitigations (*cont.*)

Pitfall: Documentation workflows impacting data completeness and accuracy

- **Mitigation**: Recommend staff training to improve accuracy and completeness of data fields
- **Mitigation:** Provide implementation guidance as a path to push the field forward
- **Mitigation:** Consider a clinical decision support system

Pitfalls and Mitigations (*cont. 2*)

Pitfall: Low response rates (instrument-based measures)

- **Mitigation:** Consider methods and timing of survey administration
- **Mitigation:** Evaluate for burden (e.g., reduce the number of fields)
- **Mitigation:** Align with other questionnaires or surveys to minimize duplicative effort and survey fatigue

Resources

- **Endorsement and Maintenance Guidebook** – This document contains comprehensive information pertaining to the Endorsement and Maintenance process, including conditions and non-negotiables for each domain, as well as a detailed measure evaluation rubric that outlines a successful submission for each domain.

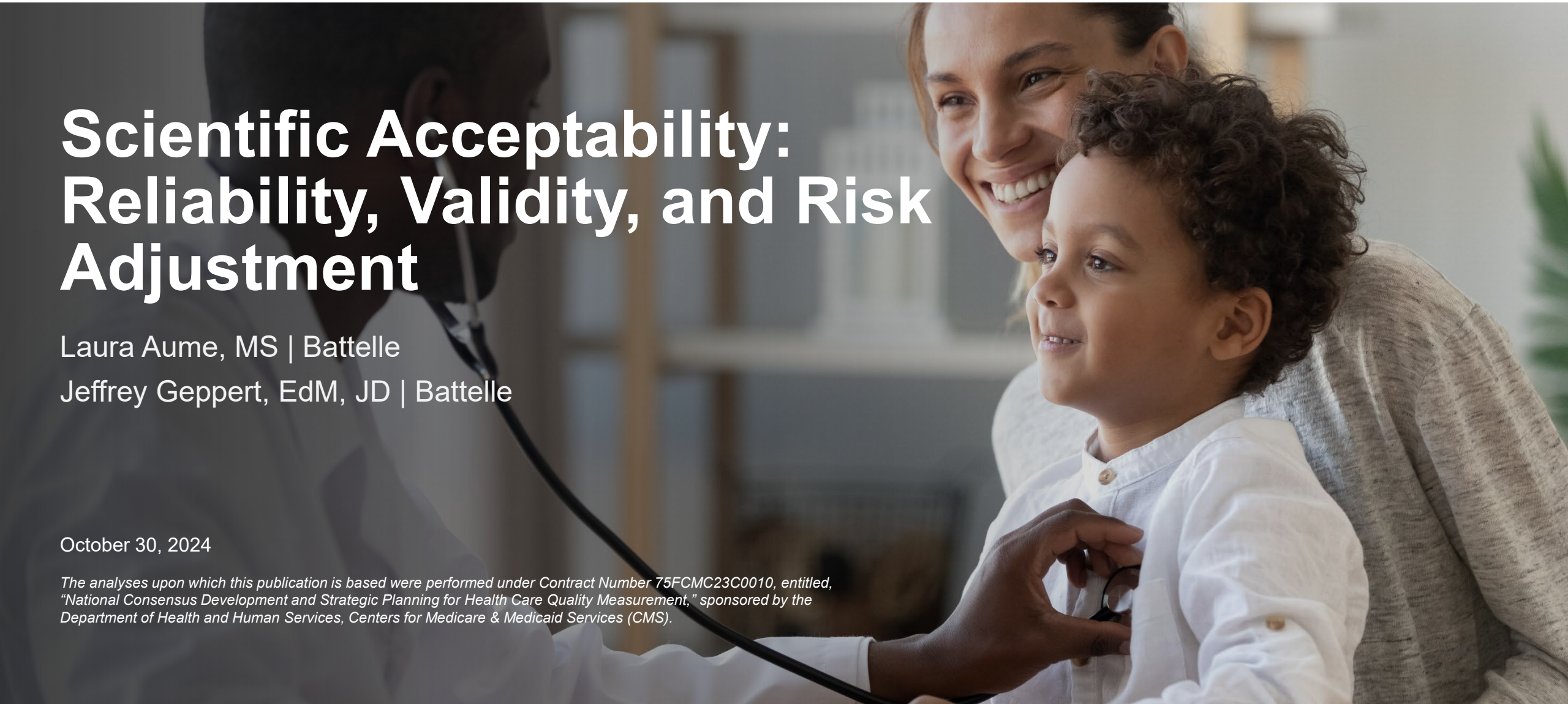
- **PQM Measure Evaluation Rubric Worksheet** – This document is used by reviewers to guide their assessments of measures under review for initial endorsement or maintenance. It includes PQM evaluation criteria and key considerations for reviewers' assessment of measures.

Questions & Answers

Partnership for Quality Measurement

Powered by Battelle

Partnership for Quality Measurement
Powered by Battelle



Scientific Acceptability: Reliability, Validity, and Risk Adjustment

Laura Aume, MS | Battelle
Jeffrey Geppert, EdM, JD | Battelle

October 30, 2024

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

Meet the Presenters

Laura Aume | Data Scientist

- Leads Endorsement and Maintenance (E&M) reliability methods evaluation and development
- 29+ years of quality and data science experience

Jeffrey Geppert | Sr. Research Leader

- Leads Measurement Science team for E&M
- 27+ years' measurement science, health care, and quality experience

Session Objectives

Purpose	To discuss expectations and common approaches to substantiating reliability and validity claims at both the person or encounter level and accountable entity level, explore common pitfalls and mitigation strategies, and relate these elements to the logic model to enhance understanding and application of testing approaches.
Agenda	- Reliability - Validity - Risk adjustment - Q&A

Reliability

Laura Aume

What is Reliability?

"Validity is measuring the right thing; reliability is measuring the thing right."
– Thissen (2001)

Reliability is the degree to which a measure repeatedly and consistently produces the same result — ISO/IEC 25020 — Quality measurement framework (2019)

Person- or Encounter-Level Reliability Testing

Table 1. For Person- or Encounter-Level Reliability

Approach	Threshold
Internal consistency	> 0.7
Inter-rater agreement	> 0.6
Test-retest reliability (ICC or Pearson correlation)	> 0.5
Linear relationship	> 0.6

(Source: Acceptable Reliability Thresholds, December 2021)

- Reliability testing is only required for data elements collected by humans or natural language processing (NLP)
  - Reliability testing is not required for electronically captured data elements (e.g., electronic clinical quality measure [eCQM] or digital quality measure [dQM])
- For NLP, report reliability statistics for at least two sites using a "template" of common person features
  - Provide an explanation for any discrepancies

Accountable Entity Reliability Testing

Table 2. For Accountable Entity-Level Reliability*

Approach	Threshold
Signal to noise/inter-unit reliability	> 0.6
Split-half reliability (ICC)	> 0.6

*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.

(Source: Acceptable Reliability Thresholds, December 2021)

- Common methods for estimating accountable entity reliability
  - Binary: Adams (beta-binomial)
  - Complex (risk-adjusted ratio or rate): ICC (with and without bootstrap) with Spearman-Brown adjustment
  - Hierarchical (clustering and shrinkage): mixed logistic regression, empirical Bayes
- Other
  - Spearman rank-order correlation (with and without bootstrap)

Example 1: Reliability by Deciles

Table 3. Signal-to-Noise Estimates of CBE #4125 - Thirty-day Risk-Standardized Death Rate among Surgical Inpatients with Complications (Failure-to-Rescue)

	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Reliability	0.7039 (mean)	**0.2314**	**0.2571**	**0.3248**	**0.3879**	**0.4671**	**0.5379**	0.6016	0.6697	0.7384	0.8106	0.8861	0.973
N of Entities	2055	21	205	206	206	206	205	205	205	206	205	206	1
N of Persons/ Encounters/ Episodes	1087624	525	15853	21776	29419	40024	53027	69384	92901	129893	199744	435603	8099

- Reliability is a feature of the entity(s), not the measure.
- Majority (51%) of entities have a reliability <0.6.
- Developer/steward should consider mitigation for entities with low reliability estimates.

Mitigation of Low Reliability

- Start by understanding the features of the persons and entities at low-reliability and high-reliability entities

Features	Low-reliability entities	High-reliability entities
Person-level features		
Entity-level features		
Average person-level features by entity		
Geographic features		

- Low reliability might be acceptable in certain contexts:
  - Measure focus with a direct causal association (e.g., complications, errors)
  - No alternative for choice-making. A low reliability measure is usually better than no measure (e.g., Farmer's Almanac)
  - A related high-reliability measure to inform the choice (structure, process, or outcome)
  - Trade-offs in harm favor low reliability over low validity or lower quality program participation

Example 2: Reliability Methods

- There are multiple methods for estimating reliability. However, the relationship between the methods and the context in which each method works best is unclear.
- Battelle is conducting a simulation study to address these two issues, and to provide guidance to developers on selecting the most appropriate method.

Pitfalls and Mitigations

Pitfall: Not demonstrating person- or encounter-level reliability

- **Mitigation: Make an argument about the potential harms to persons and entities associated with low reliability**
  - Present plans current and future to address, including implementation and feedback
- **Mitigation: Make an argument about the trade-offs between reliability and feasibility**
  - Provide an explicit articulation of the people, processes, and technology required for data collection and reporting
- **Mitigation: Present any plans for future digital capture and reporting**
  - In particular, the use of interoperability standards (e.g., USCDI+ and FHIR)

Pitfalls and Mitigations (*cont.*)

Pitfall: Not demonstrating accountable entity-level reliability

- **Mitigation: Use the appropriate method for the performance score**
  - Some developers calculate reliability on the observed rate rather than the risk-adjusted ratio or rate (which tends to have lower reliability)
- **Mitigation: Consider mitigation approaches for low-reliability persons and entities**
  - "Borrow strength" from related structure, process, or outcome measures
  - At a minimum, report characteristics of low-reliability persons and entities
- **Mitigation: Make an argument about the trade-offs between reliability and validity or stability**
  - Low reliability does not mean the measure is not useful. It depends on the alternatives and the consequences

Expectations for New vs. Maintenance

New Measures

- Person- or encounter-level reliability testing
- Data within the last 5 years
- Explicit articulation of harms associated with low reliability and plans for mitigation of such harms

Maintenance Measures

- Accountable entity-level reliability testing
- Data within the last 5 years
- Reliability table by decile
- Plan for mitigation for low-reliability (< 0.6) entities and persons

Resources

- **Ash, A. S., Fienberg, S. F., Louis, T. A., Normand, S. L. T., Stukel, T. A., & Utts, J. (2012). Statistical issues in assessing hospital performance** – provides the basic rationale for the use of shrinkage to increase reliability in accountability uses (however with the trade-off of decreased validity)
- **Adams, J. L. (2009). The reliability of provider profiling: a tutorial** – provides the basic rationale for the reliability metric as related to the risk of misclassification and provides a method of calculation for binary variables (i.e., beta-binomial)
- **Nieser, K. J., & Harris, A. H. (2024). Comparing methods for assessing the reliability of health care quality measures. Statistics in Medicine** – recent paper presents simulation results comparing alternative methods for calculating the reliability metric
- **Nieser, K. J., & Harris, A. H. (2024). Split-sample reliability estimation in health care quality measurement: Once is not enough. Health Services Research** – recent paper demonstrating that multiple iterations of the ICC improves robustness

Validity

Jeffrey Geppert

What is Validity?

