Putting the Test in Measure Testing

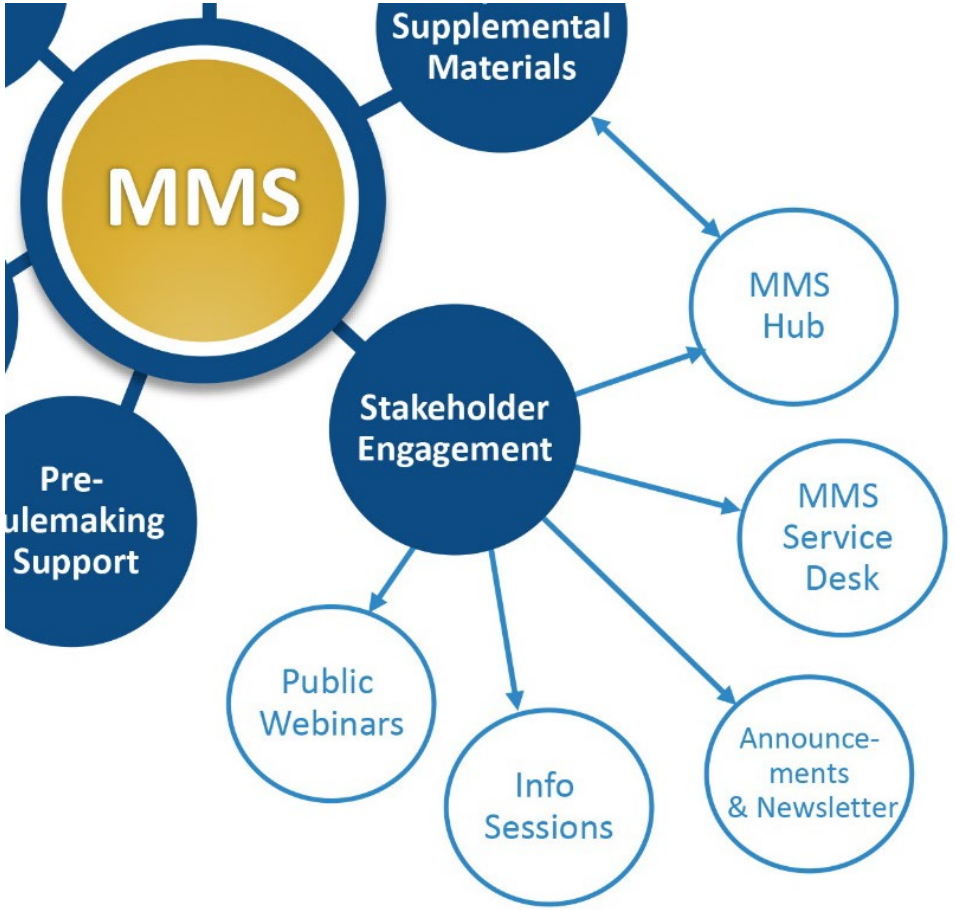
Addressing a Collection of Measure Testing Challenges

Presenters:
Meridith Eastman, Battelle
Mujahed Khan, Battelle

Welcome

The purpose of CMS's Measures Management System (MMS) Information Sessions are:

- To educate about quality measurement,
- To promote a standard approach to measure development and maintenance, and
- To encourage public involvement throughout the Measure Lifecycle.



Presentation Objectives

Discuss an assortment of measure testing challenges, brought to light through pre-rulemaking activities and other interactions with measure developers:

