



NATIONAL
QUALITY FORUM

Scientific Methods Panel Fall 2019 Measure Evaluation Meeting

October 28-29, 2019

Welcome, Introductions, and Disclosures of Interest

Agenda for Day 1

- Welcome, Introductions, and Disclosures of Interest
- Meeting Overview: Process and Criteria
- Measure Evaluations
 - ▣ *Subgroup 1*
 - ▣ *Subgroup 2*
 - ▣ *Subgroup 3*

NQF Scientific Methods Panel Team

- Content Leads
 - ▣ *Karen Johnson, MS*
 - ▣ *Ashlie Wilbon, MS, MPH, FNP-C*
 - ▣ *Sam Stolpe, PharmD, MPH*
 - ▣ *Michael Abrams, MPH, PhD*
 - ▣ *Andrew Lyzenga, MPP*
- Project Management
 - ▣ *Yetunde Ogunbemi, BS*

Scientific Methods Panel Members

J. Matt Austin, PhD	Jack Needleman, PhD
Bijan Borah, MSc, PhD	David Nerenz, PhD, Co-chair
John Bott, MBA, MSSW	Eugene Nuccio, PhD
David Cella, PhD, Co-chair	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	Michael Stoto, PhD
Joseph Hyder, MD*	Christie Teigland, PhD
Sherrie Kaplan, PhD, MPH	Ronald Walters, MD, MBA, MHA, MS
Joseph Kunisch, PhD, RN-BC, CPHQ	Terri Warholak, PhD, RPh, CPHQ, FAPhA*
Paul Kurlansky, MD	Eric Weinhandl, PhD, MS*
Zhenqiu Lin, PhD	Susan White, PhD, RHIA, CHDA

Meeting Overview: Process and Criteria

Background: Scientific Methods Panel

- Formed in 2017
 - ▣ *Direct outcome of May 2017 Kaizen event to redesign NQF's Consensus Development Process (CDP)*
- Charge of the panel
 - ▣ *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*

SMP Measure Evaluation Process

- Developers submit measure specifications and testing attachment
- NQF staff determine which measures are sent to SMP
- NQF team reviews measures for the following:
 - ▣ *Testing is performed at requisite levels (data element and/or measure score)*
 - ▣ *Testing is aligned with specifications*
 - ▣ *Administrative claims measures are specified and/or tested using ICD-10 codes*
 - ▣ *All required submission form items have a response*
 - ▣ *Maintenance measures submitted with a rationale when providing face validity only*

SMP Measure Evaluation Process

- SMP completes measure-specific disclosures of interest
- NQF staff assigns measures to SMP subgroups
 - ▣ *Based on number of measures, number of panelists, expertise, need for recusal, and previous experience on panel*
- Panel given four weeks to complete initial evaluations
 - ▣ *Apply the NQF Scientific Acceptability criteria for reliability and validity via a standardized form (“preliminary analysis” form)*
- NQF staff collates initial evaluations and ratings, and provided to developers
 - ▣ *Developers provided additional information to SMP, based on their initial evaluations, as desired*

SMP Measure Evaluation Process

- SMP members given opportunity to pull measures for discussion
- NQF staff develops discussion guide and determines meeting agenda
 - ▣ *Based on initial ratings, provision of additional information, and whether pulled by SMP members or staff*
- NQF staff make meeting materials available to the public
 - ▣ *Specifications, testing attachment, SMP preliminary analysis, discussion guide, agenda, slides*
- SMP meets in-person to discuss and vote on measures
- NQF staff provide summaries of the SMP discussion and votes to standing committees

Fall 2019 Evaluation Cycle Statistics

- A total of 55 measures submitted
 - ▣ *Of these, 22 were evaluated by the SMP*
 - » 10 new, remainder are maintenance measures being considered for continued endorsement
- Type
 - ▣ *Health outcomes: 10*
 - ▣ *Intermediate clinical outcomes: 6*
 - ▣ *composite measures: 3*
 - ▣ *PRO-PMs: 3*
- Initial SMP results
 - ▣ *Passed both reliability and validity: 13*
 - ▣ *Did not pass either reliability or validity: 3*
 - ▣ *Consensus not reached (CNR) on reliability and/or validity: 6*

