NATIONAL QUALITY FORUM
Driving measurable health improvements together

http://www.qualityforum.org

Scientific Methods Panel Web Meeting

Sai Ma
Mike DiVecchia
Hannah Ingber
Caitlin Flouton

December 8, 2020

Funded by the Centers for Medicare and Medicaid Services under contract HHSM-500-2017-00060I –75FCMC19F0007.

Welcome

NQF Scientific Methods Panel Team

- Senior Lead
  - Sai Ma, PhD
- Project Management
  - Mike DiVecchia, MBA, PMP
  - Hannah Ingber, MPH
  - Caitlin Flouton, MS

Scientific Methods Panel Members

David Nerenz, PhD, Co-chair	Paul Kurlansky, MD
Christie Teigland, PhD, Co-Chair	Zhenqiu Lin, PhD
J. Matt Austin, PhD	Jack Needleman, PhD
Bijan Borah, MSc, PhD	Eugene Nuccio, PhD
John Bott, MBA, MSSW	Sean O'Brien, PhD
Daniel Deutscher, PT, PhD	Jennifer Perloff, PhD
Lacy Fabian, PhD	Patrick Romano, MD, MPH
Marybeth Farquhar, PhD, MSN, RN	Sam Simon, PhD
Jeffrey Geppert, EdM, JD	Alex Sox-Harris, PhD, MS
Laurent Glance, MD	Ronald Walters, MD, MBA, MHA, MS
Joseph Hyder, MD	Terri Warholak, PhD, RPh, CPHQ, FAPhA
Sherrie Kaplan, PhD, MPH	Eric Weinhandl, PhD, MS
Joseph Kunisch, PhD, RN-BC, CPHQ	Susan White, PhD, RHIA, CHDA

Meeting Overview

NATIONAL QUALITY FORUM

Meeting Objectives

- To clarify and improve the current NQF evaluation guidance for future measure evaluation cycles
  - Identify and attempt to resolve discrepancies in the current guidance regarding reliability and validity criteria
  - Clarify evaluation expectations for testing of various measure types
- To consider changes to Scientific Acceptability criteria evaluation

NATIONAL QUALITY FORUM

Meeting Agenda

- Debrief of the SMP evaluation review meeting in October
- Short-term action needed: improve the current evaluation guidance for measure developers
  - Expectations for validity testing and correlation analyses
  - Using formative and reflective models for composite measures
  - Testing requirements for instrument-based measures
- Longer-term action needed: consider changes for the future evaluation cycles
  - Discrepancies in evaluation policy and processes
  - Consideration of Use during Scientific Acceptability evaluation
  - Re-evaluating Landis & Koch
  - Clarifying expectations of risk adjustment in the SMP evaluations

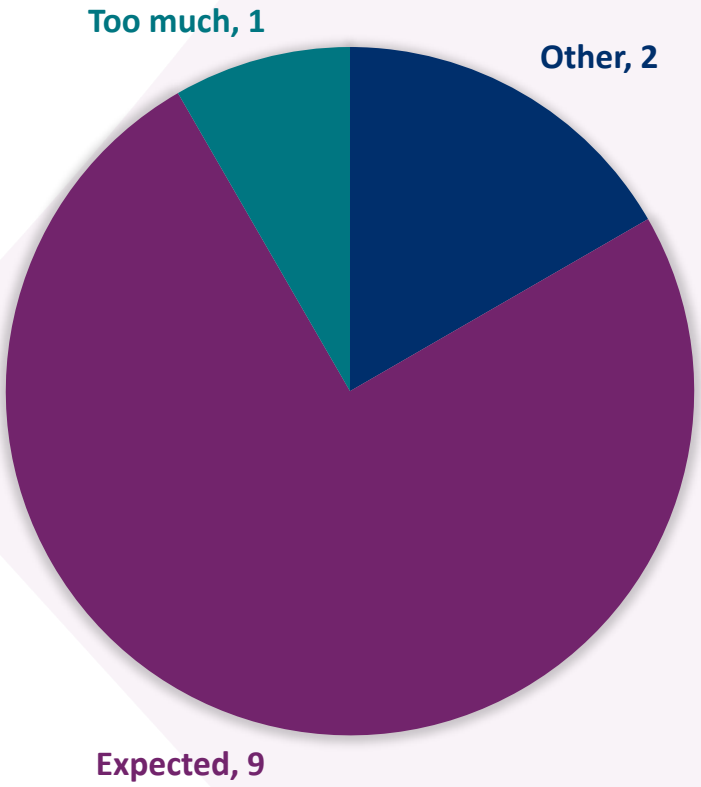
NATIONAL QUALITY FORUM

Process for Changing Evaluation Criteria

1. Obtain consensus from SMP
2. Gather input from stakeholders
3. Present recommendations to CSAC
4. CSAC votes on changes
5. If changes are enacted, incorporate into guidance and evaluation criteria
   - NOTE that NQF often allows up to a 1-year gap between changing criteria and implementing the changes
6. Disseminate changes, provide education resources, and clarify changes for measure developers