"Validity is measuring the right thing; reliability is measuring the thing right."
– Thissen (2001)

Validity is "an overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy of appropriate interpretations and actions on the basis of [the measure]."
– Messick (1989); Standards for Educational and Psychological Testing (2014)

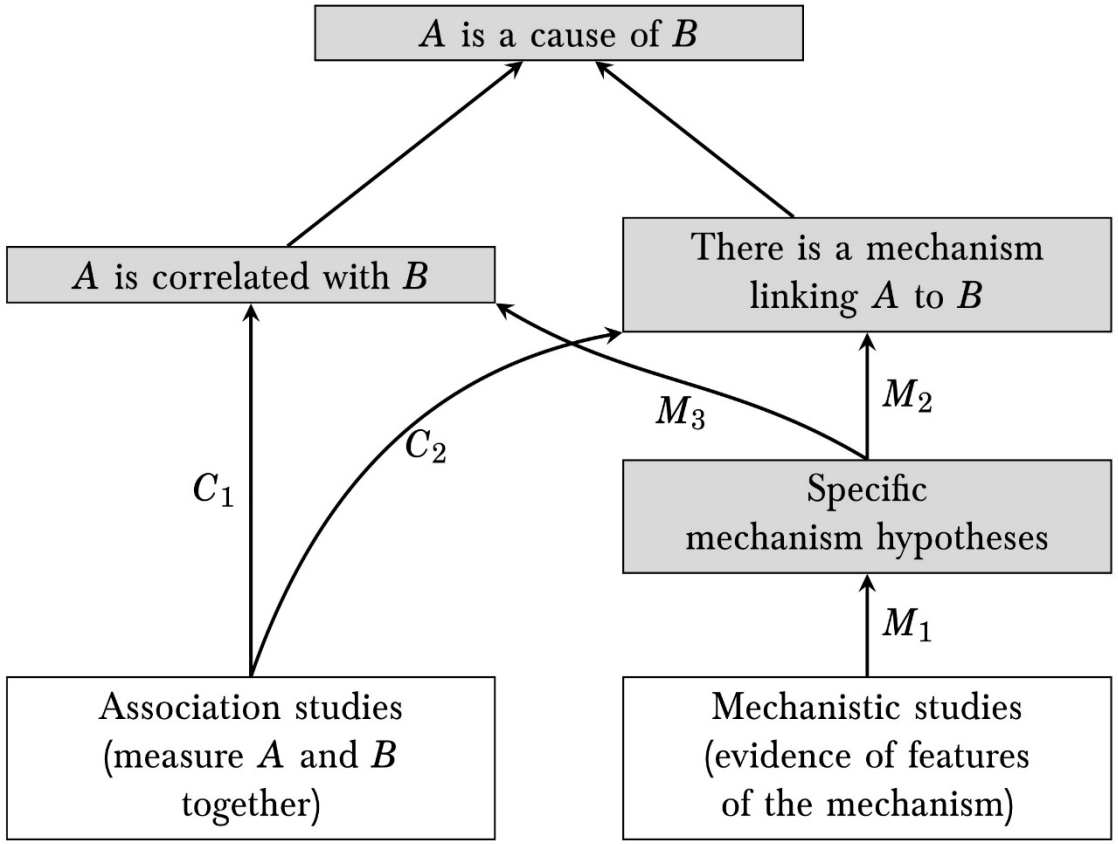
Person- or Encounter-Level Validity Testing

Table 4. For Person- or Encounter-Level Validity

Approach	Threshold
Sensitivity (true positive)	TBD
Specificity (true negative)	TBD
PPV or NPV	TBD
Accuracy	TBD
F1 Score	TBD
Area under the ROC curve	TBD

- Validity testing is only required for data elements that are not the “gold standard”
- There are no a priori thresholds for validity testing results
  - Rather make explicit any potential harms to persons and entities of low validity and argue how those harms might be mitigated
- Note trade-offs between validity and conformance to the intent of the concept of interest (e.g., exclusions)

Accountable Entity Validity Claims



- **A is a cause of B**
  - A: quality program and entity response
  - B: measure focus
- **Association claims**
  - A is correlated with B
- **General mechanism claims**
  - A is responsible for B
  - Accounts for the association

Source: Shan, Y., Williamson, J. (2023). Evidential Pluralism in the Social Sciences. United States: Taylor & Francis.

Accountable Entity Validity Claims (*cont.*)

Table 5. Substantiating Causal Claims

Claim	Association studies	Mechanism studies
Causal claim (A is a cause of B)	(Prone to bias)	(Prone to complexity)
Correlation claim (A and B are probabilistically dependent conditional on potential confounders C)	Test for an association between A and B (C1)	Features + unsystematic observations (M3)
General mechanistic claim (there is a complex of mechanisms that invokes A as partially responsible for B and that can account for the extent of the correlation)	Non-causal connections and confounding are ruled out (IBE) (C2)	Confirm existence of suitable mechanism (M2)
Specific mechanism hypothesis (posit features of such a mechanism complex)		Confirm presence of hypothesized features (M1)

Source: Shan, Y., Williamson, J. (2023). Evidential Pluralism in the Social Sciences. United States: Taylor & Francis.

Accountable Entity Validity Claims (*cont.*)

Table 3.1. Quality Levels of Evidence

Quality Level	Interpretation
High	Further research is highly unlikely to have a significant impact on our confidence in the claims
Moderate	Further research is moderately unlikely to have a significant impact on our confidence in the claims
Low	Further research is moderately likely to have a significant impact on our confidence in the claims
Very Low	Further research is highly likely to have a significant impact on our confidence in the claims
Unavailable	Further research is not possible

Table 3.3 Confidence Levels of Evidence

Quality Level	Interpretation		
	Independence	**Consistency**	**Robust**
	Multiple studies using different methods demonstrate similar associations	Multiple studies using similar methods demonstrate similar associations	Multiple studies across different contexts demonstrate similar associations
High	Yes	Yes	Yes
More likely than not	Yes	Yes	No
	Yes	No	Yes
	Yes	No	No
	No	Yes	Yes
	No	Yes	No
	No	No	Yes
Low	No	No	No

Accountable Entity Validity Claims (*cont. 2*)

Table 3.2. Status of a Claim			
Status	**Interpretation**	**Quality Level**	**Confidence Level**
Established	Community standards are met for adding the claim to the body of evidence (i.e., as evidence for other claims)	High	High
Provisional established		Moderate	High
Arguably true		Moderate	Claim more likely than not
Speculative	None of the other categories		
Arguably false		Moderate	Negation more likely than not
Provisionally ruled out		Moderate	High
Ruled out	Community standards are met for adding the negation of the claim to the body of evidence	High	High

Document	Description
Endorsable	Importance, validity, and usability are all either established, provisionally established, or arguably true
Potentially endorsable	Niether endorsable nor unlikely endorsable
Unlikely endorsable	Importance, validity, and usability are all speculative or ruled out, provisionally ruled out, or arguable false

Accountable Entity Validity Claims (*cont. 3*)

Table 6. Other Possible Explanations to Rule Out

Possible Explanations	Description
Causation	A is a cause of B (mechanism)
Reverse causation	B is a cause of A
Confounding	C is a common cause of both A and B (risk-adjustment)
Performance bias	A group identified and treated differently than not A group
Detection bias	B is measured differently in A group than in not A group
Chance	Random (reliability)
Fishing	Association between A and some B
Temporal trends	A and B change over time for independent reasons
Semantic relationships	A and B have overlapping meaning
Constitutive relationships	A is a component of B
Logical relationships	A and B are logically overlapping
Nomological (law) relationships	Association between A and B due to a natural law
Mathematical relationships	A = B + C

Example 3: Validity and Importance

Table 7. Importance of CBE #4125 - Thirty-day Risk-Standardized Death Rate among Surgical Inpatients with Complications (Failure-to-Rescue)

	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
Performance Score	0.0119	0.0299	0.0407	0.0518	0.0615	0.071	0.0824	0.0954	0.1137	0.1713
N of Entities	227	226	226	227	226	226	227	226	226	226
N of Persons/ Encounters/ Episodes	26,978	64,721	64,121	68,896	61,062	59,418	58,308	66,919	38,603	30,274
Adoption (assumed)	100%	90%	80%	70%	60%	50%	40%	30%	20%	10%

- If entities performed at the benchmark (Decile 3) or better, a potential decrease in the number of adverse events of 59.6% (37,172 to 20,474)
- However, assuming an existing adoption-implementation rate and effect size of 20%, a potential decrease in the number of adverse events of only 8.2% (37,172 to 34,251)

Pitfalls and Mitigations

Pitfall: Not demonstrating person- or encounter-level validity

- **Mitigation:** Make an argument about the potential harms to persons and entities associated with low validity
  - Present plans current and future to address, including implementation and feedback
- **Mitigation:** Make an argument about the trade-offs between validity and feasibility
  - Provide an explicit articulation of the people, processes, and technology required for data collection and reporting
- **Mitigation:** Present any plans for future digital capture and reporting
  - In particular, the use of interoperability standards (e.g., USCDI+ and FHIR)

Pitfalls and Mitigations (*cont.*)

Pitfall: Not demonstrating accountable entity-level validity

- **Mitigation:** Face validity: Begin with a logic model of plausible mechanisms associated with an increase in the likelihood of the measure focus
  - Leverage the expertise and experience of your technical expert panel (TEP), including patients, to provide evidentiary support for your logic model
  - Best practices of doing so will be focus of future scientific methods panel (SMP)
- **Mitigation:** Use the logic model to inform the association studies and the mechanism studies to support the causal claim
  - Correlating two measures: Is there a common mechanism complex sufficient to explain?
  - Performance gap or known groups: Are other explanations ruled out?

Expectations for New vs. Maintenance

New Measures

- Person- or encounter-level validity testing
- Data within the last 5 years
- Explicit articulation of harms associated with low validity and plans for mitigation of such harms

Maintenance Measures

- Accountable entity-level validity testing
- Data within the last 5 years
- Importance table by decile (implementation, adoption, effect)
- Plan for mitigation for low-validity persons and entities (i.e., further association and/or mechanism studies)

Resources

- **Messick, S. (1989b). Validity. In R. L. Linn (Ed.), Educational measurement (3rd ed., pp. 13-103). New "York: Macmillan** – articulation of “modern” validity theory
- **Shan, Y., Williamson, J. (2023). Evidential Pluralism in the Social Sciences. United States: Taylor & Francis** – “evidential pluralism” argues that both association and mechanism evidence are necessary to support casual claims
- **Cook, David A., and Thomas J. Beckman. "Current concepts in validity and reliability for psychometric instruments: theory and application." The American journal of medicine 119.2 (2006): 166-e7** – further reading on “modern” validity theory (“argument-based validity”)
- **Edwards, Michael C., et al. "Fit for purpose and modern validity theory in clinical outcomes assessment." Quality of life research 27 (2018): 1711-1720** – further reading on “modern” validity theory (“argument-based validity”)

Risk Adjustment

Jeffrey Geppert

Risk Adjustment (cont.)