Measure Discussion Process

- Staff will introduce the measure
- Lead discussant will summarize key concerns
- Other subgroup members are invited to comment
- Developers given 2-3 minutes for an initial response
- Discussion opened to full panel
 - ▣ *A few people are recused: they cannot discuss or vote*
 - ▣ *Developers can respond to questions from panelists*
- Subgroup members will cast final vote; other panelists will cast “shadow vote”

Meeting Materials

- Annotated agenda (provided to SMP members)
 - ▣ *Identifies subgroup members, lead discussants, and those recused for specific measures*

- Discussion Guide
 - ▣ *Includes pertinent information from the submission*
 - » Goal is to minimize need for back-and-forth with submission materials and to guide discussion so that we address critical questions/concerns
 - ▣ *Measures are include in same order as the agenda*
 - » By subgroup, then by rating (CNR, non-passing, passed but pulled, passed but not pulled)
 - ▣ *Appendix B: Additional information provided by developers*

The Voting Process

- Official subgroup votes
 - ▣ *Done via Poll Everywhere*
 - ▣ *Only members of the subgroup who provided a preliminary analysis is eligible to vote*
 - ▣ *Results from this vote will be the official vote of the SMP*
- Shadow voting: An information-gathering exercise
 - ▣ *Done via SurveyMonkey*
 - ▣ *Any panelist who did not provide a preliminary analysis is eligible to vote*
 - ▣ *Results from this vote will aid NQF staff in determining if changes to the voting process should be made in the future*

Overall Ratings

- High
 - ▣ *Score-level testing is required*
 - ▣ *A measure may be eligible for “HIGH” but the sampling method/results may make you choose “MODERATE” instead*
- Moderate
 - ▣ *The highest eligible rating if only data element testing or face validity testing is conducted*
 - ▣ *A measure may be eligible for “MODERATE” but the sampling method/results may make you choose “LOW” instead*
- Low
 - ▣ *Used primarily if testing results are not satisfactory*
- Insufficient
 - ▣ *Use when you don’t have sufficient information to assign a HIGH, MODERATE, or LOW rating*
 - » *Example: unclear specifications; unclear testing methodology*

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'd prefer both, but currently do not require both
 - » Impacts rating, as described above
 - » Exception: face validity for new measures accepted
 - ▣ *If data element validity testing provided, we **do not** require additional reliability testing*
 - » In this case, use the rating you give for validity as the rating for reliability
 - » This is not as common as it used to be

Differences in Testing Requirements by Measure Type

Composite measures

- NQF has specific definitions for “composite” measures
 - ▣ *“Traditional” composites*
 - ▣ *All-or-none measures*
 - ▣ *Does NOT include multi-item scales in surveys/questionnaires*
- Require reliability testing of the composite measure score
 - ▣ *Can also show reliability testing of the components, but this is not sufficient to pass the criterion*
- Score-level validity testing not required until maintenance
- Additional subcriterion: Empirical analyses to support the composite construction
 - ▣ *How this is addressed by the developer will depend on the type of composite*

Differences in Testing Requirements by Measure Type

Instrument-based measures

- For reliability and validity, require testing at **both** levels
 - ▣ *Data element level → must demonstrate R/V of the relevant items in the instrument*
 - ▣ *Measure score level → testing of the actual performance measure*
- We do allow multiple performance measures under same NQF number: need only one PA form, but may require multiple ratings

Differences in Testing Requirements by Measure Type

eCQMs (eMeasures):

- Testing from >1 EHR system required
 - ▣ *While more would be great, it is not required*
- Reliability testing not required if based on data from structured data fields; unstructured fields require both reliability and validity testing
- New (summer 2019) requirement: data element validation
 - ▣ *If data element testing is not possible, justification is required and must be accepted by the Standing Committee*

Some Additional Reminders...

