NATIONAL QUALITY FORUM
Driving measurable health improvements together

https://www.qualityforum.org

Scientific Methods Panel Advisory Web Meeting

June 14, 2022

Funded by the Centers for Medicare & Medicaid Services under contract HHSM-500-2017-00060I Task Order HHSM-500-T0001

Welcome

NATIONAL QUALITY FORUM
Driving measurable health improvements together

Housekeeping Reminders

- This is a Webex meeting with audio and video capabilities
- The system will allow you to mute/unmute yourself and turn your video on/off throughout the event
- Please mute yourself when not speaking
- We encourage you to keep the video on throughout the event
- We encourage you to use the following features
  - Chat box: to message NQF staff or the group
  - Raise hand: to be called upon to speak
- We will conduct roll call once the meeting begins

If you are experiencing technical issues, please contact the NQF project team at **methodspanel@qualityforum.org**

NATIONAL QUALITY FORUM
Driving measurable health improvements together

Meeting Ground Rules

- Be prepared, having reviewed the meeting materials beforehand
- Respect all voices
- Remain engaged and actively participate
- Keep your comments concise and focused
- Be respectful and allow others to contribute
- Share your experiences
- Learn from others

NATIONAL QUALITY FORUM
Driving measurable health improvements together

Agenda

- Roll call and review of meeting objectives
- Impact analysis - preliminary results
- Further guidance on divergent testing results at the patient/encounter and accountable entity levels
- Face validity testing requirements and acceptability
- Ratings and their use
- NQF member and public comment
- Next steps
- Adjourn

NATIONAL QUALITY FORUM
Driving measurable health improvements together

NQF Scientific Methods Panel Team

- Elizabeth Drye, MD, SM, Chief Scientific Officer
- Tricia Elliott, DHA, MBA, CPHQ, FNAHQ, Senior Managing Director
- Matthew Pickering, PharmD, Senior Director
- Poonam Bal, MHSA, Senior Director
- Mike DiVecchia, MBA, PMP, Director
- Hannah Ingber, MPH, Manager
- Gabby Kyle-Lion, MPH, Analyst

NATIONAL QUALITY FORUM
Driving measurable health improvements together

Scientific Methods Panel Members

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

NATIONAL QUALITY FORUM
Driving measurable health improvements together

Meeting Objectives

- Review and discuss preliminary results of the impact analysis
- Discuss and provide further guidance on divergent testing results for the patient/encounter and accountable entity levels
- Review and consider face validity testing requirements and its acceptability for maintenance endorsement
- Review and consider updates to the ratings assigned to the scientific acceptability criteria

Impact Analysis – Preliminary Results

NATIONAL QUALITY FORUM
Driving measurable health improvements together

Impact Analysis Overview

In December 2021, the SMP made three recommendations:

1. Establishing reliability testing thresholds, which measures must meet at initial and maintenance endorsement.
2. Requiring accountable entity level reliability testing at maintenance endorsement.
3. Requiring accountable entity level empirical validity testing at maintenance endorsement (i.e., not accepting face validity for maintenance review nor patient/encounter level empirical testing only).

NATIONAL QUALITY FORUM
Driving measurable health improvements together

Methods

- NQF staff conducted an analysis of all endorsed measures reviewed by the SMP to ascertain the proportion of measures that already:
  - meet these recommendations
  - *do not* meet these recommendations
- Staff reviewed testing attachments for relevant information to classify measures as meeting or not meeting the proposed recommendations

Category	Count
Total reviewed by SMP	163
Total available for impact analysis review	**138***

**Measures not yet endorsed (i.e., new measures in the spring 2022 and fall 2021 review cycle) were not included in the impact analysis. Additionally, documentation could not be found for three measures.*

- Staff also stratified the results by measures that are in use within Federal programs

NATIONAL QUALITY FORUM
Driving measurable health improvements together

SMP Recommended Thresholds for Person/Encounter Level Reliability Testing

Approach (Test)	Purpose	Range	Threshold
Internal consistency (e.g., Cronbach's Alpha)	The internally consistency of items in a multi-item scale.	0 to 1	0.7
Inter-rater agreement e.g., (Cohen's Kappa)	The inter-rater agreement of qualitative items correcting for chance.	−1 to +1	0.6
Test-Retest Reliability (Intraclass coefficient [ICC] or Pearson correlation)	Extent to which two measurements of the same concept at different times agree.	−1 to +1	0.5
Linear Relationships (e.g., Pearson correlation coefficient)	Agreement between two measures of the same concept.	−1 to +1	0.6