- The purpose of risk adjustment is to “rule out” that the variation in the measure focus (B) between entities (A) is due to other factors (besides the mechanism)
  - C is a common cause of both A and B (risk adjustment)
- Other factors may include (conceptual model demonstrates a good model)

Factors	Factors
Pre-ecosystem (built environment)	Behavioral
Demographic	Access (selection)
Clinical	Post-ecosystem (e.g., community supports)
Functional	-

- Empirical metrics are largely relative (e.g., calibration, discrimination)

Risk Adjustment (*cont.2*)

Table 8. Comparative Risk-Adjustment Descriptive Table

Category	Feature	Values
Measure	Measure type	Outcome, intermediary outcome, patient-reported outcome
	Independent variable	Binary, continuous
Degrees of Freedom	Model type	Logistic, linear
	Model level	Person, entity, hierarchical
	Number of persons	N
	Number of entities	K_1
	Number of model factors	K_2
Performance Metrics	Discrimination	C-statistic
	Calibration	Plot, Brier score, Hosmer-Lemeshow
	Goodness-of-fit	R-squared, AIC, deviance, chi-square

Pitfalls and Mitigations

Pitfall: Not justifying the risk-adjustment approach

- **Mitigation:** Provide a conceptual model for the risk-adjustment model
  - Include justification for the factors included and not included (i.e., independence from the mechanism complex)
- **Mitigation:** Distinguish between non-hierarchical and hierarchical models
  - Report descriptive statistics on those entities and persons “shrunk” to the shrinkage target (including the degree of shrinkage)
  - For hierarchical models (mixed logistic regression), report the “between entity variance” to enable calculation of entity-level reliability metrics

Expectations for New vs. Maintenance

New Measures

- Risk-adjustment conceptual model
- Measure logic model

Maintenance Measures

- Risk-adjustment model performance statistics
- Data within the last 5 years

Resources

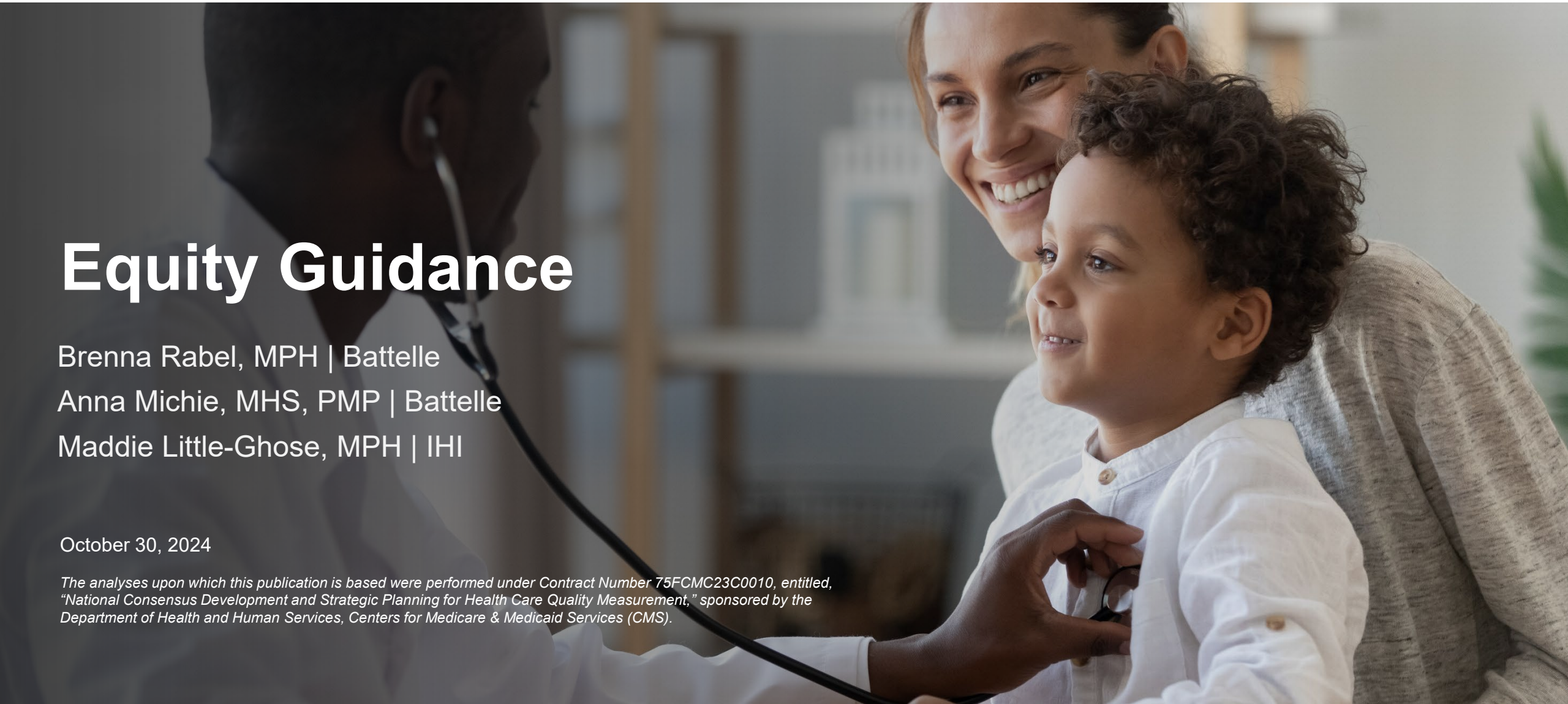
- **Iezzoni, L. I. (2013). Risk Adjustment for Measuring Health Care Outcomes. United States: Health Administration Press** – a definitive text on risk adjustment.
- **MMS Hub (https://mmshub.cms.gov)** – useful references on developing and implementing risk-adjustment models.
- **Endorsement and Maintenance Guidebook** – This document contains comprehensive information pertaining to the Endorsement and Maintenance process, including conditions and non-negotiables for each domain, as well as a detailed measure evaluation rubric that outlines a successful submission for each domain.
- **PQM Measure Evaluation Rubric Worksheet** – This document is used by reviewers to guide their assessments of measures under review for initial endorsement or maintenance. It includes PQM evaluation criteria and key considerations for reviewers' assessment of measures.

Questions & Answers

Partnership for Quality Measurement

Powered by Battelle

Partnership for Quality Measurement
Powered by Battelle



Equity Guidance

Brenna Rabel, MPH | Battelle
Anna Michie, MHS, PMP | Battelle
Maddie Little-Ghose, MPH | IHI

October 30, 2024

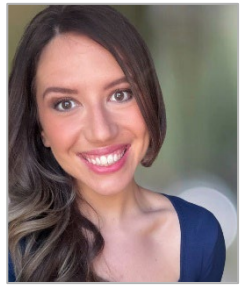
The analyses upon which this publication is based were performed under Contract Number 75FCMC23C0010, entitled, "National Consensus Development and Strategic Planning for Health Care Quality Measurement," sponsored by the Department of Health and Human Services, Centers for Medicare & Medicaid Services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

Meet the Presenters

Brenna Rabel | CBE Technical Director

- Facilitates collaboration across Consensus-Based Entity (CBE) activities to ensure consistency and excellence
- 10+ years' health care, public health, and quality experience

Anna Michie | E&M Deputy Task Lead



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

Maddie Little-Ghose | IHI Project Manager

- Supports guidance development related to the Equity domain
- 9+ years as a public health professional

Session Objectives

Purpose	To introduce equity guidance in preparation for the transition to a required domain in Spring 2025.
Agenda	- Introduce and describe the importance of an Equity domain - Explain elements of equity - Q&A

Objectives and Principles of Equity in Quality Measurement

Brenna Rabel, Technical Director

Introduction

- Recent federal priorities underscore the importance of health equity in quality measurement.[1,2]
- The Partnership for Quality Measurement (PQM) plays an integral role in shaping the health care quality measurement landscape.
- To that end, an Equity domain is now part of PQM's measure endorsement and maintenance (E&M) process.
- While previously an optional domain, **beginning in Spring 2025, Equity is required for all new and maintenance measures** submitted to Battelle for PQM endorsement consideration.

1. Executive Order 13985, "Advancing Racial Equity and Support for Underserved Communities Through the Federal Government," issued in January 2021.
2. Centers for Medicare & Medicaid Services. (2022). CMS Framework for Health Equity 2022-2032. Retrieved from https://www.cms.gov/files/document/cms-framework-health-equity.pdf

Importance of An Equity Domain

1. Supports recent national priorities of achieving health equity.
2. Encourages the collection of comprehensive data across different subpopulations, so that developers can assess where disparities exist.
3. Promotes accountability among health care providers and systems to deliver equitable care by highlighting existing disparities.
4. Builds trust and credibility among the patient community (e.g., patients, caregivers, advocates) by identifying potential disparities.

Using the Equity Guidance, Developers/Stewards Will Be Able To:

- Describe health equity in the context of quality measurement.
- Describe the intent and key elements of the Equity domain.
- Identify examples of common methods, pitfalls, and mitigation approaches for supporting the claim that the measure contributes to reducing disparities in health care and health outcomes.
- Describe how quality measures can be used by interested parties (providers, patients, implementers, policymakers) to discern disparities in care, thereby reducing disparities and improving equity.

Defining Health Equity

> ***Health equity*** *is the attainment of the highest level of health for all people, where everyone has a fair and just opportunity to attain their optimal health regardless of race, ethnicity, disability, sexual orientation, gender identity, socioeconomic status, geography, preferred language, or other factors that affect access to care and health outcomes.*

– Centers for Medicare & Medicaid Services. (n.d.). Health equity. CMS.gov. https://www.cms.gov/priorities/innovation/key-concepts/health-equity

Clinical quality measurement is one of many potential strategies to address inequality, inequity, and injustice in health care. **By evaluating measures in terms of their net benefit for subpopulations of persons and entities**, the PQM community can promote the development and implementation of measures that **drive improvements in care for all persons**.

Guiding Principles for Equity

PQM's guiding principles for equity are as follows:

Principle	Description
1. Alignment with national efforts	The Equity domain should align with national efforts to advance health equity and reduce health care disparities.
2. Feasibility for developers	The domain should take into consideration what is feasible for developers and stewards, noting any limitations with data sources, methods, etc., and how developers can overcome these limitations.
3. Usable for accountable entities	Developers and stewards should consider and describe how accountable entities will use the measure to improve health disparities.
4. Comprehensible for users	Users (e.g., measure reviewers, measure implementers, the public) should be able to review the information presented under the Equity domain and understand how the measure can advance health equity and identify existing gaps in disparate populations.

Elements of the Equity Domain

Anna Michie, Deputy E&M Task Lead

Element 1

1. Describe how this measure contributes to efforts to address inequities in health care delivery and health outcomes.

a. Describe how the measure aligns with the historical background of the measure focus area and existing health disparities, including historically underserved populations.

b. Identify the subpopulation(s) that should be considered to address inequities in this measurement area.

c. Describe current or past efforts to address these inequities and if disparities remain.

d. Summarize relevant literature, internal empirical analyses, and qualitative data for this element.

Pitfalls and Mitigations

Pitfall: Determining which subpopulations to consider

- **Mitigation:** Element 2 includes a minimum list of the sociocontextual variables that can be used to define subpopulations. Additional subpopulations (e.g., that have been noted in the literature as being historically underserved or marginalized) should also be considered.

Pitfalls and Mitigations (*cont.*)

Pitfall: Lack of Evidence on Disparities

- **Mitigation:** Measure developers should provide an explanation of how they determined that there was no evidence of disparities (e.g., provide a summary of findings from a literature review, internal empirical analyses, and/or qualitative data).
- **Mitigation:** Available evidence may be lacking for certain subpopulations, or the literature may draw conclusions reflective of only non-marginalized populations. Measure developers and stewards should make a conscious effort to identify evidence that has an inclusive population or identify this as a potential weakness in the data available.

Element 2

2. *[For maintenance measures]* Provide a description of your methodology and approach to empirical testing of differences in performance scores across identified sociocontextual variables.

- Required for maintenance measures, optional for new measures if the developer has available data/information.
  a. Provide an overview of the methodology for stratifying data across various sociocontextual variables.
  b. When selecting which method to use, consider the type of data (e.g., continuous, discrete/categorical, proportions) and how those data are distributed.

Sociocontextual Variables

- Age
- Race
- Ethnicity
- Sex
- Language
- Gender Identity
- Insurance coverage
- Sexual orientation
- Indicator of urbanicity/rurality (e.g., zip code)
- Indicator of disability (e.g., frailty)
- Indicator of socioeconomic status (e.g., SES indices)

Pitfalls and Mitigations

Pitfall: Data limitations with sociocontextual variables

- **Mitigation:** Collaborate with other health care providers, research organizations, or community groups to access additional data.
- **Mitigation:** Assess if available variables can be used as proxies for the missing data.
- **Mitigation:** Apply statistical methods to estimate the missing values based on the patterns and relationships observed in available data (e.g., multiple imputation or regression imputation).

Element 3

3. *[For maintenance measures]* Provide the results and an interpretation of the results, including an explanation of any differences in performance scores across subpopulations, how the results relate to the evidence in Element 1, any limitations of the results, and the potential impact of these differences on the identified subpopulations.

- Required for maintenance measures, optional for new measures if the developer has available data/information
  a. Describe any differences found, including whether they are statistically significant or not.
  b. Consider the clinical significance (i.e., practical importance of the difference) and the community significance (i.e., relevance to identified subpopulations).
  c. Summarize how identified differences correspond to the evidence summarized in Element 1, providing rationale for why their findings are starkly different than the evidence in Element 1.
  d. Discuss any potential limitations of the results, including whether the results are generalizable/representative.