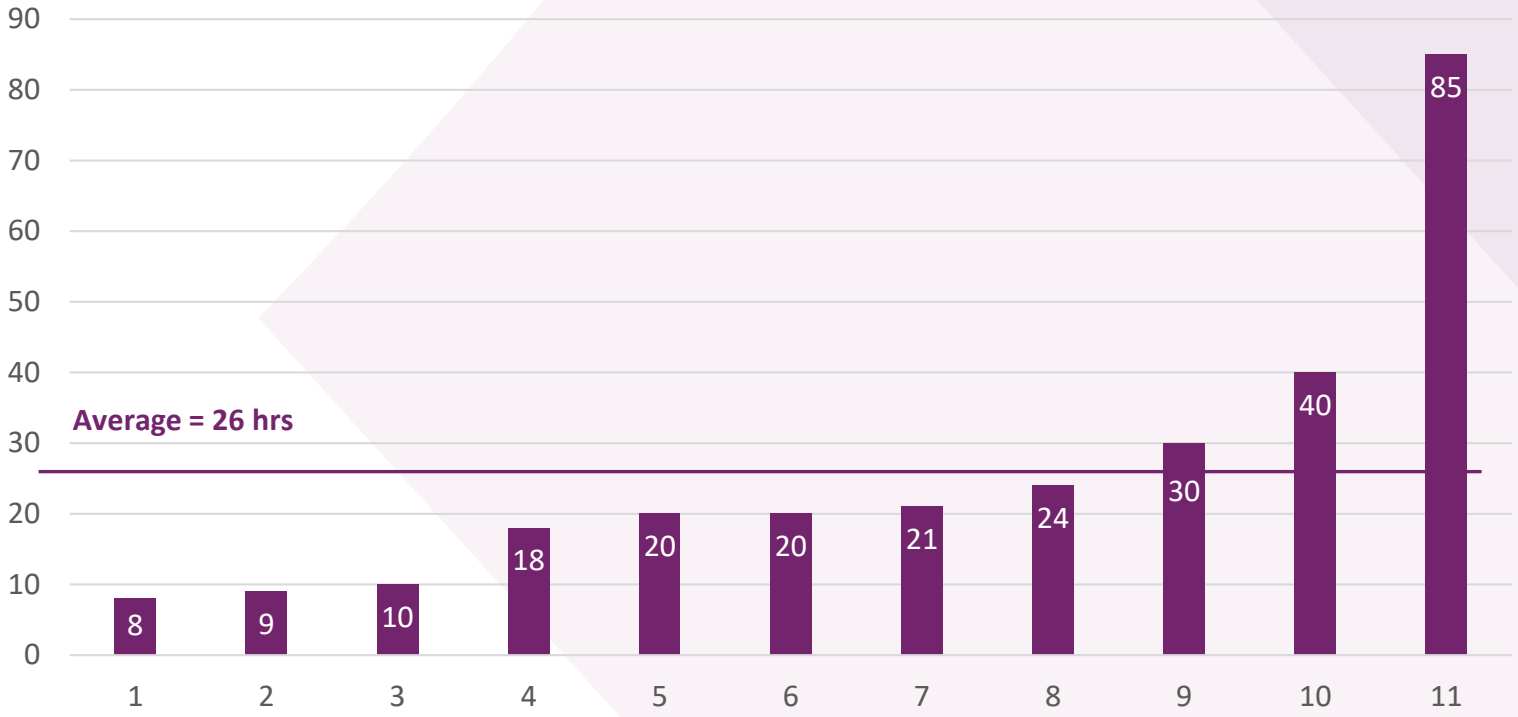
Debrief of SMP Measure Evaluation Meeting