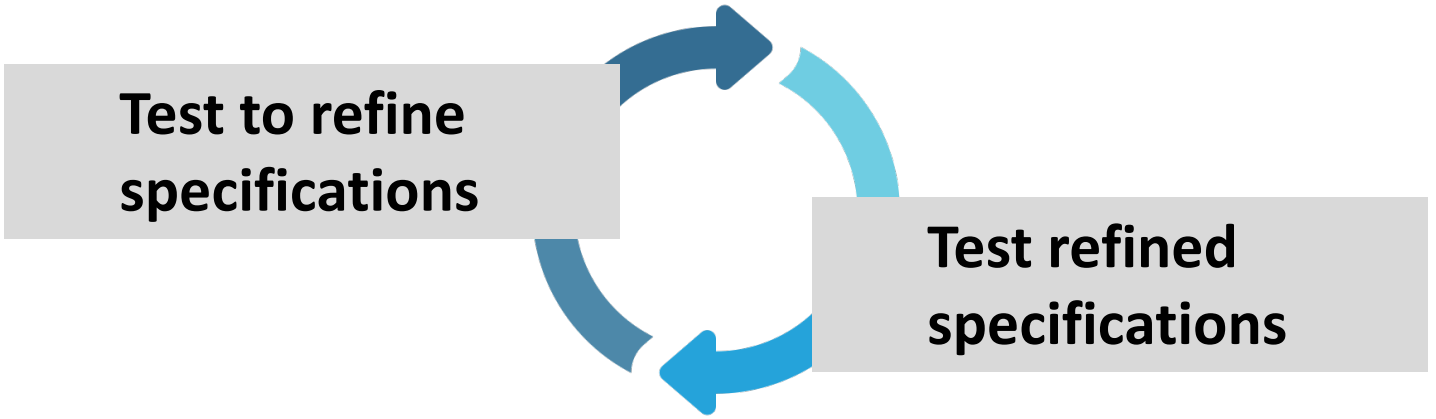
- Understand differences between alpha and beta testing
- Discuss tradeoffs between validity and feasibility
- Understand the difference between patient-/person-level data and patient-reported data
- Understand the benefits of referencing existing data element repositories
- Learn strategies for dealing with low-frequency (rare) events
- Discuss promising practices for collaborating with testing sites

Measure Testing

https://mmshub.cms.gov/measure-lifecycle/measure-testing/overview

Interactions with Specification

Although measure testing interacts with multiple stages of the measure lifecycle (e.g., conceptualization, implementation) measure testing often occurs iteratively with specification



Alpha vs. Beta Testing

What's in a Name?

Terms "alpha" and "beta" are borrowed from other fields to describe versions of a product (e.g., software) in development

- Other terms include "formative" and "field"

Useful in measure development because these terms provide an organizing framework for dividing the measure testing stage according to varying characteristics including:

- Scope, timing, and purpose
- Sampling considerations
- Measure evaluation criteria addressed

Scope, Timing, & Purpose

Alpha Testing	Beta Testing
Limited scope **Timing**: • Before detailed measure specifications are fully developed • During information gathering, empirical analysis • Iteratively **Purpose**: • Influence measure specification decisions	**Larger scope** **Timing**: • After the development of initial technical specifications **Purpose**: • To assess scientific acceptability and usability of a measure • To evaluate the measure's suitability for risk adjustment or stratification • To expand previous importance and feasibility evaluations

Sampling Considerations

Alpha Testing	Beta Testing
• Comprehensive enough that all elements needed for the measure are included in the data set • Enough data to reach “saturation” (universe of variation is identified) • Leverages convenience sampling (e.g., in house data)	• Statistically adequate sample size for scientific acceptability (i.e., reliability and validity) testing • Strives to be representative—may require inclusion of multiple sites with differing characteristics that matter for the measure • Uses statistically appropriate sampling techniques (e.g., a probability sampling technique such as simple random or cluster sampling)

Importance

Alpha Testing	Beta Testing
• Obtains a sense of the overall volume or frequency of the quality issue • Affirms that the measure can identify a gap in care • Provides support for further measure development	• Identifies performance thresholds, disparities, and variation in the outcome of interest • Identifies statistically significant variability among comparison groups (indicating an opportunity for improvement) • Ensures the measure is not “topped out”

Scientific Acceptability

Alpha Testing	Beta Testing
• Potentially assesses face validity	• Assesses empirical validity (including threats to validity such as the impact of missing data) and reliability • Tests any risk adjustment model