- Testing must align with specifications
 - ▣ *Not a new requirement, but NQF is more rigorously upholding this requirement, particularly for level of analysis and minimum sample sizes*
 - » If multiple LoAs specified, each must be tested separately
 - ▣ *It is possible for you to “pass” part of the measure*
- Often there are several performance measures included under one NQF number
 - ▣ *Each must be evaluated separately; some might pass and others not pass*

Some Additional Reminders...

- For risk-adjusted measures
 - ▣ *Inclusion (or not) of certain factors in the risk-adjustment approach should not be a reason for rejecting a measure*
 - » Concerns with discrimination, calibration, or overall method of adjustment are still grounds for rejecting a measure
- For all measures
 - ▣ *Incomplete or ambiguous specifications are grounds for rejecting a measure—but remember that there is an option to get clarifications, although this must be done early on*
- Empirical validity testing is expected at time of maintenance evaluation
 - ▣ *If not possible, justification is required and must be accepted by the Standing Committee*

Some Additional Reminders...

- Recently, the SMP articulated additional **guidance** for submissions
 1. *Desire for more detail when describing methodology*
 2. *Desire for more than one overall statistic if reporting on signal-to-noise reliability*
 3. *Desire for detail in description of construct validation (i.e., narrative describing the hypothesized relationships; narrative describing why you think examining these relationships would validate the measure; expected direction of the association; expected strength of the association; specific statistical tests used; results; interpretation of those results (including how they related to hypothesis and whether they have helped to validate the measure))*
- Lack of #2 and #3 should not be grounds for rejecting a measure

Voting Test

Achieving Consensus

- Quorum: 66% of the Subgroup Pass/Recommended: Greater than 60% “Yes” votes (high + moderate ratings) of the quorum
- Consensus not reached (CNR): 40-60% “Yes” votes (inclusive of 40% and 60%) of the quorum
- Does not pass/Not Recommended: Less than 40% “Yes” votes of the quorum

Measure Evaluations: Subgroup 1

2456 Medication Reconciliation: Number of Unintentional Medication Discrepancies per Patient

- Preliminary Voting Result: Consensus Not Reached
 - ▣ *Reliability: H-0; M-4; L-2; I-0*
 - ▣ *Validity: H-0; M-3; L-2; I-1*
- Staff Lead: Andrew Lyzenga
- Lead Discussant: Jennifer Perloff
- Measure Developer: Brigham and Women's Hospital

1623 Bereaved Family Survey

- Preliminary Voting Result: Fail
 - ▣ *Reliability: H-3; M-2; I-0; I-1*
 - ▣ *Validity: H-1; M-1; L-3; I-1*
- Staff Lead: Karen Johnson
- Lead Discussant: Christie Teigland
- Measure Developer: Department of Veterans Affairs / Hospice and Palliative Care

Break

0575 Comprehensive Diabetes Care: Hemoglobin A1c (HbA1c) Control (<8.0%)

- Preliminary Voting Result: Pass
 - ▣ *Reliability: H-1; M-4; L-0; I-0*
 - ▣ *Validity: H-2; M-2; L-0; I-1*
- Staff Lead: Sam Stolpe
- Lead Discussant: Sherrie Kaplan
- Measure Developer: NCQA

0059 Comprehensive Diabetes Care: Hemoglobin A1c (HbA1c) Poor Control (>9.0%)

- Preliminary Voting Result: Pass
 - ▣ *Reliability: H-2; M-3; L-0; I-0*
 - ▣ *Validity: H-1; M-3; L-0; I-1*
- Staff Lead: Sam Stolpe
- Lead Discussant: Sherrie Kaplan
- Measure Developer: NCQA

