NATIONAL QUALITY FORUM

Scientific Methods Panel Orientation Meeting

Elisa Munthali, MPH
Helen Burstin, MD, MPH
Karen Johnson, MS
May Nacion, MPH

September 22, 2017

Welcome, Introductions/DOI, and Review of Meeting Objectives

Today's Speakers

Elisa Munthali
Acting Senior
Vice President

Helen Burstin
Chief
Scientific Officer

Karen Johnson
Senior
Director

May Nacion
Project
Manager

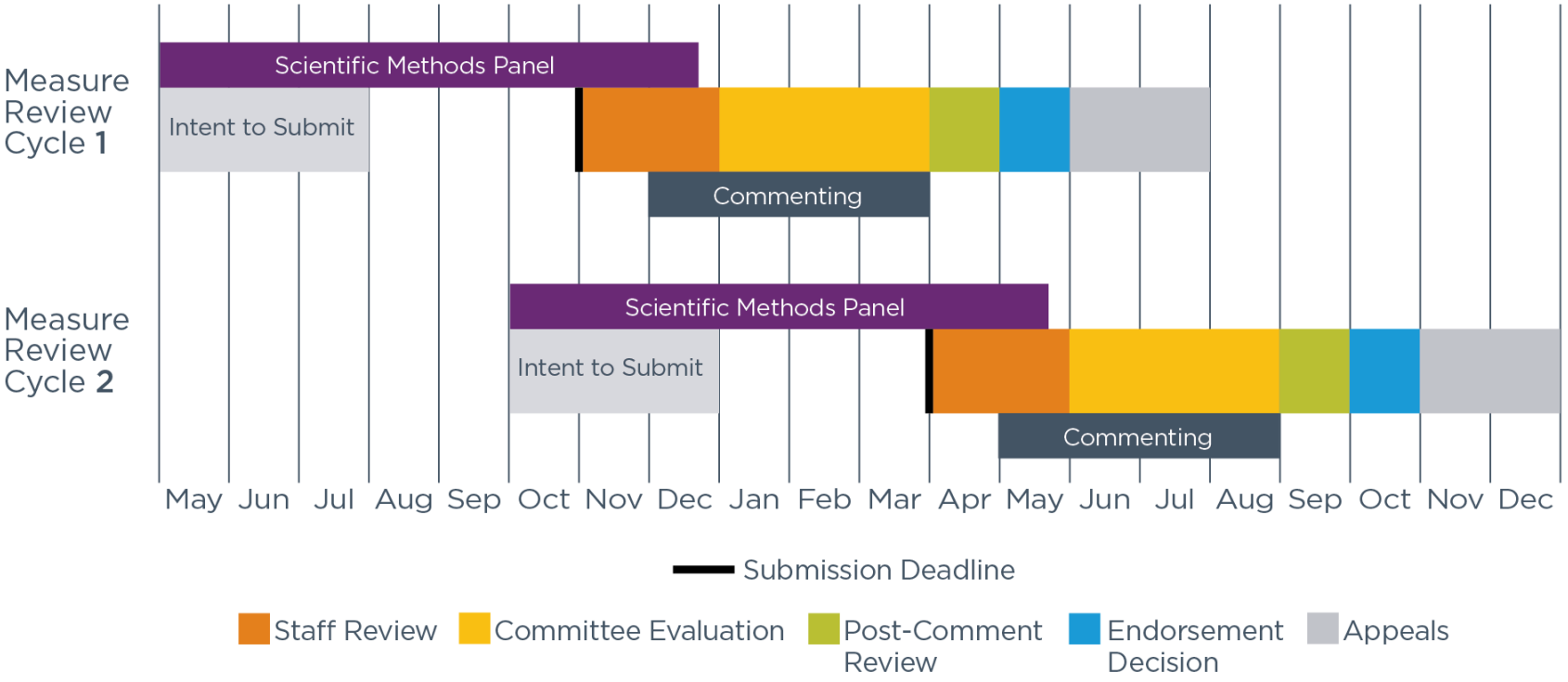
2017 CDP Redesign Changes

- Ongoing measure submission opportunities
- A newly formed NQF Scientific Methods Panel
- Expanded and continuous commenting period—with support/non-support
- Change in the content and structure of the measure evaluation technical report
- Final endorsement decision by the Standing Committee
- Shift in the role of the CSAC and the Appeals Board in the endorsement process
- Enhancements in stakeholder training and education
- Improvements in information exchange and access

NQF Consensus Development Process (CDP)

Consensus Development Process:

Two Cycles Every Contract Year



Scientific Methods Panel Members

J. Matt Austin, PhD	Joseph Kunisch, PhD, RN-BC, CPHQ
Bijan Borah, MSc, PhD	Paul Kurlansky, MD
John Bott, MBA, MSSW	Zhenqiu Lin, PhD
David Cella, PhD	Jack Needleman, PhD
Lacy Fabian, PhD	David Nerenz, PhD
Marybeth Farquhar, PhD, MSN, RN	Eugene Nuccio, PhD
Jeffrey Geppert, EdM, JD	Jennifer Perloff, PhD
Paul Gerrard, BS, MD	Sam Simon, PhD
Laurent Glance, MD	Michael Stoto, PhD
Stephen Horner, RN, BSN, MBA	Christie Teigland, PhD
Karen Joynt Maddox, MD, MPH	Ronald Walters, MD, MBA, MHA, MS
Sherrie Kaplan, PhD, MPH	Susan White, PhD, RHIA, CHDA

Meeting Objectives

- Review Scientific Methods Panel Charge
- Overview of NQF and the Consensus Development Process
- Overview of NQF's Measure Evaluation Criteria
- Overview of Potential Methodological Issues

Scientific Methods Panel Charge

Background

- Recommendation from our May 2017 Kaizen event to redesign our Consensus Development Process (CDP)
  - *Promote more consistent evaluations of Scientific Acceptability criterion*
  - *Reduce standing committee burden*
  - *Hopefully—promote greater participation of consumers, patients, and purchasers on NQF standing committees*

Methods Panel Charge

- Conduct evaluation of complex measures for the criterion of Scientific Acceptability, with a focus on reliability and validity analyses and results.
- Serve in an advisory capacity to NQF on methodologic issues, including those related to measure testing, risk adjustment, and measurement approaches.

Evaluation of the Scientific Acceptability Criterion

- Provide evaluation and ratings for reliability and validity subcriteria
  - *This information will help to inform the standing committee's endorsement decision*
  - *The Scientific Methods Panel will not render endorsement recommendations*
  - *Standing committees may raise concerns with the specifications of the measure or with potential threats to validity (e.g., selection of variables for risk adjustment model) and can overturn the Scientific Methods Panel ratings*

Workflow for Evaluations

- A minimum of three panel members will independently evaluate each measure
- The majority recommendation from the three evaluations will serve as the overall assessment of reliability and validity
- If there is substantial disagreement in the ratings between the three reviewers, the panel co-chairs will evaluate the measure and determine the overall recommendation
- NQF staff will compile the method's panel's ratings, evaluation, and commentary on reliability and validity and provide it to NQF's standing committees

A Few More Details…

- Complex measures
  - *Outcome measures, including intermediate clinical outcomes*
  - *Instrument-based measures (e.g., PRO-PMs)*
  - *Cost/resource use measures*
  - *Efficiency measures (those combining concepts of resource use and quality)*
  - *Composite measures*
- Workload ~15-20 measures per year (per panel member)
  - *We will try to match you based on expertise, availability, and need for recusal*

A Few More Details…

- Disclosure of Interest policy is the same as for standing committees
  - *Annual disclosure (general)*
  - *Measure-specific (as needed, but probably at least twice per year)*
- Terms
  - *Initial 2 or 3 year appointment (randomly assigned)*
  - *Optional 3-year follow-on appointment*

Advisory Function

- Advice on methodologic issues related to measure testing, risk adjustment, and measurement approaches
  - *Thresholds or rules of thumb for rating reliability and validity*
  - *Approaches to testing*
  - *Approaches for risk-adjustment*
  - *Testing requirements and ratings for reliability and validity*
- Recommendations are non-binding
  - *Changes to criteria/guidance subject to review and approval by the Consensus Standards Approval Committee (CSAC)*
- Advisory discussions will be the focus of monthly calls

Overview of NQF and the CDP

The National Quality Forum: A Unique Role

Established in 1999, NQF is a nonprofit, nonpartisan, membership-based organization that brings together public and private sector stakeholders to reach consensus on healthcare performance measurement. The goal is to make healthcare in the U.S. better, safer, and more affordable.