SMP Workload for this Review Cycle

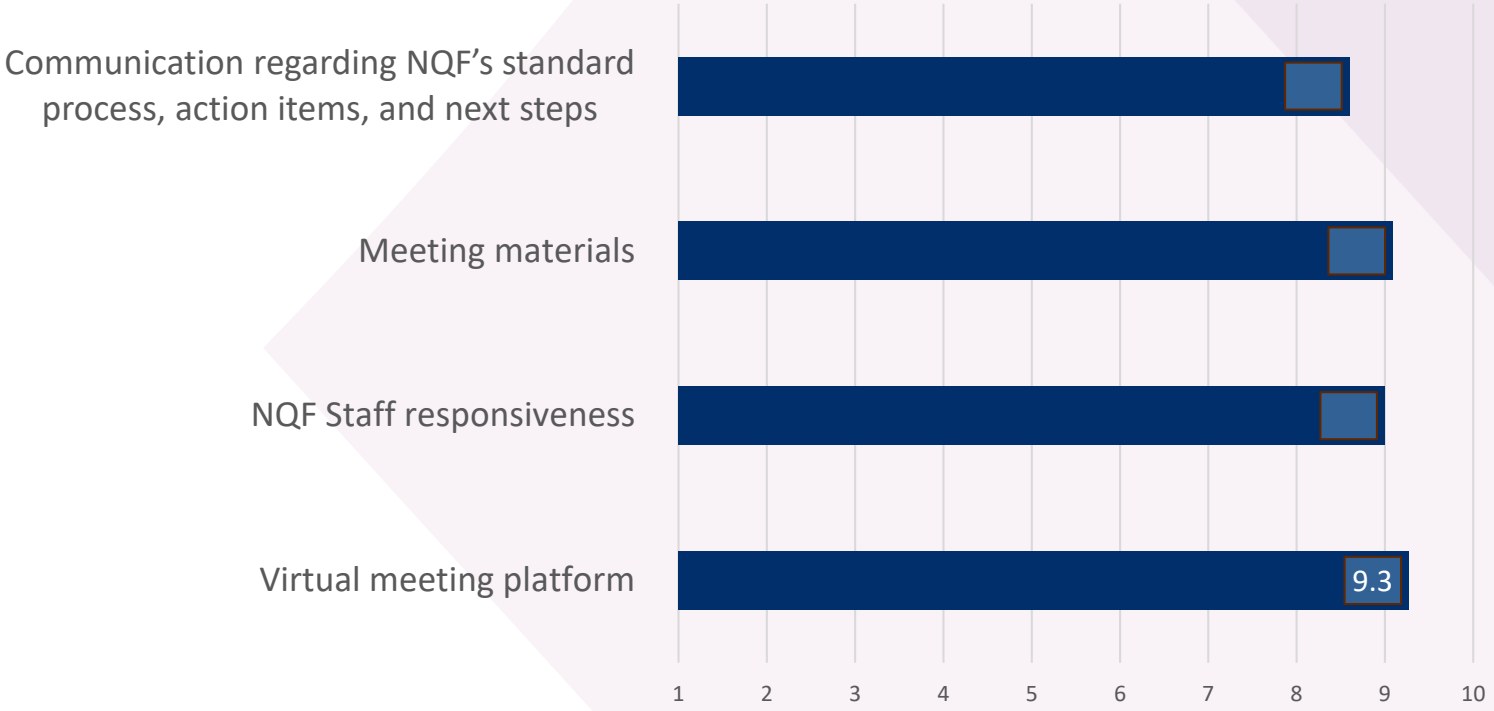


NATIONAL QUALITY FORUM

Approximately, how many hours did you spend reviewing all assigned measures, writing up the reviews, and preparing for and participating in the meetings?



Average Ratings of Meeting Logistics



Improving Guidance

Improving Guidance on Expectations for Validity Testing and Correlation Analyses

- Currently, the written NQF guidance on composite measures only states that the developer had to show a relationship between process, composite, and outcome, not a relationship in the "right" direction.
- Questions for the SMP:
  - Should the NQF guidance include language on expected directional relationship between a process/composite measure and outcomes?
  - Should each of the components show the same direction as the composite measure when correlating with an outcome?

NATIONAL QUALITY FORUM

Composite Measures

- Current NQF guidance does not require the developer to specify a reflective or formative model in the submission form
  - Reflective models represent the classical concept of measurement used in psychometrics: a single latent variable causing all the indicators
  - Formative models assume the indicators jointly determine the meaning of the construct.
  - Reference: Avila et al. *BMC Res Notes* (2015) 8:612
- Questions for the SMP:
  - Should we require a formative or reflective model in the submission?
  - How does this inclusion affect evaluation?

NATIONAL QUALITY FORUM

Further Clarifying Testing Requirements for Instrument-Based (Including PRO-PMs) Measures

- For reliability and validity, testing is required at **both** levels
  - *Data element level*: must demonstrate reliability and validity of the multi-item scales (e.g. at the patient level)
  - *Measure score level*: testing of the actual performance measure (e.g. at the practice level)
- Question for the SMP:
  - What language can be included in the NQF guidance to make the two required levels clearer?
  - Can we identify good examples from past submissions?