0061 Comprehensive Diabetes Care: Blood Pressure Control (<140/90 mm Hg)

- Preliminary Voting Result: Pass
 - ▣ *Reliability: H-2; M-3; L-0; I-0*
 - ▣ *Validity: H-3; M-1; L-0; I-1*
- Staff Lead: Sam Stolpe
- Lead Discussant: Sherrie Kaplan
- Measure Developer: NCQA

Opportunity for Public Comment

Lunch Break

0425 Functional status change for patients with lumbar impairments

- Preliminary Voting Result: Pass
 - ▣ *Reliability: H-3; M-1; L-0; I-1*
 - ▣ *Validity: H-4; M-1; L-0; I-0*
- Staff Lead: Sam Stolpe
- Lead Discussant: Sherrie Kaplan
- Measure Developer: Focus on Therapeutic Outcomes, Inc.

Measure Evaluations: Subgroup 2

0696 STS Composite CABG Score

- Preliminary Voting Result: Consensus Not Reached
 - ▣ *Reliability: H-0; M-6; L-1, I-0*
 - ▣ *Validity: H-2; M-1; L-3; I-1*
 - ▣ *Composite: H-3; M-2; L-1; I-1*
- Staff Lead: Michael Abrams
- Lead Discussant: Jeff Geppert
- Measure Developer: The Society of Thoracic Surgeons (STS)

3537 Intraoperative Hypotension among Non-Emergent Noncardiac Surgical Cases

- Preliminary Voting Result: Consensus Not Reached
 - ▣ *Reliability: H-4; M-0; L-2; I-1*
 - ▣ *Validity: H-1; M-3; L-3; I-0*
- Staff Lead: Andrew Lyzenga
- Lead Discussant: Larry Glance
- Measure Developer: Mathematica

Break

0018 Controlling High Blood Pressure

- Preliminary Voting Result: Consensus Not Reached
 - ▣ *Reliability: H-4; M-1; L-0; I-2*
 - ▣ *Validity: H-0; M-4; L-2; I-1*
- Staff Lead: Ashlie Wilbon
- Lead Discussant: Daniel Deutscher
- Measure Developer: National Committee for Quality Assurance (NCQA)

3534 30 Day All-cause Risk Standardized Mortality Odds Ratio following Transcatheter Aortic Valve Replacement (TAVR)

- Preliminary Voting Result: Fail
 - ▣ *Reliability: H-0; M-2; L-3; I-2*
 - ▣ *Validity: H-0; M-4; L-2; I-1*
- Staff Lead: Karen Johnson
- Lead Discussant: Alex Sox-Harris
- Measure Developer: American College of Cardiology

Measure Evaluations: Subgroup 3

3478 Surgical Treatment Complications for Localized Prostate Cancer

- Preliminary Voting Result: Consensus Not Reached
 - ▣ *Reliability: H-0; M-3; L-1; I-1*
 - ▣ *Validity: H-0; M-3; L-1; I-1*
- Staff Lead: Ashlie Wilbon
- Lead Discussant: John Bott
- Measure Developer: Alliance of Dedicated Cancer Centers

Opportunity for Public Comment

Wrap-Up and Recap of Day 1

Adjourn for Day 1



NATIONAL
QUALITY FORUM

Scientific Methods Panel Fall 2019 Measure Evaluation Meeting

October 28-29, 2019

Welcome and Review of Objectives

Agenda for Day 2

- Welcome
- Measure Evaluations
 - ▣ *Subgroup 3 measures*
 - ▣ *Subgroup 4 measures*
- Process Review
- Methods Discussion
- Informational Update
- Next Steps
- Adjourn

Measure Evaluations: Subgroup 3 (continued)

3492 Acute Care Use Due to Opioid Overdose

- Preliminary Voting Result: Consensus Not Reached
 - ▣ *Reliability: H-4; M-0; L-1; I-2*
 - ▣ *Validity: H-1; M-3; L-0; I-3*
- Staff Lead: Michael Abrams
- Lead Discussant: Marybeth Farquhar
- Measure Developer: Yale CORE/CMS