Mission: To lead national collaboration to improve health and healthcare quality through measurement

- An Essential Forum
- Gold Standard for Quality Measurement
- Leadership in Quality

NQF Activities in Multiple Measurement Areas

- **Performance Measure Endorsement**
  - *600+ NQF-endorsed measures across multiple clinical areas*
  - *19 empaneled standing committees*
- **Measure Applications Partnership (MAP)**
  - *Advises HHS on selecting measures for 20+ federal programs, Medicaid, and health exchanges*
- **National Quality Partners**
  - *Convenes stakeholders around critical health and healthcare topics*
  - *Spurs action on patient safety, early elective deliveries, and other issues*
- **Measurement Science**
  - *Convenes private and public sector leaders to reach consensus on complex issues in healthcare performance measurement such as attribution, alignment, sociodemographic status (SDS) adjustment*
- **Measure Incubator**
  - *Facilitates efficient measure development and testing through collaboration and partnership*

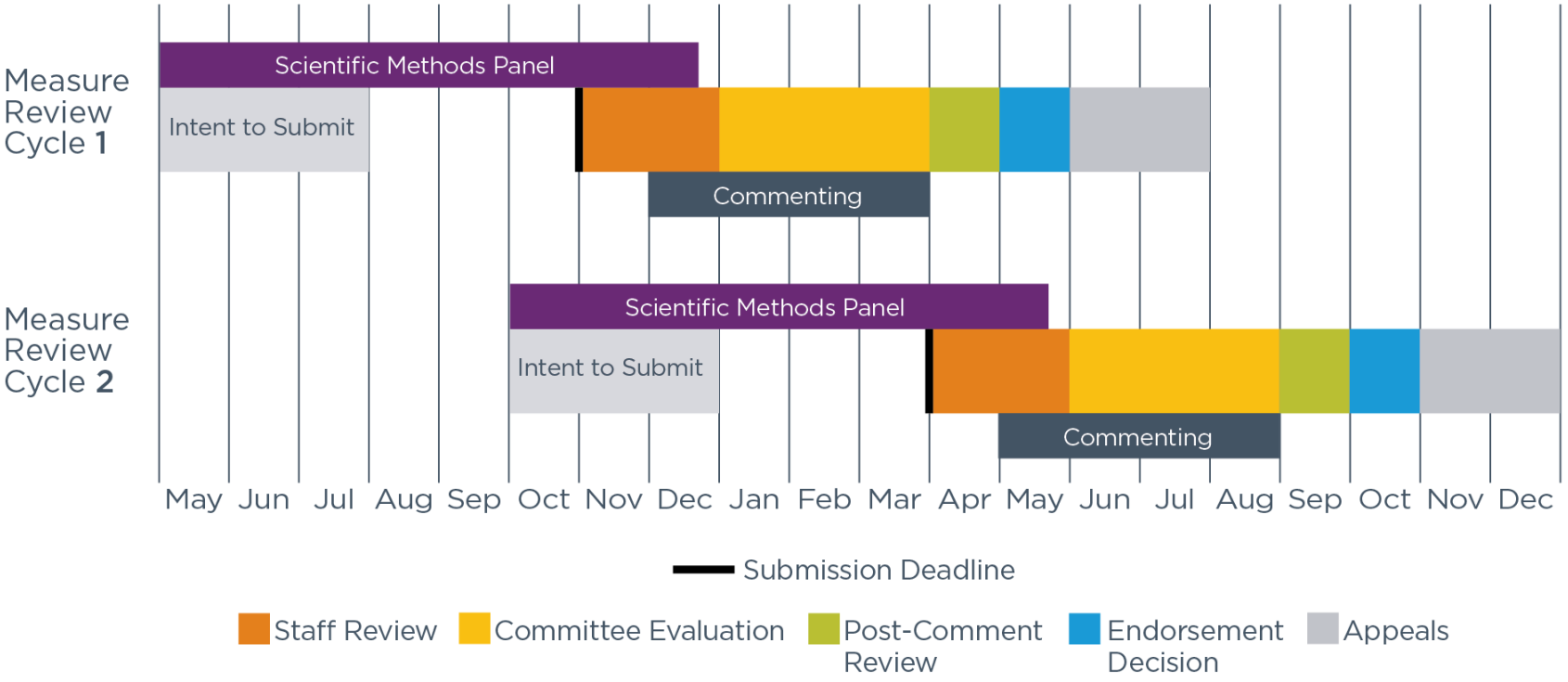
NQF Consensus Development Process (CDP)
6 Steps for Measure Endorsement

- Call for nominations for Standing Committee
- Call for candidate standards (measures)
- Candidate consensus standards review
- Public and member comment
- Consensus Standards Approval Committee (CSAC) ratification and endorsement
- Appeals