NATIONAL QUALITY FORUM
Driving measurable health improvements together

SMP Recommended Thresholds for Accountable Reporting Entity Level Reliability Testing

Approach (Test)	Testing Purpose	Range	Threshold
Signal to Noise Ratio (SNR) or Inter-Unit Reliability (IUR)	The precision attributed to an actual construct versus random variation.	0 to 1	0.6
Split-half reliability (Intraclass coefficient, with correction for full sample with Spearman-Brown formula)	Agreement between two measures of the same concept derived from split samples drawn from the same entity at a single point in time.	0 to 1	0.6

NATIONAL QUALITY FORUM
Driving measurable health improvements together

Preliminary Results of SMP-Reviewed Measures

Patient/Encounter Category	Count (n=138)
Patient/Encounter reliability testing was not done	81
Patient/Encounter reliability testing was done	57
Meets patient/encounter level threshold	27*
Does not meet patient/encounter level threshold	**7***

Accountable Entity (AE) Category	Count (n=138)
Accountable entity level reliability testing was not done	11
Accountable entity level reliability testing was done	127
Meets accountable entity level threshold	89†
Does not meet accountable entity level threshold	**13†**

Empirical Validity Testing Category	Count (n=138)
AE level empirical validity testing was done	117
AE level empirical validity testing was not done	**21**

*there were 23 measures for which some testing met the threshold and some did not (9) OR the threshold could not be confirmed (14).
†there were 25 measures for which some testing met the threshold and some did not (13) OR the threshold could not be confirmed (12).

NATIONAL QUALITY FORUM
Driving measurable health improvements together

SMP-Reviewed Measures That Did Not Meet Thresholds

Patient/Encounter Category	Count (n=7)
Does not meet patient/encounter level threshold	7
Outcome	7

Accountable Entity (AE) Category	Count (n=13)
Does not meet accountable entity level threshold	13
Outcome	10
Composite	2
Process	1

NATIONAL QUALITY FORUM
Driving measurable health improvements together

Preliminary Results for SMP-Reviewed Measures in Use

Patient/encounter category	Count (n=59)
Patient/Encounter reliability testing was not done	33
Patient/Encounter reliability testing was done	26
Meets patient/encounter level threshold	10*
Does not meet patient/encounter level threshold	**5***

Accountable entity category	Count (n=59)
Accountable entity level reliability testing was not done	4
Accountable entity level reliability testing was done	55
Meets accountable entity level threshold	34†
Does not meet accountable entity level threshold	**10†**

Empirical validity testing category	Count (n=59)
AE level empirical validity testing was done	50
AE level empirical validity testing was not done	**9**

*there were 11 measures for which some testing met the threshold and some did not (3) OR the threshold could not be confirmed (8).
†there were 11 measures for which some testing met the threshold and some did not (8) OR the threshold could not be confirmed (3).

NATIONAL QUALITY FORUM
Driving measurable health improvements together

Impact Analysis – Reliability Discussion Questions

- One developer reported profile inter-unit reliability (PIUR) in addition to an IUR.
  - ***How would we classify if this meets the threshold when one method is good (PIUR) and when one isn't as good (IUR) and vice versa?***
- There are cases where the overall/mean reliability is high, yet there are a non-trivial percentage of accountable entities where reliability is below the threshold.
  - ***How would the SMP envision applying thresholds to cases like this one?***
- There are some measures in which the reliability testing was deemed satisfactory, yet it does not fall clearly into the methods outlined in the thresholds table (e.g., PIUR).
  - ***How would the SMP envision applying thresholds to cases like this one?***

Guidance on Divergent Testing Results

NATIONAL QUALITY FORUM
Driving measurable health improvements together

Divergent Testing Results at the Patient/Encounter and Accountable Entity Levels

- It is up to the discretion of the NQF-convened body to determine if the testing results presented are satisfactory per NQF criteria.
- There are instances in which a developer submits multiple levels of testing (e.g., face validity and empirical validity; patient/encounter level and accountable entity level).
- NQF criteria does not provide prescriptive guidance for how NQF-convened bodies should consider multiple testing results that are divergent (i.e., when patient/encounter level is good, but accountable entity level is poor).
- The SMP, in the past, has expressed that additional guidance on these situations is needed.

Discussion:

- **What are suggestions for improvement?**
- **Do you have recommended changes to NQF criteria?**

Face Validity Testing Requirements and Acceptability at Maintenance Endorsement

NATIONAL QUALITY FORUM
Driving measurable health improvements together

Face Validity Requirements and Acceptability

- NQF's criteria allow a developer to submit face validity testing only for new measures.
  - Requirements: the process done must be systematic and transparent, by identified experts, that explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.
- For maintenance endorsement, empirical validity testing is expected at time of maintenance review; **if not possible, justification is required**.