Pitfalls and Mitigations

Pitfall: Lack of statistically significant results

- **Mitigation:** Explain the clinical significance (practical importance of the measure) and community significance of findings (relevance to identified subpopulations), even if they aren't statistically significant.
- **Mitigation:** Dataset may not be of sufficient size or generalizable (see mitigations for Element 2).

Element 4

4. *[For maintenance measures]* Describe or provide evidence indicating how accountable entities can use these results to reduce identified disparities.

- Required for maintenance measures, optional for new measures if the developer has available data/information
  a. Summarize how accountable entities can use the stratified results to implement interventions to reduce disparities and assess effectiveness.
  b. To inform this summary, consider conducting an evaluation of the literature or engaging with accountable entities to identify interventions that have been or could be implemented.

Pitfalls and Mitigations

Pitfall: Lack of evidence to demonstrate how entities can improve underlying inequities

- **Mitigation:** Engage with accountable entities to gain insights into their challenges/experience. This engagement can involve interviews, surveys, or workshops to gather insights on past experiences, challenges faced, and successful strategies. This real-world input is crucial for understanding the practical aspects of implementing interventions.
- **Mitigation**: Describe policy or system changes needed within communities based on identified disparities.

Expectations for New vs. Maintenance

New Measures

- Element 1: An assessment (using relevant literature, internal empirical analyses, qualitative data) of how the measure contributes to efforts to address inequities in health care delivery and health outcomes.

Maintenance Measures

- Elements 1-4: An assessment of any new disparities and whether performance scores across previously identified subpopulations have improved. If no improvement, developers/stewards should provide a rationale as to why.

Questions & Answers

Breakout Sessions: Elements of Equity - Developer/Steward Insights on Enhancing Guidance and Potential Challenges

Maddie Little-Ghose, IHI Project Manager

Breakouts: Purpose and Logistics

- **Purpose:** To obtain developer/steward feedback on potential obstacles and areas to strengthen guidance for the four elements in the Equity domain
- **Length:** 20 minutes
- Participants will be randomly assigned to one of four breakout rooms
  - Each room will have a moderator to collect feedback
  - Each room will discuss all elements of the Equity domain
    - What challenges do you foresee in completing each element?
    - Are the new and maintenance requirements feasible?
    - What is missing from the guidance that should be added for clarity?
  - Feedback will be used to update and finalize Equity guidance
  - Key takeaways will be shared at the 4:45pm “Days End Review” session

Resources

- **Endorsement and Maintenance Guidebook** – This document contains comprehensive information pertaining to the Endorsement and Maintenance process, including conditions and non-negotiables for each domain, as well as a detailed measure evaluation rubric that outlines a successful submission for each domain.

- **PQM Measure Evaluation Rubric Worksheet** – Reviewers use this document to guide their assessments of measures under review for initial endorsement or maintenance. The worksheet includes PQM evaluation criteria and key considerations for reviewers' assessment of measures.

Partnership for Quality Measurement

Powered by Battelle

Partnership for Quality Measurement
Powered by Battelle

Break (3:00 – 3:15 PM ET)

Partnership for Quality Measurement
Powered by Battelle

Instrument-Based Measure Guidance

Matthew Pickering, PharmD | Battelle
Anna Michie, MHS, PMP | Battelle

October 30, 2024

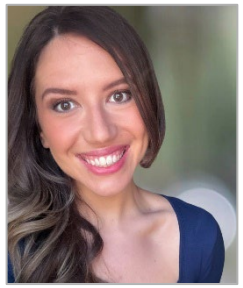
The analyses upon which this publication is based were performed under Contract Number 75FCMC23C0010, entitled, "National Consensus Development and Strategic Planning for Health Care Quality Measurement," sponsored by the Department of Health and Human Services, Centers for Medicare & Medicaid Services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

Meet the Presenters

Matthew Pickering | E&M Task Lead

- Oversees endorsement & maintenance (E&M) processes
- 10+ years' quality experience

Anna Michie | E&M Deputy Task Lead



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

Session Objectives

Purpose	Review PQM's policy on instrument-based measures, including considerations for submitting measures derived from instruments individually or as a composite, and operational details for completing Intent to Submit (ITS) and Full Measure Submission (FMS).
Agenda	- Overview of consensus-based entity (CBE) policy - Aggregating into composite vs. individual measures - Feedback from Consumer Assessment of Healthcare Providers and Systems (CAHPS) developers/stewards - Q&A

What are Instrument-Based Clinical Quality Measures?

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

cahps®
Surveys and Tools to Advance Patient-Centered Care

Hospice Outcomes and Patient Evaluation
HOPE Timepoints

HOSPICE
QUALITY REPORTING PROGRAM

CBE Policy on Instrument-based Clinical Quality Measures

- The CBE does not review or endorse instruments or surveys. Rather, the CBE reviews and endorses clinical quality measures derived from instruments or surveys.
- Endorsing the individual measures derived from a survey instrument allows:
  - More specific evaluations without conflating endorsement decisions
  - Actionable feedback and opportunities for improvements at the instrument-derived measure level
  - Alignment with CMS initiatives (e.g., CMIT, MERIT)
- The CBE Policy on Instrument-based Clinical Quality Measures can be found in Appendix G of the E&M Guidebook.

Policy Specifics

- There are no differences in the requirements or criteria for endorsement between instrument-based clinical quality measures and other clinical quality measures.
- Each clinical quality measure derived from an instrument or survey is reviewed and endorsed separately.
- These measures must be specified and tested at the accountable entity level (e.g., clinician or facility).
- For person- or encounter-level (i.e., data element) reliability and validity, measure developers/stewards may cite existing literature to substantiate those properties.
- Measures developers/stewards are also encouraged to attest that the instrument or survey was developed using a best practice protocol (e.g., Holmbeck, 2009).

1. 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.

Composite Measures

- 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.
- Composite measures:
  - Combine two or more measures to provide an even more effective glimpse into multiple dimensions of quality
  - Combine results into a single score
    - Types of scoring methods including all-or-none/any-or-none, sum, average, weighted average, or opportunity scoring
  - Individual component measures should have consistent score types and scales

Feedback from Developers and Stewards

Concerns with Duplicative Work

- Some developers were concerned that this approach will lead to duplicative work, as some information is at the survey-instrument level and is the same across instrument-derived measures (e.g., feasibility).
  - Battelle will use logic within the ITS/FMS forms to auto-populate relevant information from the survey instrument to the instrument-derived measure forms.

Splitting Measures Across Cycles

- Due to the large number of instrument-derived measures (e.g., 19) within some surveys, Battelle discussed reviewing instrument-derived measures across two E&M cycles.
- Most developers preferred to review measures linked to the same instrument in a single cycle, citing changes in committee members, data availability, and burden as reasons.
  - Battelle will work with developers to determine what approach would work best for them.

Endorsement Status

- Data may not be as robust for all instrument-derived measures within a survey instrument. Endorsing instrument-derived measures individually prevents one instrument-derived measure's limitations from affecting other instrument-derived measures.
- Enables focused feedback concerning instrument-derived measures, accurately assesses each instrument-derived measure's effectiveness, and highlights necessary changes at the measure level that impact endorsement decision.

How Will Measures be Submitted?

- Each instrument-derived measure will have a unique ID tied to a survey instrument, which will have its own unique ID
  - Example:
    - Survey Instrument ID: CBE #1234
    - Instrument-derived Measures: CBE #1234-1; CBE #1234-2; CBE #1234-X
- Developers will fill out an ITS for the survey instrument with information specific to the survey. They will then be able to open a new ITS form for each instrument-derived measure, which when linked to the survey-instrument via ID number will allow relevant information from the survey information to be inherited.
- The goal of inheriting data is to decrease the amount of duplicative information across the instrument-derived measures.

How Will Measures be Submitted? (*cont.*)

ITS Submission Process

1. Create survey instrument ITS.
2. After the survey instrument ITS is created, you will be able to link instrument-derived measure ITS submissions to the survey instrument ITS.
3. Complete ITS fields and make sure to save progress. Relevant content from the survey instrument ITS will be transferred to the measure ITS automatically after saving.
4. Submit survey instrument and instrument-derived measure ITS for completeness check.

FMS Submission Process

1. Begin FMS for survey instrument by completing FMS fields and saving progress. Relevant content from the survey instrument FMS will be transferred to each measure FMS automatically after saving.
2. Begin FMS for each instrument-derived measure by completing FMS fields and saving progress.
3. Submit survey instrument FMS and the FMS for each instrument-derived measure for completeness check.

Key Takeaways

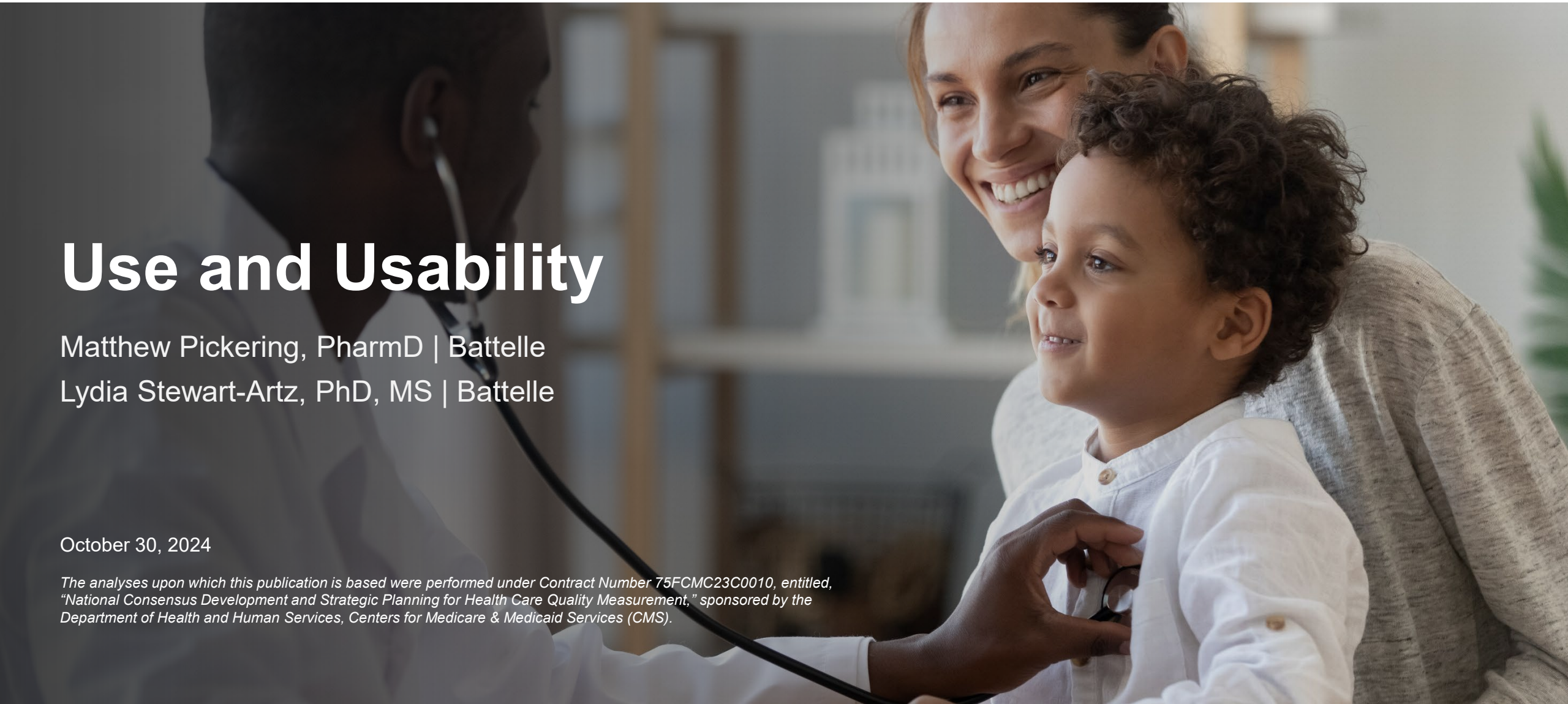
- Each clinical quality measure derived from an instrument or survey is reviewed and endorsed separately.
- The submission process will start with the creation of a survey-instrument ITS with relevant information carried over to instrument-derived measures.
- Instrument-derived measures will be allowed to split across cycles.
- Measure developers/stewards are encouraged, where appropriate, to combine individual instrument or survey items into a person/respondent-level “composite.”
- As the new process is implemented, we will monitor potential pitfalls over time for continued improvement/education.
- E&M project staff are available for technical assistance to measure developers/stewards in the application of this policy. Please reach out to PQMSupport@battelle.org.