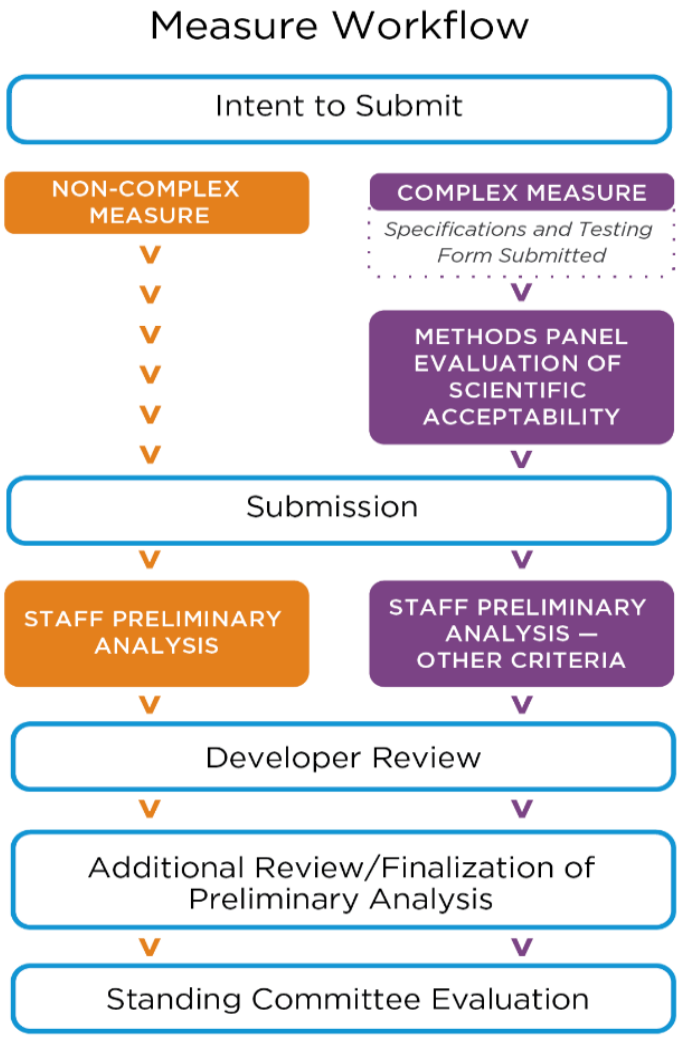
NQF Consensus Development Process (CDP)

Consensus Development Process:

Two Cycles Every Contract Year



Technical Review



A Few More Details…

- Developers provide submission information to NQF
- NQF staff check for completeness/responsiveness
- NQF staff prepare preliminary analysis for each measure
  - *This is where the Scientific Methods Panel comes in*
- Standing committee evaluation meeting
  - *Consider each measure, one criterion at a time*
  - *Voting on criteria/subcriteria*
    - » Process may differs a little if evaluating a maintenance measure
  - *Overall vote on suitability for endorsement*
- Public comment, followed by post-comment call
- CSAC decision
- Appeals

Questions?

Measure Evaluation Criteria Overview

NQF Measure Evaluation Criteria for Endorsement

NQF endorses measures for accountability applications (public reporting, payment programs, accreditation, etc.) as well as quality improvement.

- Standardized evaluation criteria
- Criteria have evolved over time in response to stakeholder feedback
- The quality measurement enterprise is constantly growing and evolving – greater experience, lessons learned, expanding demands for measures – the criteria evolve to reflect the ongoing needs of stakeholders

Major Endorsement Criteria
Hierarchy and Rationale (page 28)

- ***Importance to measure and report**: Goal is to measure those aspects with greatest potential of driving improvements; if not important, the other criteria are less meaningful (**must-pass**)*
- ***Reliability and Validity-scientific acceptability of measure properties** : Goal is to make valid conclusions about quality; if not reliable and valid, there is risk of improper interpretation (**must-pass**)*
- ***Feasibility**: Goal is to, ideally, cause as little burden as possible; if not feasible, consider alternative approaches*
- ***Usability and Use**: Goal is to use for decisions related to accountability and improvement; if not useful, probably do not care if feasible*
- ***Comparison to related or competing measures***

Criterion #1: Importance to Measure and Report (page 30-39)

1. Importance to measure and report - Extent to which the specific measure focus is evidence-based and important to making significant gains in healthcare quality where there is variation in or overall less-than-optimal performance.