Feasibility

Alpha Testing	Beta Testing
• Assess if the required data can be collected and through what methods • Identifies implementation barriers • Estimates costs or burden of data collection and analysis	• Adds to evidence of feasibility obtained during alpha testing • Evaluates the feasibility of stratification (adequacy of cell size)

Usability and Use

Alpha Testing	Beta Testing
Potentially includes qualitative testing (e.g., focus groups, interviews) with: • Patients • Measured entities • Technical expert panel(s)	• Identifies unintended consequences, including susceptibility to inaccuracies and errors • Includes qualitative testing on refined specifications

Alpha Testing Activity Example

Measure developer drafts specifications based on results of environmental scan (e.g., literature and guidelines) and technical expert input then extracts data from the health system database to:

- Ensure that required elements are available in the format needed for measure calculation
  - **Example**: Can you have a patient refusal exclusion in an eCQM measure, if these data are only captured in unstructured notes?
- Assess frequency of exclusions and other data elements
  - **Example**: An exclusion that occurs with very low frequency may not be worth including

Beta Testing Activity Example

Work with multiple test sites (with different EHR systems) to extract data from EHRs, compare against manual abstraction, and calculate measure scores

- Calculate patient/encounter-level reliability (electronic vs. manual abstraction)
- Additional feasibility test for measure implementation
- Identify differences in scores between sites and within subgroups

Validity and Feasibility Tradeoffs

Validity and Feasibility Tradeoffs

- Decisions about validity and feasibility are often encountered during alpha testing, as specifications are drafted
  - **Validity**: Are you measuring what you want to measure?
  - **Feasibility**: Are data readily available? Can the measure be implemented?

- Many times, the “perfect” data element can not be feasibly and reliably obtained (e.g., in scanned documents or unstructured fields)

Weighing Tradeoffs

For example, selecting a data element that is more clinically appropriate, but more difficult to access vs. selecting a data element that is less clinically appropriate, but easier to access.

Tradeoff Decision Point

Aim to develop the most clinically valid measure possible, with the understanding that changes in code systems and technologies may very well increase feasibility over time.

- **Example**: As of this past year, LOINC® codes now distinguish between germline and somatic testing of the *BRCA1* and *BRCA2* genes

Be clear and transparent about how your measure might differ from the ideal.

Patient/Encounter Data vs. Patient-Reported Data

Terminology

"Patient/person- or encounter-level" testing incorporates what was formerly referred to as "data element-level" testing

- Refers to discrete information captured about a person or encounter (could include disease states, medical history, healthcare services performed)
  - Can be obtained from health records, claims, administrative data, etc.

Patient-/person- or encounter-level data ≠ patient-reported data

Importance

Completed patient-/person-/encounter-level testing of each critical data element, when appropriate, is one aspect of CMS's definition of a fully developed measure

- Could include comparing electronic vs. manual extract of the element (i.e., validity)

Note: This is not the same as testing a patient-reported data collection tool or survey.

For more information on fully developed measures, visit CMS MMS Hub:
https://mmshub.cms.gov/measure-lifecycle/measure-implementation/selection

Testing Data Elements

Are You Testing a New Data Element?

- The CMS Data Element Library (DEL) is the centralized resource for CMS assessment instrument data elements (e.g., questions and responses) and their associated health information technology (IT) standards
  - https://del.cms.gov/DELWeb/pubHome
- Electronic Clinical Quality Measure (eCQM) Data Element Repository (DERep)
  - https://ecqi.healthit.gov/mc-workspace-2/data-element-repository

Benefits of Referencing Existing Elements

- If other measures are using the same critical data elements, you may cite prior evidence to support validity and reliability assertions

- Checking the DEL and DERep also supports CMS's efforts to align and harmonize measures where possible

Low-Frequency (Rare) Events

Low-Frequency (Rare) Events

- Sometimes, the measure focus you want to assess occurs infrequently
- This creates challenges for achieving acceptable reliability for measured entities
  - Determining the best solution should take into account strengths and weaknesses of the approach as well as the ultimate intended use of the measure