Questions & Answers

Partnership for Quality Measurement

Powered by Battelle

Partnership for Quality Measurement
Powered by Battelle



Use and Usability

Matthew Pickering, PharmD | Battelle
Lydia Stewart-Artz, PhD, MS | Battelle

October 30, 2024

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

Meet the Presenters

Matt Pickering | E&M Task Lead

- Oversees Endorsement and Maintenance (E&M) processes
- 10+ years' quality experience

Lydia Stewart-Artz | PRMR/MSR Evaluation Lead

- Oversees Pre-Rulemaking Measure Review (PRMR)/Measure Set Review (MSR) measure evaluation and committee initial review
- 5+ years' quality experience

Session Objectives

Purpose	To explain expectations of the Use & Usability domain, including new versus maintenance measures; define “accountability application”; and discuss common pitfalls and mitigations.
Agenda	- Describe Use and Usability submission requirements - Explain expectations for new vs. maintenance measures - Describe common pitfalls and discuss mitigation approaches - Q&A

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.
- The purpose the E&M process is not to state whether a measure is appropriate for a specific use but rather to determine if the measure is safe and effective.
  - This means that 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.
  - The PQM Measure Evaluation Rubric looks to see if the measure is currently or has a plan for use in an **accountability application**.
  - Other processes evaluate measure fit for a certain program (e.g., Pre-Rulemaking Measure Review [PRMR]).

Examples of Accountability Uses

Measure scores can be used to drive decision making. Some examples include:

- **Public Reporting:** The results of health care quality measures for different hospitals are published on a public website. Patients can use this information to choose hospitals for their medical needs, influencing hospital selection based on publicly reported performance data.
- **Performance-based Payment:** A health insurance company implements a value-based purchasing program where provider reimbursement levels are directly tied to their performance on quality measures such as patient outcomes and efficiency.
- **Accreditation:** A health care accreditation body uses performance on quality measures as a criterion for accreditation. Hospitals must meet certain thresholds on these measures to receive accreditation, which is crucial for maintaining their operational licensing and patient trust.
- **Network Inclusion/Exclusion:** A health insurance network includes only those primary care providers in its network who meet specific quality benchmarks for preventive care and chronic disease management. Providers failing to meet these benchmarks may be excluded from the network, affecting their patient base and revenue.
- **Confidential Reporting:** A hospital system internally reports the performance of each department on quality measures related to patient safety. This confidential report is used by hospital management to identify areas needing improvement and to develop targeted interventions without making the information public.

Use

- Starting for Fall 2024 cycle, Use is collected during Intent to Submit (ITS)
- Select whether the measure is currently in use (yes/no)
- Identify all current or planned uses of the measure:
  - 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)

Use (*cont.*)

- For measures currently in use, describe the program including:
  - Name of the program/sponsor
  - URL of the program
  - Purpose of the program
  - Geographic area and percentage of accountable entities and patients included
  - Applicable level of analysis and care setting

Name of the program and sponsor

Centers for Medicare & Medicaid Services (CMS) Accountable Care Organization Realizing Equity Access, and Community Health (ACO REACH) Model

URL of the program

https://www.cms.gov/priorities/innovation/innovation-models/aco-reach

Purpose of the program

The ACO Realizing Equity, Access, and Community Health (ACO REACH) Model provides novel tools and resources for health care providers to work together in an accountable care organization (ACO) to improve the quality of care for people with Traditional Medicare in underserved communities and make measurable changes to address health disparities. Additionally, the model uses an innovative payment approach to better support care delivery and coordination for people in underserved communities.

Geographic area and percentage of accountable entities and patients included

The ACO REACH model for 2023 consisted of 132 ACOs, including 131,772 providers and 2.6 million patients, across the United States (click here for map of currently participating ACOs). The TFU measure is calculated for all eligible ACOs in the ACO REACH model.

Applicable level of analysis and care setting

Level of Analysis: Accountable Care Organization Care Settings: Hospital: Outpatient, Clinician Office/Clinic, Home Health, Hospital: Critical Access, Emergency Department, Hospital: Inpatient, Rural Emergency Hospital.

Pitfalls and Mitigations

Pitfall: Program information (lacking detail/inconsistent)

- **Mitigation:** Clearly explain the program purpose: Who is it serving and what are the goals?
- **Mitigation:** Describe the number of accountable entities and where they are located.
- **Mitigation:** Ensure level of analysis and care setting(s) for the program are consistent with the measure specification and testing.

Expectations for New vs. Maintenance

New Measures

- There is a plan for use in at least one accountability application after initial endorsement but before the measure's first maintenance review.*

Maintenance Measures

- The measure is currently in use in at least one accountability application.
- If the measure is not currently in use in at least one accountability application, a short-term plan (i.e., within 1 year) must be described.

**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.*

Usability – All Measures

- A measure needs to be practical and provide clear information that entities can use to enhance their performance, thereby improving the quality of health care provided.
- Describe actions measured entities can take to improve performance
  - Evidence-based actions (e.g., clinical practice guidelines, studies from the literature)
  - How difficult are these actions to achieve?
    - Level of burden in relation to appropriate clinical care
    - Resource or policy constraints/considerations
  - How can a measured entity overcome these difficulties?

Usability – Measures Currently in Use

- Feedback on measure performance
  - Describe implementation and measure performance feedback
    - When was the measure implemented?
    - Any implementation issues reported by measured entities? How many?
  - How was feedback obtained? (e.g., interviews, surveys, helpdesk)
  - How often is feedback collected? (e.g., ongoing, annually)
- Consideration of measure feedback
  - Explain how feedback was considered/incorporated into the development and revision of the measure
    - Was the measure modified? Why or why not?

Usability – Measures Currently in Use (cont.)

- Progress on improvement
  - Detail any improvements in performance score trends, numbers/percentages that have occurred over time. Include:
    - Number and percentage of accountable entities, patients included, and geographic area
    - Number and percentage of patients receiving high-quality care
  - Discuss how improvements apply to sub-populations (e.g., if stratified)
  - If no demonstrated improvements, provide an explanation as to why
- Unexpected findings
  - Unintended consequences or impacts to patients

Pitfalls and Mitigations

Pitfall: Measure does not show substantive improvement over time

- **Mitigation:** Offer an explanation as to why (e.g., resource issues, sample size/data limitations).
- **Mitigation:** Explore whether additional implementation guidance is needed for measured entities.

Expectations for New vs. Maintenance

New Measures

- Performance scores yield actionable information that can be used to improve performance among measured entities.

Maintenance Measures

- 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.

Resources

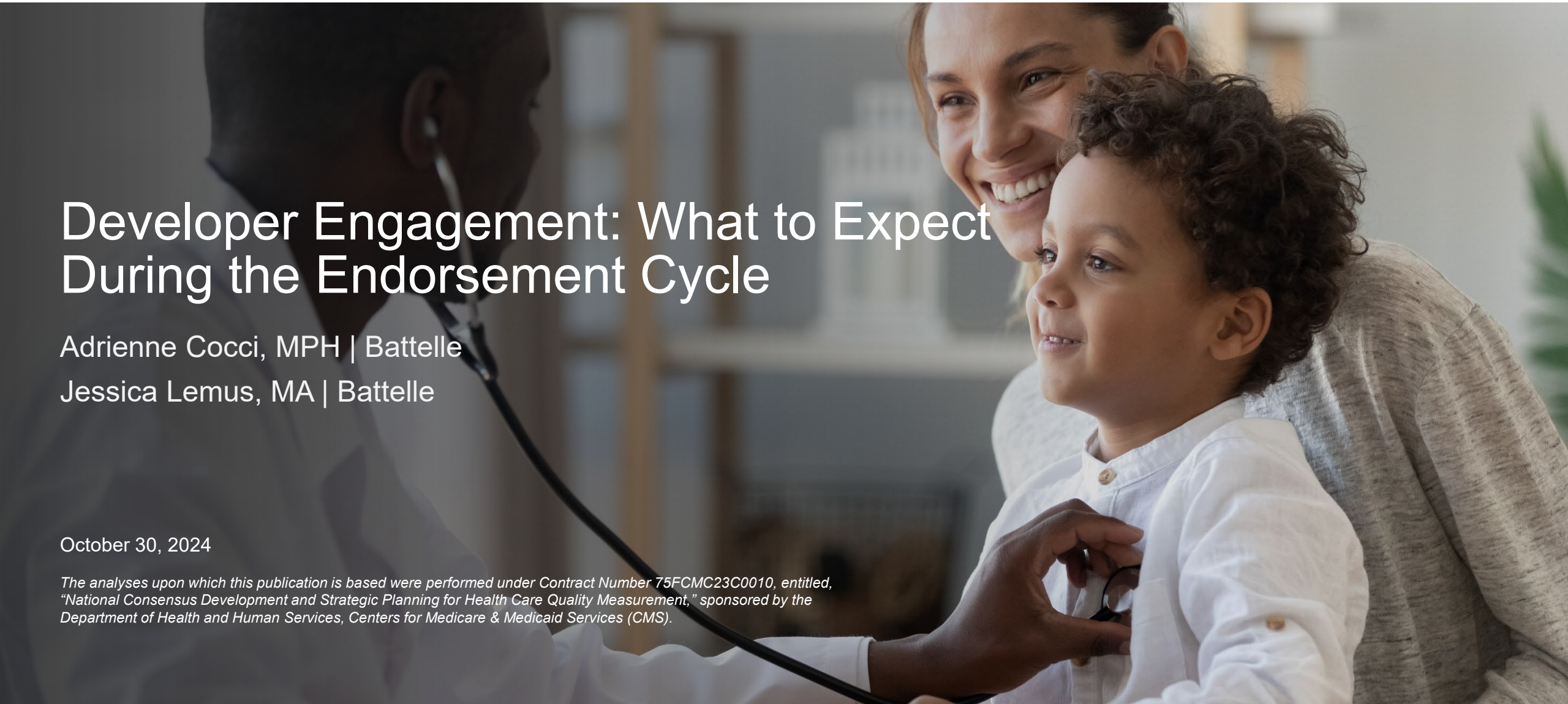
- **Endorsement and Maintenance Guidebook** – This document contains comprehensive information pertaining to the Endorsement and Maintenance process, including conditions and non-negotiables for each domain, as well as a detailed measure evaluation rubric that outlines a successful submission for each domain.

- **PQM Measure Evaluation Rubric Worksheet** – Reviewers use this document to guide their assessments of measures under review for initial endorsement or maintenance. It includes PQM evaluation criteria and key considerations for reviewers' assessment of measures.

Questions & Answers

Partnership for
Quality Measurement
Powered by **Battelle**

Partnership for Quality Measurement
Powered by Battelle



Developer Engagement: What to Expect During the Endorsement Cycle

Adrienne Cocci, MPH | Battelle

Jessica Lemus, MA | Battelle

October 30, 2024

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

Meet the Presenters

Adrienne Cocci | Database Administrator

- Maintains Submission Tool and Repository (STAR) measure database
- 10 years' public health research and evaluation

Jessica Lemus | Engagement Lead

- Supports Endorsement and Maintenance (E&M) stakeholder engagement
- 7 years' health research experience

Session Objectives

Purpose	To review the steps within the endorsement cycle, specify engagement opportunities for developers/stewards, and provide tips for effective navigation.
Agenda	- Review the Endorsement and Maintenance (E&M) cycle timeline and steps - Discuss how developers can engage in the E&M process throughout the endorsement cycle - Present tips and suggestions for navigating the E&M cycle - Q&A

What is Developer Engagement?

Measure Developer Engagement	Developer engagement refers to the active participation of measure developers/ stewards with the E&M process. This includes communicating with E&M project staff, attending endorsement meetings, responding to requests for information/review, etc.