1a. Evidence: *the measure focus is evidence-based (page 34-39)*

1b. Opportunity for Improvement: *demonstration of quality problems and opportunity for improvement, i.e., data demonstrating considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or disparities in care across population groups (pages 41-42)*

1c. Quality construct and rationale *(composite measures only)*

Criterion #2: Reliability and Validity – Scientific Acceptability of Measure Properties (page 39 -48)

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of health care delivery

2a. **Reliability (must-pass)**

2a1. Precise specifications including exclusions

2a2. Reliability testing—data elements or measure score

2b. **Validity (must-pass)**

2b1. Validity testing—data elements or measure score

2b2. Justification of exclusions—relates to evidence

2b3. Risk adjustment—typically for outcome/cost/resource use

2b4. Identification of differences in performance

2b5. Comparability of data sources/methods

2b6. Missing data

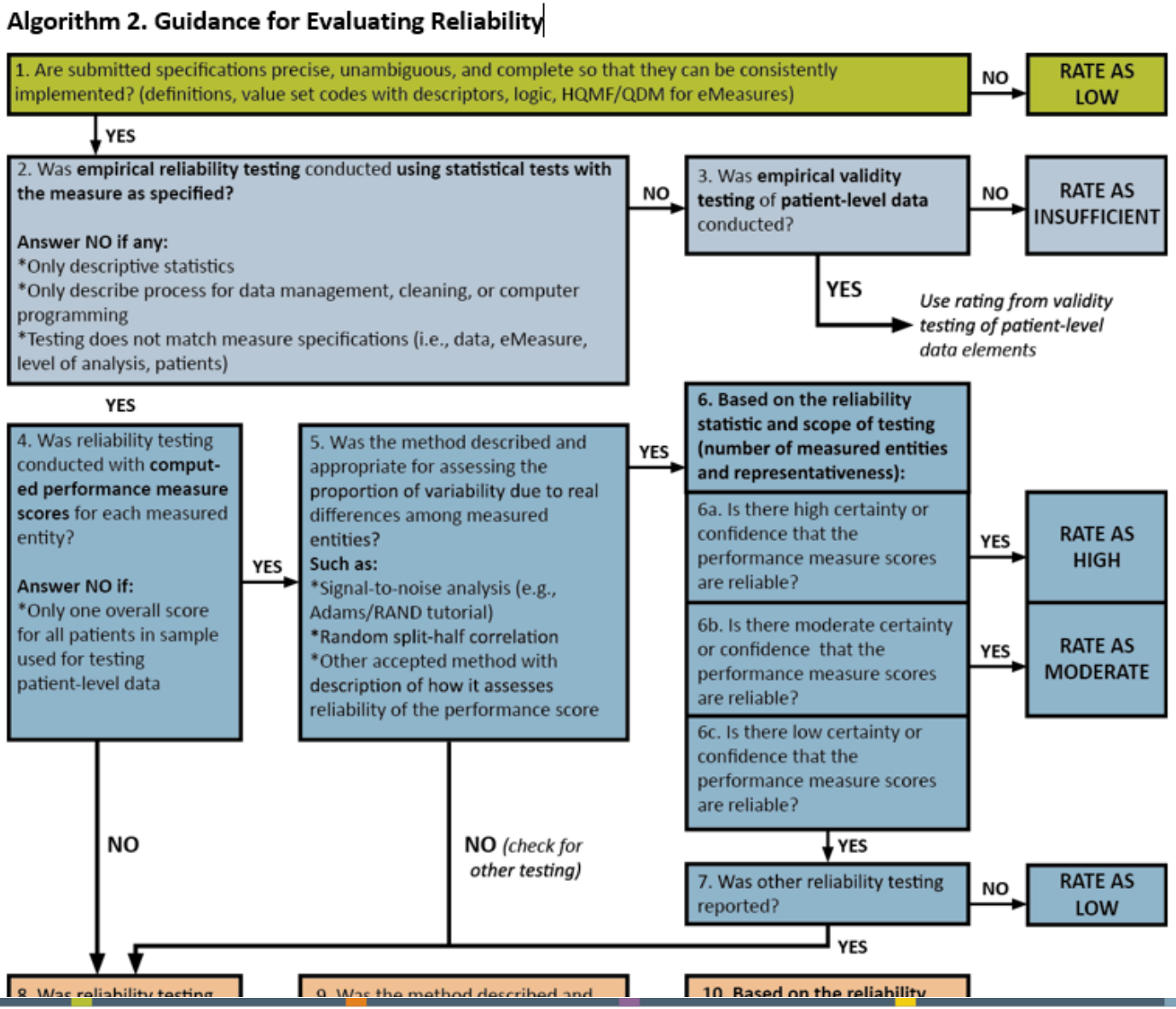
Evaluating Scientific Acceptability – Key Points (page 41)

Empirical analysis to demonstrate the reliability and validity of the *measure as specified,* including analysis of issues that pose threats to the validity of conclusions about quality of care such as exclusions, risk adjustment/stratification for outcome and resource use measures, methods to identify differences in performance, and comparability of data sources/methods.