Example
Wound infection from surgical lung biopsy

Possible Solutions

"Partial Pooling" or “Borrowing Strength"

- Pool data across longer time frames (e.g., lengthen the measurement period) to capture greater frequency
- Pool data of providers with similar structural characteristics
  - Data-driven weights are applied based on the amount of noise (i.e., measurement error) in the data

Possible Solutions

- Measure time between events rather than count of events (i.e., continuous variable), or create a measure without a denominator (e.g., number of infections per month)
- Broaden measure focus to increase frequency of measured event by developing a composite of similar events

Example
Instead of measuring wound infections, include any/all possible complications of lung biopsies

Collaborating with Testing Sites

Why Collaborate with Testing Sites?

Measure developers often engage with testing sites that can provide data for alpha and beta testing

Collaboration with sites can help ensure:

- Efficiency of testing process and adherence to timelines
- Data quality
- Insights about feasibility and implementation are shared
- Identification of best practices with testing
- Establishment of a network of experts to pool resources

Collaboration Strategies

Hold Kick-Off Meeting(s)

- Fully inform sites about the measure intent and ultimate implementation and use goals of the measure

Hold Training(s)

- Develop training materials (e.g., abstraction manuals, templates) to assist site staff in data collection

Hold Periodic Check-In Meetings

- Keep communication open to address data quality challenges or questions that may occur

Collaboration Strategies

Report Back

- Report results of interim and final analytic findings to the site, understand that they may want to use this information to learn about quality improvement issues at their own site

Stay in Touch

- Keep sites informed about the status of the measure and important milestones (e.g., Measures Under Consideration List submission or CBE review)

Collaboration Strategies

Create a Learning Collaborative

- Rather than engage multiple sites independently during the testing process, consider group meetings and engagement, so that sites with similar challenges and questions can learn from one another
- Identify a 'measure champion' at each site who will represent the site and work cross-functionally with staff
- Setup a repository of resources that sites can regularly pull from to address challenges they may face in the testing phases
- Consider publishing a regular newsletter or e-blast to inform about site-specific news, wins, and opportunities to network

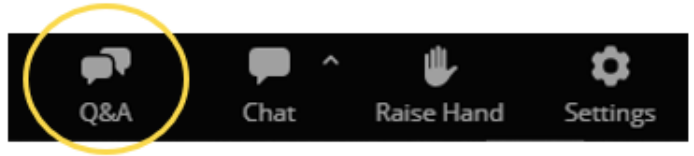
Collaboration Poll

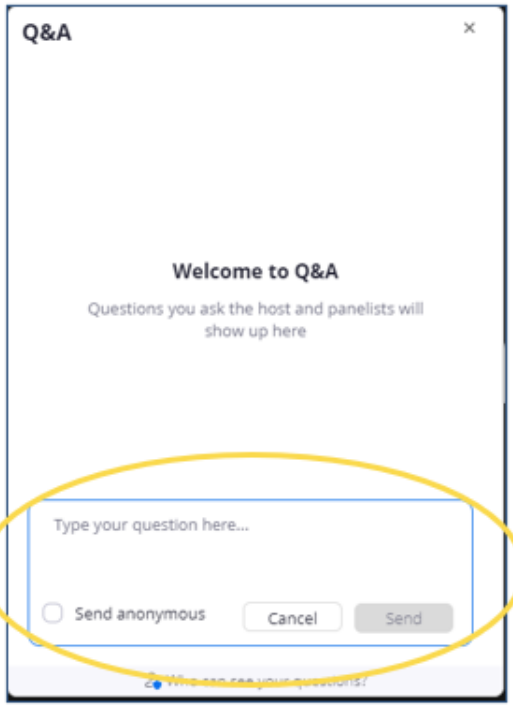
Tell us about your experience collaborating with testing sites!