Benefits of Developer Engagement

Guidance and Clarification

- E&M staff provide guidance, including explaining the endorsement criteria, timelines, and documentation required, which helps developers to navigate the process and avoid errors and/or delays.

Streamlined Communication

- Regular engagement ensures open communication for quick resolutions of any issues or questions that arise.

Feedback and Insights

- E&M staff and committee members can offer valuable feedback on how to improve measures or better align them with endorsement criteria.

Benefits of Developer Engagement (*cont.*)

Compliance and Assurance

- Engaging with project staff ensures measures comply with E&M requirements. Project staff can alert developers to compliance issues early in the process, allowing for timely adjustments.

Anticipating Challenges

- E&M Staff can help developers/stewards anticipate challenges or objections during the endorsement process, allowing developers to adjust their measures proactively.

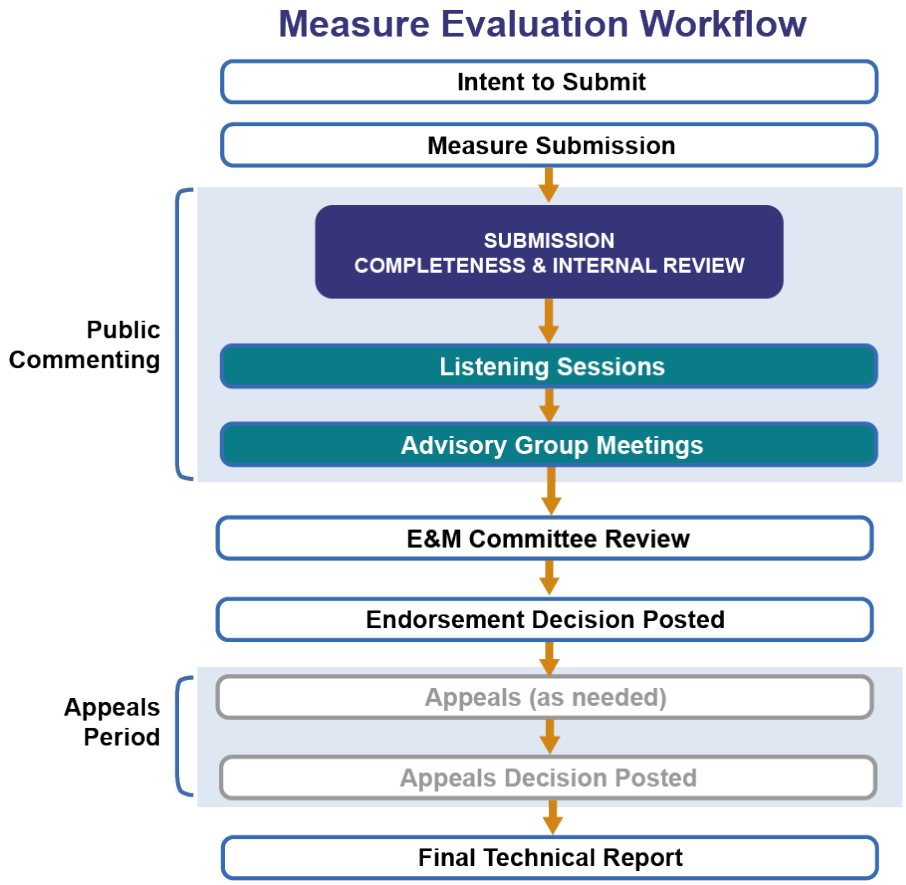
Accurate Information

- By responding to public comments and committee member feedback, developers/ stewards can maintain accuracy of information shared about their measures and offer clarifications.

E&M Process

Major steps:

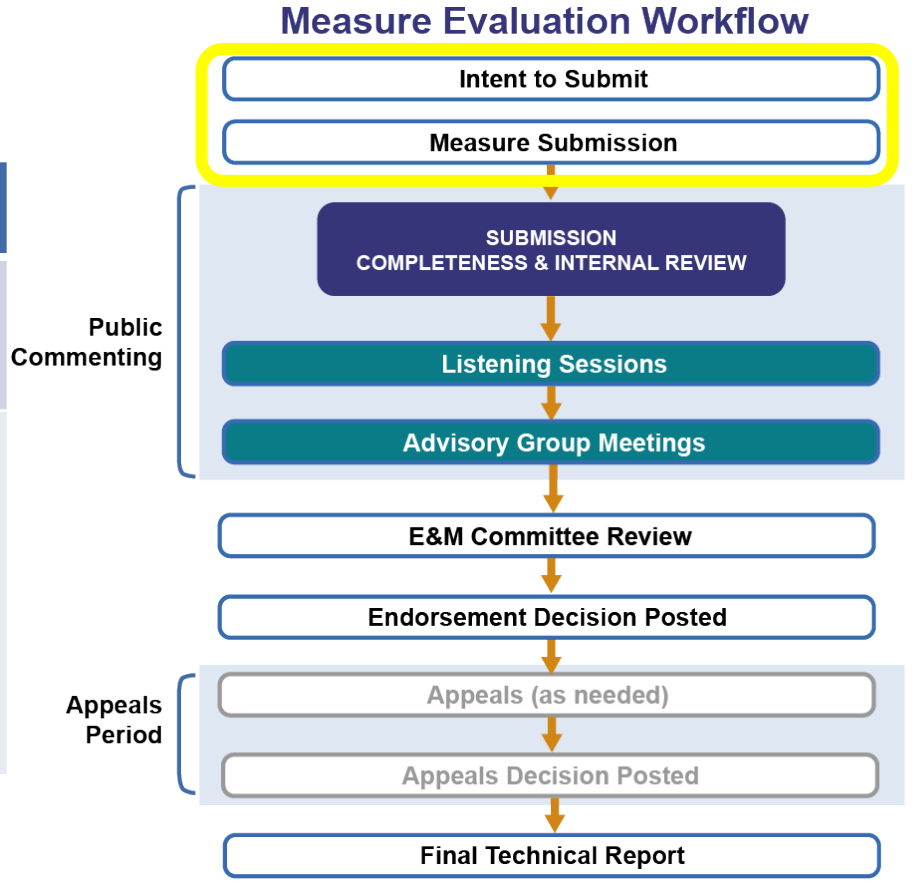
1. Intent to Submit
2. Full Measure Submission
3. Staff Assessments and Measure Public Comment Period
   a. Public Comment Listening Sessions
   b. Advisory Group Meetings
4. E&M Committee Review
5. Endorsement Decision
   a. Recommendation Group Meetings
6. Appeals Period (as warranted)



Intent to Submit & Full Measure Submission

Developers/stewards submit key measure information to Battelle.

Engagement Opportunity	*Battelle Action*	*Developer Action*
Confirm cycle assignment (maintenance measures)	Email communication confirming cycle assignments (3 months before ITS).	• Developers are asked to confirm cycle assignments for their measure(s).
Review ITS/FMS completeness check	• SMEs will conduct a completeness check at both the ITS and FMS stages. They will provide feedback through the PQM website. • Email communication of any comments on the Full Measure Submission	• Developers have the opportunity to respond to comments and update their submission.



Completeness Checks

The goal of the completeness check review is to ensure the measure submission information is complete for it to be reviewed for endorsement.

Feedback may include:

- Noting inconsistencies across fields (e.g., measure type does not align with measure title)
- Verifying differences from previous submissions (e.g., analysis level has changed)
- Highlighting missing fields
- Indicating any missing attachments (e.g., QMDSA Form, logic model)

From: **Partnership for Quality Measurement** <p4qm@battelle365.onmicrosoft.com>

Subject: CBEID X Intent to Submit Is Under Review

To: <no-reply@p4qm.org>

Hello Measure Submitter,

Thank you for submitting your Intent to Submit for CBEID X. Battelle staff are currently reviewing your submission. The team may provide feedback on your submission as a note in the system. View notes by logging in to PQM and selecting the CBEID for your measure. Once your measure opens, select "View Notes." You can reply to notes creating a thread similar to comments in a Word or PDF file.

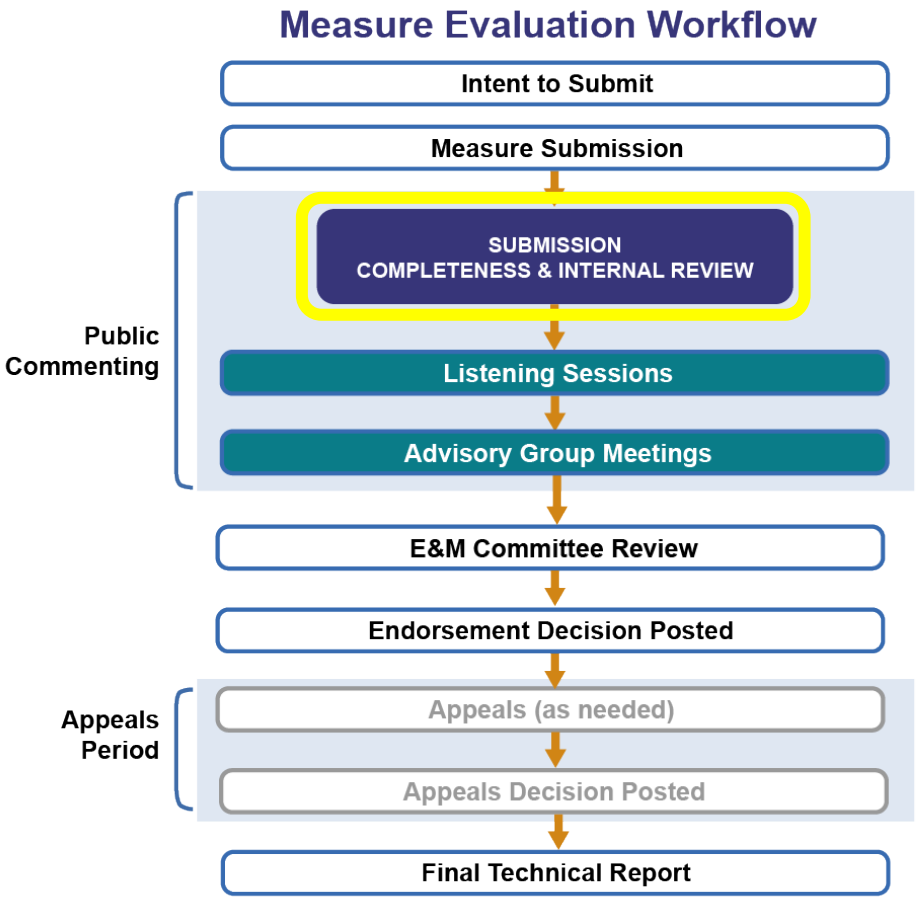
Please email PQM Support at PQMsupport@battelle.org if you have any questions.

Staff Assessments

Staff assessments are done using the PQM rubric (see E&M Guidebook) using the five domains: Importance, Equity, Scientific Acceptability (i.e., Reliability and Validity), Feasibility, and Use and Usability.