Reliability Testing
Key points - page 42

- Reliability of the ***measure score*** refers to the proportion of variation in the performance scores due to systematic differences across the measured entities in relation to random variation or noise (i.e., the precision of the measure).
  - *Example - Statistical analysis of sources of variation in performance measure scores (signal-to-noise analysis)*
- Reliability of the ***data elements*** refers to the repeatability/reproducibility of the data and uses patient-level data
  - *Example – inter-rater reliability*
- Consider whether testing used an appropriate method and included adequate representation of providers and patients and whether results are within acceptable norms
- Algorithm #2 – page 43

Rating Reliability: Algorithm #2 – page 43

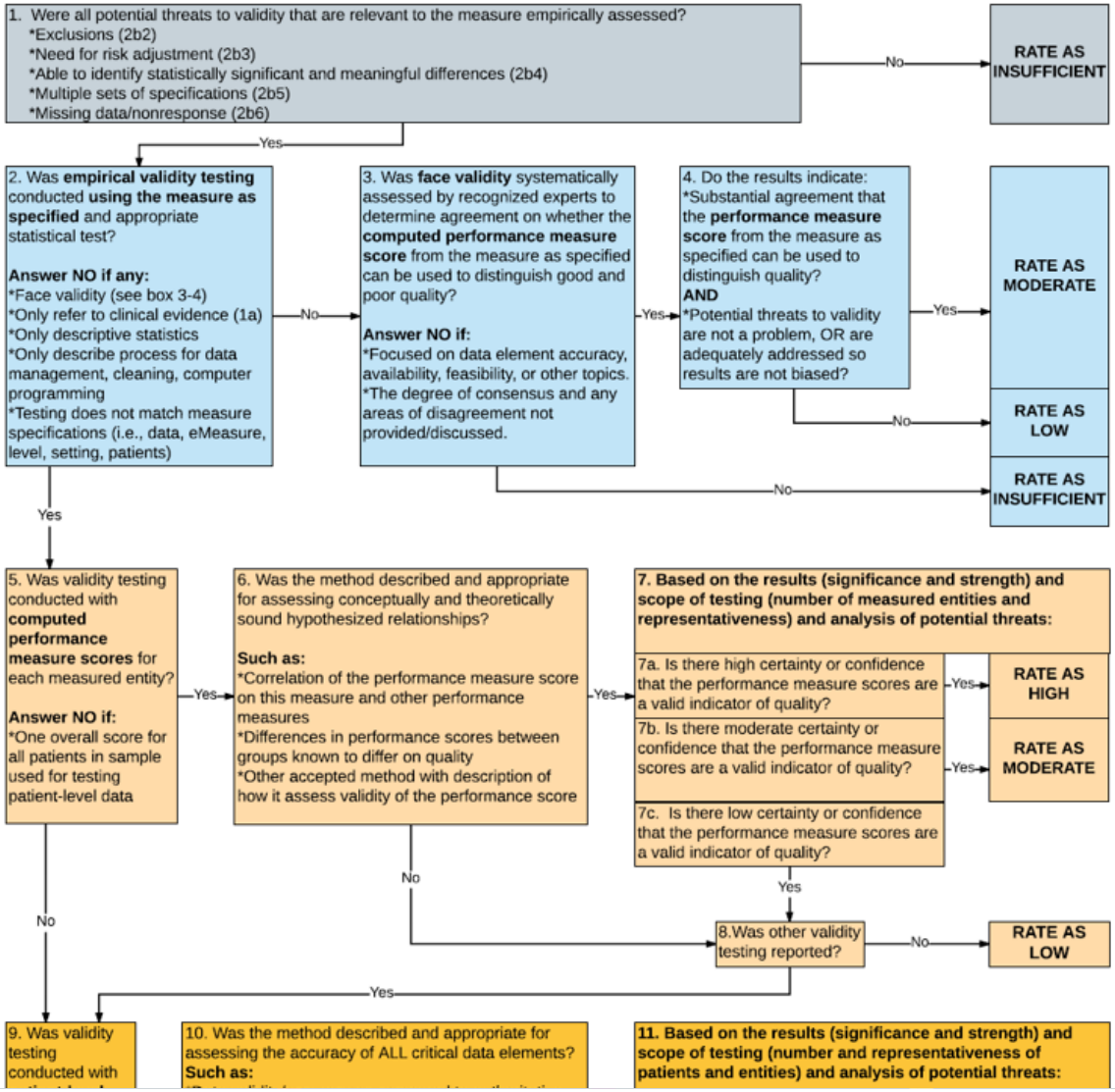


Validity testing (pages 44)
Key points – page 47

- **Empirical testing**
  - *Measure score* – assesses a hypothesized relationship of the measure results to some other concept; assesses the correctness of conclusions about quality
  - *Data element* – assesses the correctness of the data elements compared to a "gold standard"
- **Face validity**
  - *Subjective determination by experts that the measure appears to reflect quality of care*