Considerations for Future Panel Evaluations

NATIONAL QUALITY FORUM

Discrepancies in Evaluation Policy and Processes

- **Differences in Testing Requirements by Measure Type**
- Health outcomes, intermediate clinical outcomes, cost/resource use, structure, process.
  - For both reliability and validity, NQF requires **EITHER** data element testing **OR** score-level testing. We prefer both, but currently do not require both
  - Impacts rating
  - Exception: face validity for new measures accepted
- Questions for the SMP:
  - For maintenance measures, should score-level testing always be required?
  - For maintenance measures, should empirical testing always be required?

NATIONAL QUALITY FORUM

Do Scientific Acceptance Criteria (Validity and Reliability) Vary by the Purpose of Measure Use?

- Proposed purpose categories:
  1. Identify outliers
  2. Put entities into one of two groups to either get or not get financial reward or punishment (e.g., lowest 25th percentile)
  3. Group entities into quintiles or similar groupings for "star ratings"
  4. Group entities into deciles for purposes of financial rewards or punishments
  5. Support consumer choice among similar entities in a geographic or market area (this would often involve making distinctions among "three-star" entities
  6. Use continuous scores for either consumer choice or financial incentive payments
- Questions for the SMP:
  - Should evaluation criteria vary according to the measure purpose categories listed above? If so, how?
  - Should measures be endorsed for only specific uses for which the testing was completed?

NATIONAL QUALITY FORUM

Adding Settings?

- Currently the submission form lists settings as individual clinician, group/practice (hospital/facility/agency), health plan, and others
- We have increasingly seen measures submitted for ACOs, National Provider Identifiers (NPIs), hospital unit/department, etc.
- Testing must align with specifications
  - Current language: "Testing must be conducted for the measure as specified (e.g., all relevant levels of analysis, using applicable data sources, care settings, patients, providers, etc.). If more than one measure is included under one NQF number, each measure must be tested per NQF evaluation requirements. **If more than one level of analysis is specified, testing must be conducted for each level separate" (p18)**
- Questions for the SMP:
  - Should other levels be added? (e.g. NPI can range from an individual doctor to an organization (group practice, lab, hospital, etc.))
  - What guidance can we offer?

NATIONAL QUALITY FORUM

Acceptable Thresholds for Reliability

- Differing threshold values exist within the literature (Landis, Adams, others)
- Questions for the SMP:
  - What are acceptable ways to look for alternatives to Landis? What conceptual approaches are acceptable without an appropriate threshold?
    - Past precedent and existing norms?
    - Arbitrary adjective descriptions?
    - Probability of misclassification?
  - Something else?
  - In lieu of the Landis & Koch paper, or Adams tutorial, what guidance and references can we provide to developers?
    - ICC : Koo & Li 2016
    - 0.5 > poor; 0.5-0.75 = moderate; 0.75-0.9 = good; 0.9 = excellent
    - Are the thresholds suggested by Koo & Li appropriate for data-element level reliability?
    - How would the evaluation ratings be assigned based on the threshold?

NATIONAL QUALITY FORUM

Including Risk Adjustment in the SMP Evaluations

- Current guidance:
  - 2b3. For **outcome measures and other measures when indicated (e.g., resource use)**: an evidence-based risk-adjustment strategy is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care, and has demonstrated adequate discrimination and calibration
  - **OR**
  - Rationale/Data support no risk adjustment. (See section on Risk Adjustment for Social Risk Factors)
  - Risk factors that influence outcomes should not be specified as exclusions.
  - A second Social Risk Trial began in 2017 and will run until 2021. Measure developers **are required** to provide a conceptual rationale for how a social risk factor affects an outcome of interest. If a conceptual relationship exists, developers **should conduct empirical analyses** to examine the relationship between the social risk factor and the outcome of interest.
- Question to the SMP:
  - Should this guidance apply to other measure types?
  - Should a measure be rated "insufficient" solely due to a lack of or inappropriate use of risk adjustment?

Opportunity for Public Comment

Next Steps

NATIONAL QUALITY FORUM

Next Steps

- Meeting summary review
- SMP Methods Meetings (every other month)
  - Poll for availability is forthcoming
  - 2-hour duration
  - Next call: February 2021 (specific date/time TBD)
- SMP evaluation meetings are likely to be in late March and October

THANK YOU.

NATIONAL QUALITY FORUM

http://www.qualityforum.org