Engagement Opportunity	*Battelle Action*	*Developer Action*

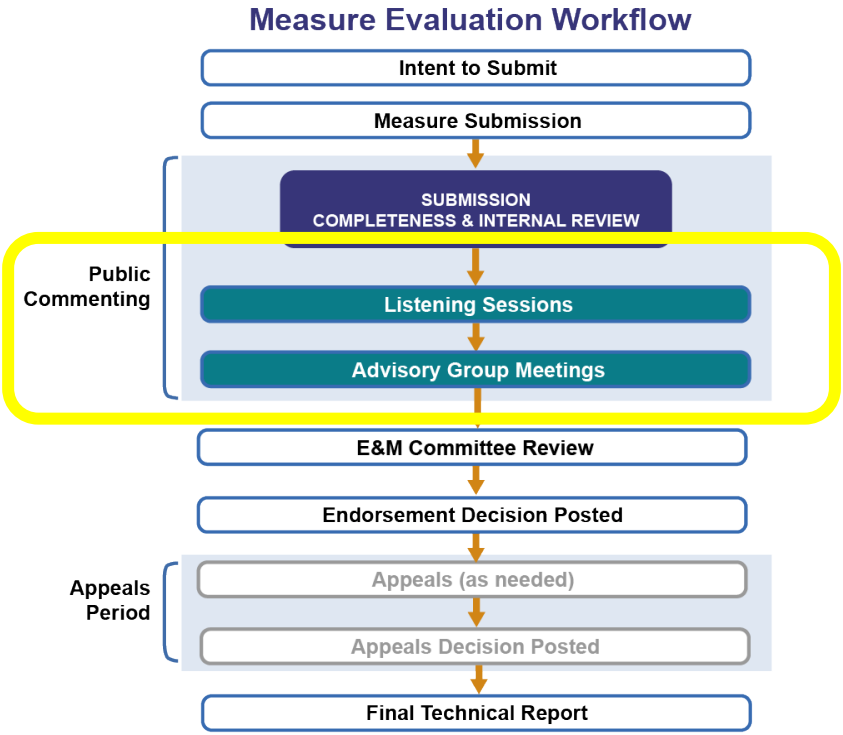
Factual review of staff assessments	Provide developers with a preliminary analysis of their submission and request developer's review for factual inaccuracies.
	
	These preliminary assessments summarize key points of the submission as they pertain to the PQM Measure Evaluation Rubric, and when appropriate, provide additional context or interpretation for certain aspects of the submission (e.g., verifying a testing methodology is appropriate).
		• Developers have an opportunity to respond to the staff assessment feedback by the set deadline.
		  • Were testing results accurately reflected?
		  • Are dates for testing or evidence correct?
		  • Is the stated type of reliability and/or validity method used and/or the level of testing (e.g., data element, measure score, facility level, clinician level) correct?



Public Comments, Listening Sessions, & Advisory Group Meetings

Public comments, listening sessions, and Advisory Group meetings are all opportunities to hear feedback on your measures including strengths and limitations.

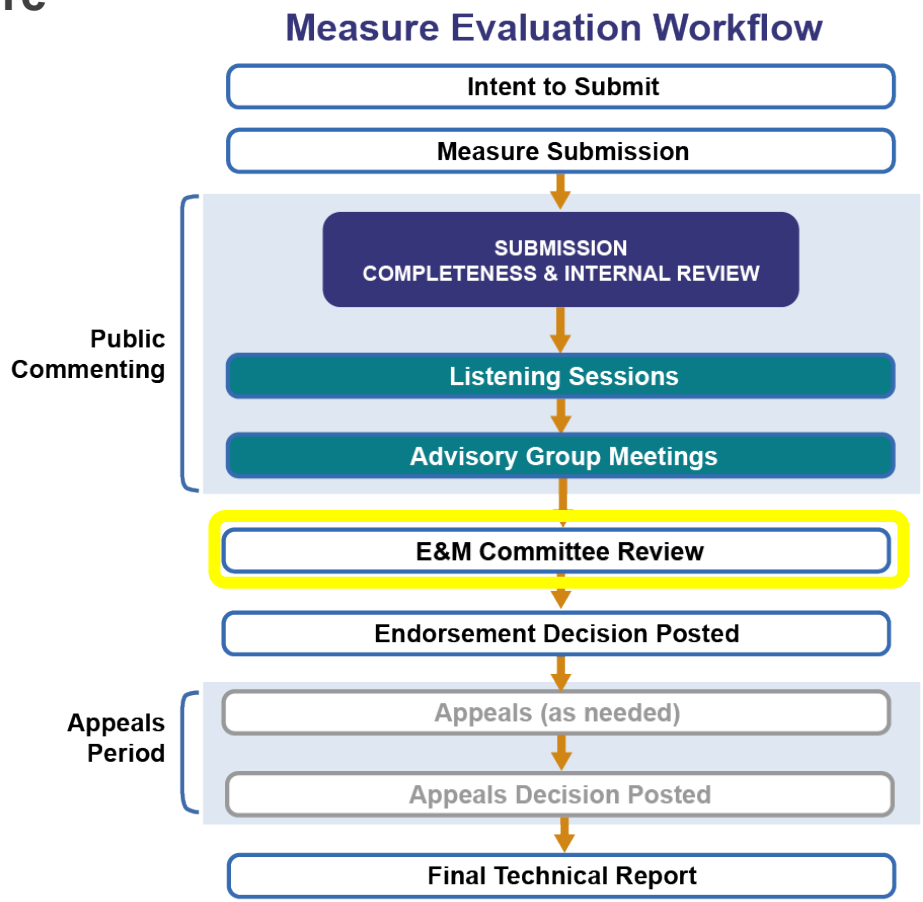
Engagement Opportunity	*Battelle Action*	*Developer Action*
Attend public comment listening session	Battelle sends meeting registration and link to posted materials. The listening session is an opportunity for any interested party to share verbal comments about measures under review.	Developers/stewards are encouraged to register and attend this call to listen to the comments shared. Developers/stewards will not be asked to address any comments during the call.
Attend Advisory Group meeting	Battelle sends meeting invite and links to posted materials. Advisory Group members raise questions and share perspectives verbally regarding the measures under review. No voting occurs.	Developers are encouraged to attend and respond in real time to Advisory Group feedback. Developers/stewards are also encouraged to invite their subject matter experts (SMEs) to support answering questions. Developers/ stewards may submit additional written feedback after the meeting.
Provide responses to public comments	Battelle will notify developers that the public comment period has closed.	Developers/stewards will have the opportunity to provide written responses to all public comments received. Please note, this is not a requirement.



Endorsement Committee Review

Recommendation Group members receive a summary of all measure feedback to date and conduct an independent review of measures using the PQM Measure Evaluation Rubric.

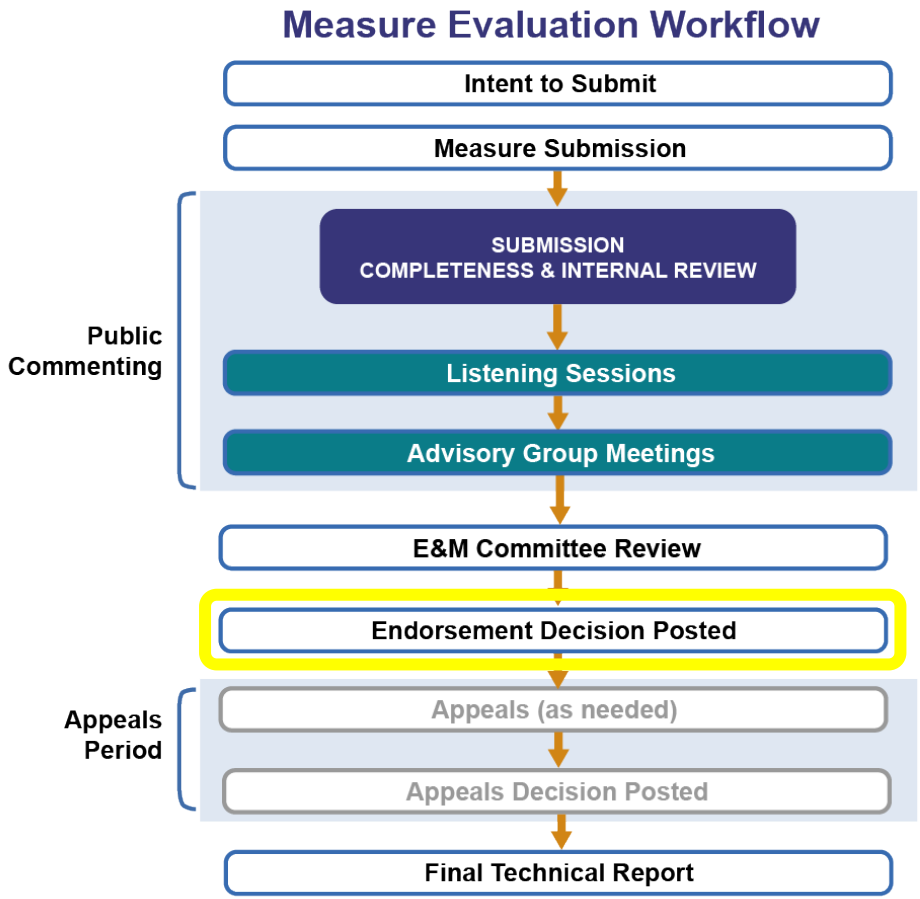
Engagement Opportunity	*Battelle Action*	*Developer Action*
Read committee independent reviews	Battelle staff aggregate information from public comments/developer responses, listening sessions, and Advisory Group meetings into a discussion guide, which is sent to the Recommendation Group prior to conducting their independent review. Committee independent reviews are then posted to the measure pages in the Submission Tool and Repository Measure Database before the endorsement meeting.	Developers review the aggregated results of the independent reviews prior to the Recommendation Group endorsement meeting to anticipate where clarification and/or rationale will be needed based on committee concerns. Developers can also include clarification in their 3–5-minute opening remarks during the Recommendation Group meeting.



Endorsement Decision:
Recommendation Group Endorsement Meeting

The Recommendation Group of each E&M committee meets to discuss the measures and aggregated feedback and render an endorsement decision via a vote.

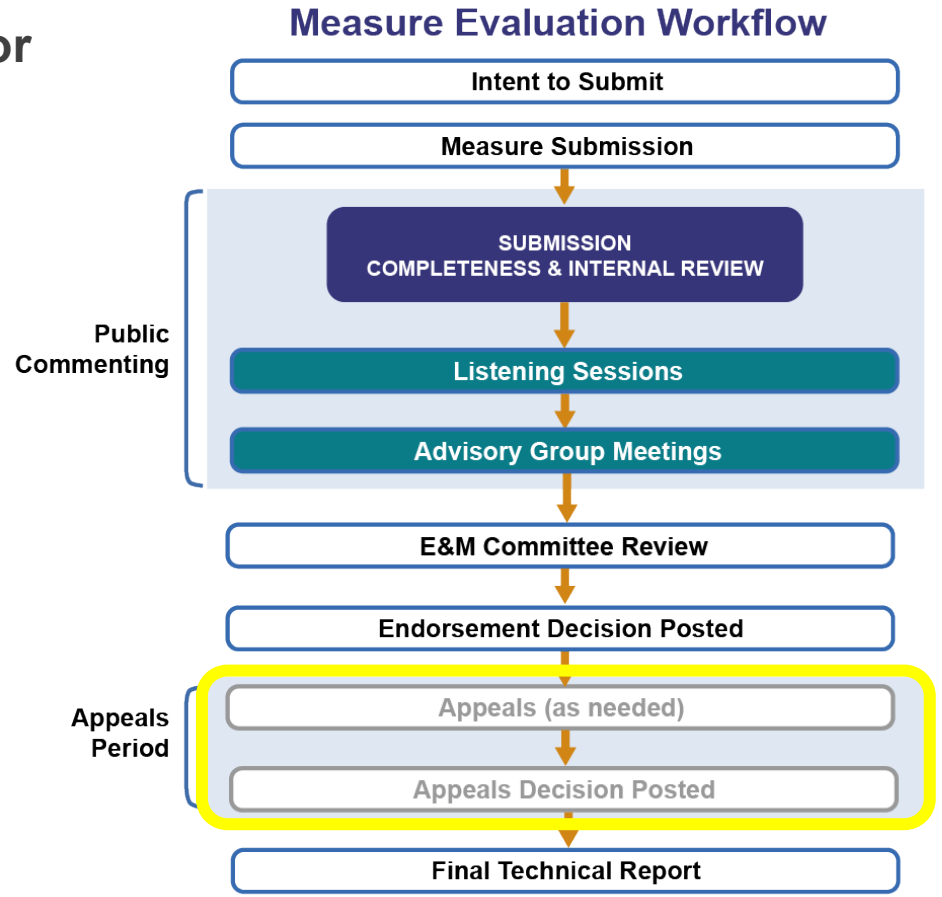
Engagement Opportunity	*Battelle Action*	*Developer Action*
Participate in Recommendation Group meeting	Battelle sends meeting invite and links to posted materials. Battelle convenes the Recommendation Group to discuss and share perspectives about the measures under review and render an endorsement decision via a vote.	Developers/stewards are invited and encouraged to attend, as they can give a 3–5-minute overview of their measure and respond to the Recommendation Group discussion during the meeting. Developers/stewards are encouraged to invite their SMEs to participate to support answering questions.