Rating Validity: Algorithm #3 – page 50

Algorithm 3. Guidance for Evaluating Validity



Threats to Validity

- Conceptual
  - *Measure focus is not a relevant outcome of healthcare or not strongly linked to a relevant outcome*
- Unreliability
  - *Generally, an unreliable measure cannot be valid*
- Patients inappropriately excluded from measurement
- Differences in patient mix for outcome and resource use measures
- Measure scores that are generated with multiple data sources/methods
- Systematic missing or “incorrect” data (unintentional or intentional)

Criterion #3: Feasibility (page 49)
Key Points – page 50

Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement.

3a: Clinical data generated during care process

3b: Electronic sources

3c: Data collection strategy can be implemented

Criterion #4: Usability and Use (page 50) Key Points – page 51

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

Use (4a)

4a1: Accountability and Transparency: *Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement.*

4a2: Feedback by those being measured or others: *Those being measured have been given results and assistance in interpreting results; those being measured and others have been given opportunity for feedback; the feedback has been considered by developers.*

Usability (4b)

4b1: Improvement: *Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.*

4b2: Benefits outweigh the harms: *The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).*

Criterion #5: Related or Competing Measures (page 51-52)

If a measure meets the four criteria and there are endorsed/new related measures (same measure focus or same target population) or competing measures (both the same measure focus and same target population), the measures are compared to address harmonization and/or selection of the best measure.

- 5a. The measure specifications are harmonized with related measures **OR** the differences in specifications are justified.
- 5b. The measure is superior to competing measures (e.g., is a more valid or efficient way to measure) **OR** multiple measures are justified.

Questions?

Methodological Issues

A Sampling of Potential Topics

- Can we identify thresholds for reliability?
  - *If not actual thresholds, how about some rules of thumb?*
- Should we think any differently about reliability when measure focus is very rare (or very prevalent)?
- What are various methods for assessing reliability?
  - *When are the various methods appropriate?*
  - *Are there any pros/cons?*
- When/how should we think about reliability testing for eMeasures?
- What are the various methods for assessing validity?
- What are ways that can be used to support composite construction?