Questions

Open the Zoom Q&A function

- Type your **question** into the question box
- Press **send** to submit





Resources and Links

- Measure Testing Lifecycle
  - **https://mmshub.cms.gov/measure-lifecycle/measure-testing/overview**
- CMS MMS Hub
  - **https://mmshub.cms.gov/measure-lifecycle/measure-testing/process/overview**
- For more information on fully developed measures
  - **https://mmshub.cms.gov/measure-lifecycle/measure-implementation/selection**
- The CMS Data Element Library (DEL)
  - **https://del.cms.gov/DELWeb/pubHome**
- Electronic Clinical Quality Measure (eCQM) Data Element Repository (DERep)
  - **https://ecqi.healthit.gov/mc-workspace-2/data-element-repository**
- Addressing Low Case-Volume in Healthcare Performance Measurement of Rural Providers
  - **https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=89672**

June Public Webinar

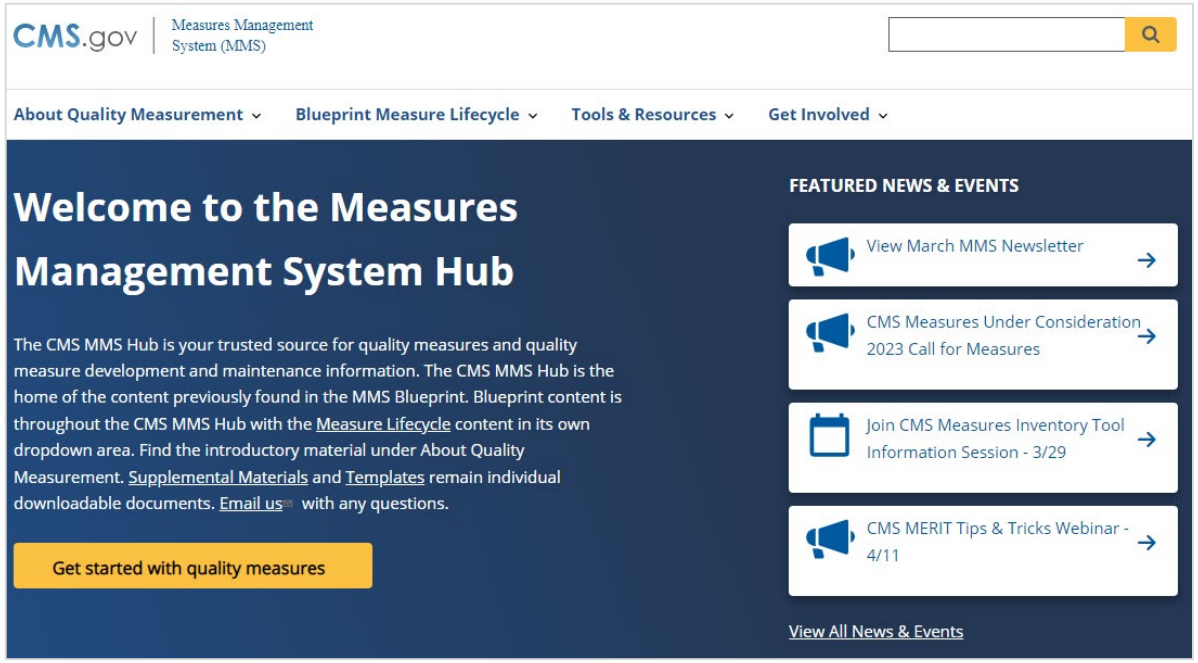
Register for June 22nd Session at 1 p.m. (ET): Register Here
Register for June 27th Session at 12 p.m. (ET): Register Here

For More Information

CMS MMS Hub

Visit **mmshub.cms.gov** for:

- Quality measurement resources,
- Latest MMS news and events
- Opportunities to get involved in quality measurement through public comments and participation in technical expert panels



CMS
Angela Wright
angela.wright@cms.hhs.gov

Gequincia Polk
gequincia.polk@cms.hhs.gov

MMS Team
Email: **MMSSupport@battelle.org**
CMS MMS Hub: **mmshub.cms.gov**