Appeals Period

Any interested party can submit an appeal request for any E&M committee endorsement decision. Battelle staff review appeals for eligibility. If an eligible appeal is received, the Ad Hoc Appeals Committee will meet to review the appeal and decide whether to uphold (i.e., overturn the endorsement decision) or deny it (i.e., uphold the endorsement decision).

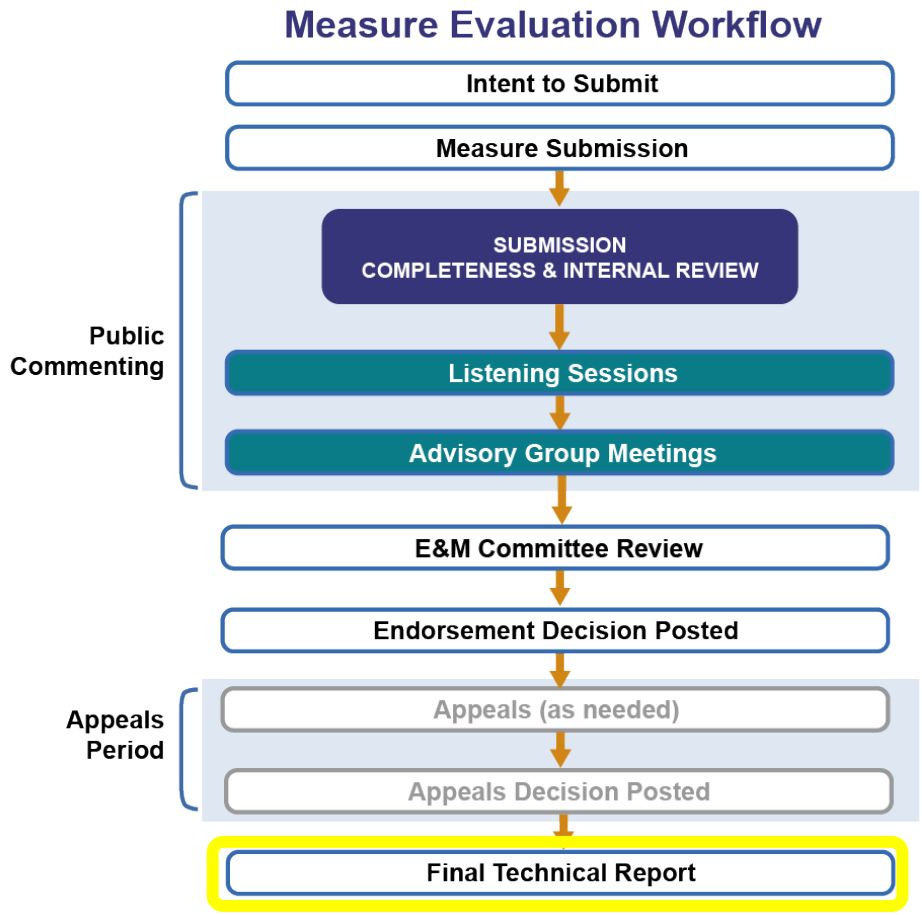
Engagement Opportunity	*Battelle Action*	*Developer Action*
Review any appeals received for your measure	Battelle notifies developers/stewards if an appeal is received for their measure (regardless if it is an eligible appeal or not). If the appeal is eligible, the Ad Hoc Appeals Committee will meet, and their decision is published to the respective measure page in STAR.	Developers are invited to attend and answer questions, as needed, for their measure.



Final Endorsement Decision Posted

Final endorsement decisions of measures are published in STAR. Final technical reports are published in April/May (Fall) and October/November (Spring).

Engagement Opportunity	*Battelle Action*	*Developer Action*
Access final technical reports	Battelle notifies developers/ stewards of final endorsement decisions via email, and final technical reports are published on the PQM Website.	Developers can access reports to learn from measure discussions/endorsement decisions and learn of any major priority gaps and/or methodological issues discussed by the committee.



Tips for Navigating the E&M Cycle

Review Measures in STAR

- Reach out to PQMSupport@battelle.org with any questions about cycle assignments.

Communication

- Ensure emails from no-reply@p4qm.org, PQM@battelle.org and PQMSupport@battelle.org are not blocked by your email provider.
  - no-reply@p4qm.org sends automated emails during ITS and FMS completeness checks.
  - PQM@battelle.org is used to by PQM staff to send calendar holds, meeting invites, cycle assignments, preliminary assessments, and updates on endorsement decisions and appeals.
  - PQMSupport@battelle.org is used to respond to any messages sent to the PQM staff.

Meeting Attendance

- Ensure dates and times for upcoming listening sessions and committee meetings are on your calendar.
- Forward or include any technical experts/SMEs who should attend from the developer's team.

Resources

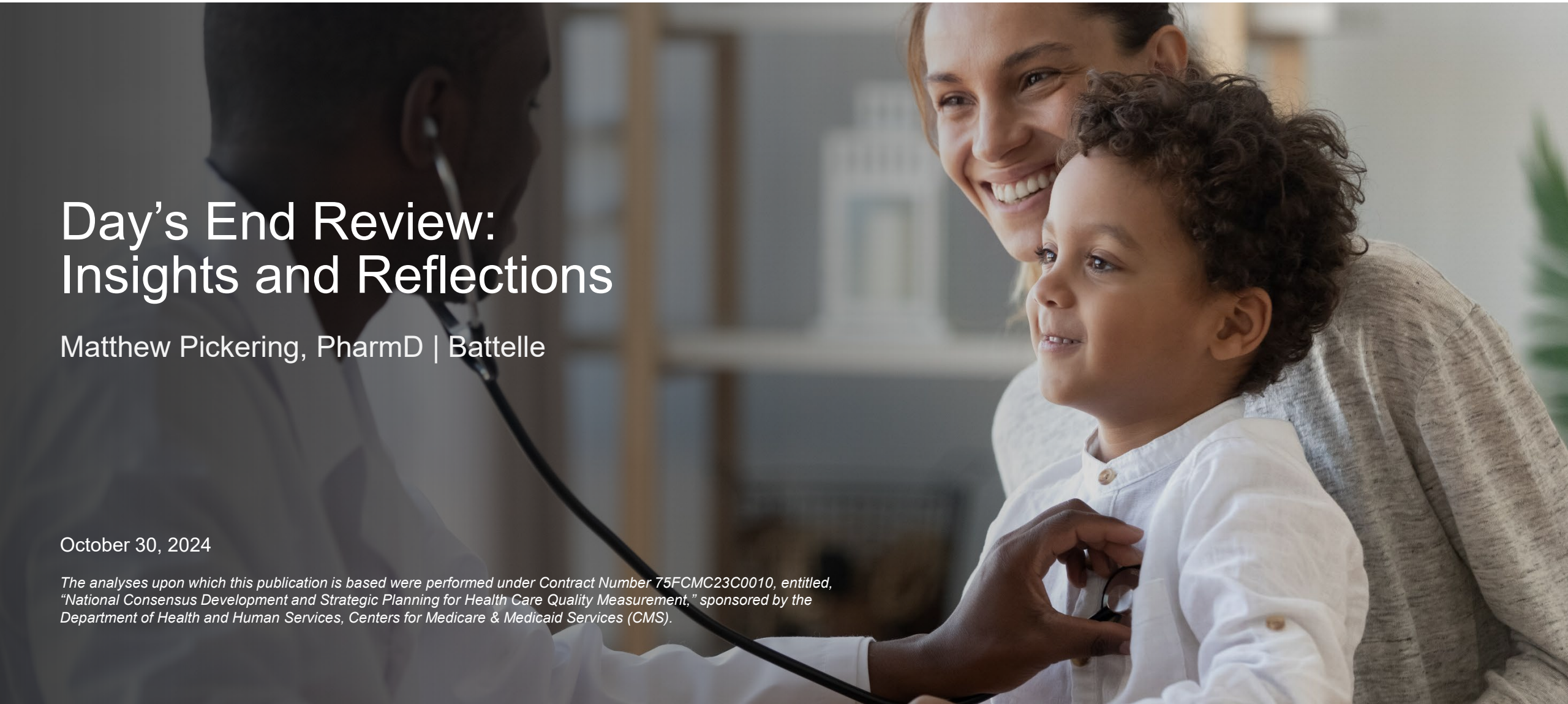
- **E&M Guidebook** – Provides information about the various steps of the Endorsement and Maintenance (E&M) process, including each phased review, possible endorsement decision outcomes, the appeals process, E&M policies and procedures, and the E&M committee structure.

- **Submission Tool and Repository Measure Database (STAR)** – An online platform where developers/stewards can submit measures, and any interested party may view measure information, including the endorsement status, in the searchable repository database. The database is updated regularly as new and maintenance measures are submitted to Battelle for PQM endorsement review.

Questions & Answers

Partnership for
Quality Measurement
Powered by **Battelle**

Partnership for Quality Measurement
Powered by Battelle



Day's End Review: Insights and Reflections

Matthew Pickering, PharmD | Battelle

October 30, 2024

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

Measure Developer Workshop

Objectives:

1. Gain insights into the key components of measure endorsement,
2. Learn about common pitfalls and challenges, and
3. Identify how to navigate the endorsement process effectively.

Workshop Agenda

Session Title	Purpose
1. Comprehensive Evaluation of Measures: A Holistic Approach	To review Battelle's **holistic** and enhanced **approach to measure evaluation** under the Partnership for Quality Measurement (PQM).
2. What Good Looks Like	To highlight **best practices for a comprehensive application submission**.
3. Importance and Logic Model Guidance	To share updated **guidance for developing a robust logic model** and review common pitfalls within the Importance domain.
4. Feasibility	To review common **pitfalls and mitigations** for the Feasibility domain and **expectations for new versus maintenance measures**.
5. Scientific Acceptability: Reliability, Validity, and Risk Adjustment	To discuss testing types based on data and measure score, explore common **pitfalls and mitigation strategies**, and relate these elements to the logic model to **enhance understanding and application of testing approaches**.
6. Equity Guidance	To introduce **Equity guidance** in preparation for the transition to a required domain in Spring 2025.
7. Instrument-Based Measure Guidance	To explain the policy on **instrument-based measures**, including considerations for submitting measures derived from instruments individually or as a composite, and operational details for completing the submission forms.
8. Use and Usability	To explain **expectations for new versus maintenance measures**, define "accountability application," and discuss **common pitfalls and mitigations**.
9. Developer Engagement: What to Expect During the Endorsement Cycle	To review the **steps within the endorsement cycle**, specify engagement opportunities for developers/stewards, and provide tips for effective navigation.

Reflection and Sharing

Key Takeaways

What are the key takeaways or insights you gained from today, especially those that will impact your approach to measure development and endorsement?

Session Topics

Were there any topics you expected to be covered that were not?

Navigating Endorsement

What is one piece of advice from today's session on navigating the endorsement process that you found particularly useful?

Challenges

What challenges do you anticipate in implementing today's learnings, and how might you address these challenges?

Helpful Resources

- **E&M Guidebook** – provides information about the various steps of the E&M process, including each phase of review, possible endorsement decision outcomes, the appeals process, E&M policies and procedures, and the E&M committee structure.
- **PQM Measure Evaluation Rubric*** – provides measure evaluation criteria for Fall 2024 cycle, as well as additional guidance for evaluating measures based on the criteria.
- **E&M Webpage** – contains additional information about E&M, including E&M project information, E&M committee meeting materials, and more.
- **Measure Management System (MMS) Hub/Blueprint** – provides a start-to-finish overview of quality measure development, implementation, and maintenance steps and processes.

* An updated PQM Evaluation Rubric effective for the Spring 2025 cycle can be found here.

Questions & Answers

Partnership for Quality Measurement
Powered by Battelle

Thank You!

Have questions? Contact us at
PQMsupport@battelle.org

Partnership for Quality Measurement

Powered by Battelle