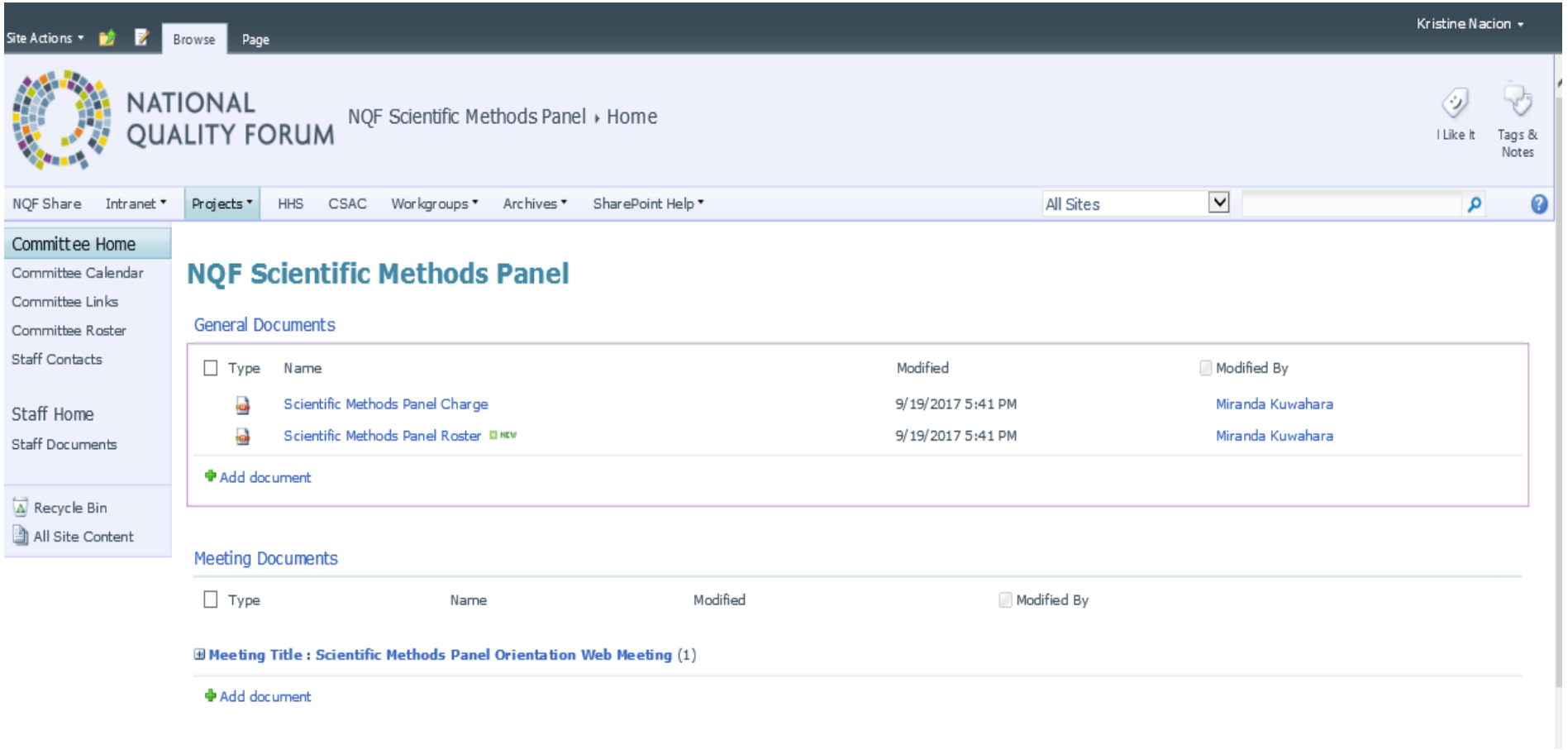
SharePoint Overview

SharePoint Overview

http://staff.qualityforum.org/Projects/NQF%20Scientific%20Methods%20Panel/SitePages/Home.aspx

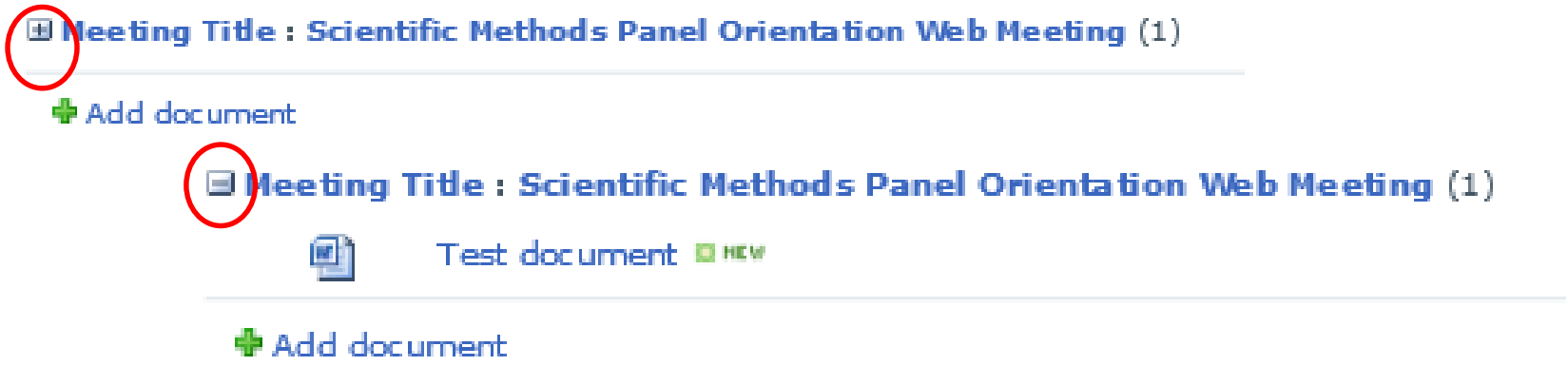
- Accessing SharePoint
- Scientific Methods Panel Charge
- Meeting and Call Documents
- Committee Roster and Biographies

Sharepoint Overview



SharePoint Overview

- **Please keep in mind:**
  - *(+) and (–) signs :*



Public Comment

Next Steps

Next Steps

- Monthly 1 hour Calls
  - *Doodle poll will be sent to determine group availability*
- Tracking Information
  - *SharePoint link will be sent to determine expertise, availability for the contract year, measure-specific DOI, etc.*
- Contact Information: methodspanel@qualityforum.org

Questions?


THANK YOU