Measure Evaluations: Subgroup 4

3528 CDC and VON Harmonized Outcome Measure for Late Onset Sepsis and Meningitis in Very Low Birthweight Neonates

- Preliminary Voting Result: Fail
 - ▣ *Reliability: H-1; M-1; L-3; I-1*
 - ▣ *Validity: H-0; M-4; L-2; I-0*
- Staff Lead: Michael Abrams
- Lead Discussant: Sam Simon
- Measure Developer: Centers for Disease Control and Prevention

3483 Adult Immunization Status

- Preliminary Voting Result: Pass
 - ▣ *Reliability: H-4; M-1; L-0; I-1*
 - ▣ *Validity: H-2; M-3; L-1; I-0*
 - ▣ *Composite: H-5; M-0; L-1; I-0*
- Staff Lead: Ashlie Wilbon
- Lead Discussant: J. Matt Austin
- Measure Developer: NCQA

3484 Prenatal Immunization Status

- Preliminary Voting Result: Pass
 - ▣ *Reliability: H-4; M-1; L-0; I-1*
 - ▣ *Validity: H-2; M-3; L-1; I-0*
 - ▣ *Composite: H-5; M-0; L-1; I-0*
- Staff Lead: Ashlie Wilbon
- Lead Discussant: J. Matt Austin
- Measure Developer: NCQA

Break

Process Review: Fall 2019 Cycle

Discussion Questions

- What is your opinion of our change to allow developers to provide additional information after your preliminary analysis?
- Was our process for allowing you to pull measures effective? If not, what should we change?
- Have you found the evaluation via in-person more effective than via webinars?
- What do you think about the “shadow voting” and the results thereof? Based on results for this cycle, would you recommend any changes in voting?
- How can we better support your evaluations?

Opportunity for Public Comment

Lunch Break

Methods Discussion

Discussion Topics

- Non-traditional “quality” measures
- Others from meeting “parking lot”

Informational Update: Other Measurement Science Efforts at NQF

Next Steps

Next Steps

- Measure submission deadline: Nov 1-15
- NQF staff will summarize the relevant measure information and discussions of the SMP, and provide to the various standing committees
 - ▣ *These committees will evaluate measures in the Jan-Feb timeframe*
 - ▣ *CSAC decisions expected in June 2020*
- Next Intent to Submit deadline: January 5, 2020

NATIONAL QUALITY FORUM ANNUAL CONFERENCE

20
20

Driving Value Through the Next Generation of Quality

March 23—25, 2020

OMNI SHOREHAM HOTEL, WASHINGTON, D.C.
[QUALITYFORUM.ORG](https://qualityforum.org)

#NQF20

[LEARN MORE ABOUT THIS PREMIER EVENT IN HEALTHCARE QUALITY](#)

Take a Peek at the Conversations Ahead

MAINSTAGE

- Approaches to Driving Improved Value in the Pharmaceutical Landscape
- The Future of Population Health: Addressing Challenges and Advancing Opportunities
- The ROI of 20 Years of Quality and the Road Ahead
- Healthcare Centers of Excellence: How Payers and Purchasers Define Success
- The Role of Healthcare Quality in Artificial Intelligence
- Hearing Directly from Patients and Consumers: Rating Systems and Activating Consumers

BREAKOUTS

- *Seeking Better Solutions for Marginalized Populations*
 - *How Quality is Responding to Public Health Crises*

Project Contact Info

- Email: MethodsPanel[@qualityforum.org](mailto:MethodsPanel@qualityforum.org)
- NQF phone: 202-783-1300
- Project page:
http://www.qualityforum.org/Project_Pages/Neurology.aspx
- SharePoint site:
<http://share.qualityforum.org/Projects/Neurology%202015/SitePages/Home.aspx>



Adjourn Day 2

Thank You