Discussion:

- **What are the benefits and pain points with the current policy?**
- **What are suggestions for improvement to the policy?**
- **How can the NQF criteria be bolstered to provide improvements to this issue?**

Updates to Ratings Assigned to the Scientific Acceptability Criteria

NATIONAL QUALITY FORUM
Driving measurable health improvements together

Updates to Ratings Assigned to the Scientific Acceptability Criteria (continued)

- **Current State:** When reviewing measures' scientific acceptability, NQF's criteria use a four-part rating scale:
  - High
  - Moderate
  - Low
  - Insufficient

- **Potential Future State:** On a prior call, SMP members presented the following alternative options:
  - Pass
  - Does Not Pass
  - Insufficient

Discussion:

- **Would changing the current ratings structure better communicate results to measure developers?**
- **To Standing Committees?**
- **To the public?**

NQF Member and Public Comment

Next Steps

NATIONAL QUALITY FORUM
Driving measurable health improvements together

Next Steps and Reminders

- NQF staff creates a meeting summary of today's meeting
- Public comment on the reliability thresholds table
- Fall 2022 Intent to Submit (ITS) deadline is August 1, 2022
  - NQF SMP team will be sending out a doodle poll for the fall 2022 measure evaluation meeting and SMP meetings for the remainder of 2022

NATIONAL QUALITY FORUM
Driving measurable health improvements together

Potential Items for Future SMP Discussion

- Formative vs. reflective composite models
- Incorporating intended use into scientific acceptability discussions
- Appropriate testing sample size requirements

NATIONAL QUALITY FORUM
Driving measurable health improvements together

Project Contact Info

- Email: methodspanel@qualityforum.org
- NQF phone: 202-783-1300
- Project page: http://www.qualityforum.org/Measuring_Performance/Scientific_Methods_Panel.aspx
- SharePoint site: https://share.qualityforum.org/portfolio/ScientificMethodsPanel/SitePages/Home.aspx

THANK YOU.

NATIONAL QUALITY FORUM

https://www.qualityforum.org

Appendix

NATIONAL QUALITY FORUM
Driving measurable health improvements together

Impact Analysis Measures That Did Not Meet Recommended Patient/Encounter Level Thresholds

- 1716 National Healthcare Safety Network (NHSN) Facility-wide Inpatient Hospital-onset Methicillin-resistant Staphylococcus aureus (MRSA) Bacteremia Outcome Measure
- 1717 National Healthcare Safety Network (NHSN) Facility-wide Inpatient Hospital-onset Clostridium difficile Infection (CDI) Outcome Measure
- 0138 National Healthcare Safety Network (NHSN) Catheter-associated Urinary Tract Infection (CAUTI) Outcome Measure
- 0139 National Healthcare Safety Network (NHSN) Central line-associated Bloodstream Infection (CLABSI) Outcome Measure
- 0174 Improvement in bathing
- 0175 Improvement in bed transferring
- 0177 Improvement in pain interfering with activity

NATIONAL QUALITY FORUM
Driving measurable health improvements together

Impact Analysis Measures That Did Not Meet Recommended Accountable Entity Level Thresholds

- 0674 Percent of Residents Experiencing One or More Falls with Major Injury (Long Stay)
- 0679 Percent of High Risk Residents with Pressure Ulcers (Long Stay)
- 0684 Percent of Residents with a Urinary Track Infection (Long Stay)
- 0230 Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute myocardial infarction (AMI) hospitalization
- 0330 Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization
- 0369 Standardized Mortality Ratio for Dialysis Facilities
- 0505 Hospital 30-day all-cause risk-standardized readmission rate (RSRR) following acute myocardial infarction (AMI) hospitalization.

NATIONAL QUALITY FORUM
Driving measurable health improvements together

Impact Analysis Measures That Did Not Meet Recommended Accountable Entity Level Thresholds (continued)

- 0506 Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization
- 1891 Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization
- 2561 STS Aortic Valve Replacement (AVR) Composite Score
- 2563 STS Aortic Valve Replacement (AVR) + Coronary Artery Bypass Graft (CABG) Composite Score
- 3366 Hospital Visits after Urology Ambulatory Surgical Center Procedures
- 3597 Clinician-Group Risk-Standardized Acute Hospital Admission Rate for Patients with Multiple Chronic Conditions under the Merit-based Incentive Payment System