NATIONAL QUALITY FORUM
Driving measurable health improvements together

http://www.qualityforum.org

Scientific Methods Panel Web Meeting

May 4, 2021

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

Welcome

NQF Scientific Methods Panel Team

- Senior Lead
  - Sheri Winsper, RN, MSN, MSHA, Senior Vice President
- Project Management
  - Sharon Hibay, DNP, BS, RN, Senior Consultant
  - Mike DiVecchia, MBA, PMP, Senior Project Manager
  - Caitlin Flouton, MS, Senior Analyst
  - Hannah Ingber, MPH, Senior Analyst

Scientific Methods Panel Members

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 improve and clarify guidance on reliability testing for future measure development and evaluation cycles
- To discuss overarching themes
  - Voting considerations for multi-item measures
  - To consider a policy change that requires reliability accountable entity (formerly known as measure score level) testing for all maintenance measures
  - Availability of evidence and conceptual framework content for scientific acceptability evaluations
  - Utility of collapsing high and moderate testing scores into one pass score
- To discuss testing considerations as related to NQF's Best Practices for Developing and Testing Risk Adjustment Models project (referred to as the "risk adjustment project")
- Topic considerations for future advisory discussions

NATIONAL QUALITY FORUM

Meeting Agenda

- Reliability
- Overarching Themes
- Risk Adjustment Project
- Opportunity for Public Comment
- Next Steps

Reliability

NATIONAL QUALITY FORUM

Reliability Objectives

- To provide a concrete reliability guidance on minimum acceptable levels of testing, methods, performance and volume thresholds, sampling, level of analysis, and other considerations
- Adjudicate and accept concepts for the Reliability Proposed Sample Table, including testing types, levels, purpose, and thresholds (i.e., unacceptable, adequate, and high)
- To provide guidance on the reliability testing guiding principles:
  - Each reliability test should have its own rule-of-thumb guideline
  - Patient-/encounter-level and accountable entity-level analyses also may require different standards and thresholds.

NATIONAL QUALITY FORUM

Additional Emerging Themes on Reliability Testing

- Minimum acceptable performance thresholds recommendations for samples and volumes
- Small or low volume testing needs:
  - Additional testing for low volumes, such as inter-quartiles or other descriptive statistics
  - Multi-year pooling for outcome measures
  - Varying sample requirements to balance acceptable precision thresholds and intended use
  - Assessing generalizability of low volume accountable entity results to generalizability to all accountable entities based on characteristics, (e.g., size, providers, rurality)
- Other considerations?

NATIONAL QUALITY FORUM

Reliability: Testing and Minimum Thresholds

- To provide ongoing technical support to the measurement community, NQF staff, in conjunction with the SMP members, proposed minimum acceptable performance thresholds recommendations for samples, volumes, and timing
- Testing Levels
  - Person-/Encounter-Level (e.g., data element)
  - Accountable/Reporting Entity Level (e.g., performance or measure score)
- Test & Use (e.g., Cronbach's Alpha for survey items)
- The purpose of the test (e.g., Cronbach's Alpha tests the internal consistency of items in a multi-item scale)
- Thresholds
  - Unacceptable, Adequate, High

NATIONAL QUALITY FORUM

Articles Offering Rule-of-Thumb Guidance for Reliability Statistics

- Adams, JL, et al. Physician cost profiling–reliability and risk of misclassification. *N Engl J Med* 2010; 362: 1014–1024.
- Bland, JM. & Altman, DG. Statistics notes: Cronbach's alpha. *BJM* 1997; 314(7080): 572.
- Eijkenaar F, van Vliet RC. (2013). Profiling individual physicians using administrative data from a single insurer: variance components, reliability, and implications for performance improvement efforts. *Med Care*. 51(8):731-9.
- He, K. et al. Inter-Unit reliability for quality measure testing. *J Hosp Adm* 2019; 8(2): 1–6.
- Koo, TK. & Li, MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016; 15(2): 155-163.
- Landis J, Koch G. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159-174.
- McGraw, KO. & Wong, SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996; 1(1): 30-46
- Staggs, VS. & Gajewski, BJ. Bayesian and frequentist approaches to assessing reliability and precision of health-care provider quality measures. *Stat Methods Med Res* 2015; 26(3).

Overarching Themes

NATIONAL QUALITY FORUM

Overarching Themes

- Voting considerations for multi-item measures
- NQF policy change to require reliability accountable entity (formerly known as measure score level) testing for all measure submissions
- Additional overarching themes
  - Identifying the utility of collapsing the high and moderate reliability or validity voting scores into a single pass score
  - Availability of measure evidence and conceptual framework for validity evaluations, especially for new measures
  - Voting challenges with high patient- encounter-level/data element testing and low accountable entity/measure score testing

NATIONAL QUALITY FORUM

Voting on Multi-Item Measures

- Multi-item measures have multiple component measures but do not roll up to an overall composite score.
- NQF's current evaluation guidance states, "Measures with multiple measure components that are assessed for each patient, but that result in multiple scores for an accountable entity, rather than a single score. These generally should be submitted as separate measures and indicated as paired/grouped measures." (p. 51)
- Voting options include:
  - Keep the above stated policy and modify the measure submission forms
  - Vote the multi-item measure as a "package"
  - Assign individual NQF measures (or sub-numbers) for each measure item to vote on individual items within the measure. NQF does not currently vote on individual items within a measure.

Accountable Entity/Measure Score Testing Policy

- SMP members requested to discuss:
  - Require accountable entity/performance score-level reliability testing for all maintenance measures
  - Require empirical validity testing for all maintenance measures
  - Ratings for maintenance measures to be based on accountable entity level reliability and empirical validity testing
- Should this change be enacted, NQF staff would engage the measurement community and query the potential needs.
  - Identify the steps and projected timeline for the policy change, including NQF staff guidance and Consensus Standards Approval Committee (CSAC) approval
  - Engaging the measure community and other external stakeholders for input on this policy change
  - Identify technical assistance and support that developers might need to meet this change
  - Assess the implications this change will have on measure submissions

NATIONAL QUALITY FORUM

Additional Overarching Themes

- Identifying the utility of collapsing the high and moderate reliability or validity voting scores into a single pass score
  - Utility to high-stakes and quality improvement purposes
  - Modifications to scientific acceptance algorithms
- Availability of measure evidence and conceptual framework for validity evaluations, especially for new and non-clinical measures
- Voting challenges with high patient- or encounter-level/data element testing and low accountable entity/measure score testing
  - The current guidance prioritizes patient- or encounter-level/data element testing.
  - How would a policy change for accountable entity/performance score level testing for all maintenance measures effect this challenge?
  - How would this change effect the scientific acceptance algorithms?

Risk Adjustment

NATIONAL QUALITY FORUM

Risk Adjustment TEP Project Objectives

- Conduct an environmental scan of data sources used for risk adjustment, functional or social risk factors available for testing, and approaches to conceptual and statistical methods for risk adjustment
  - Commenting period closed March 17, 2021
  - Draft Environmental Scan available here
  - Final version will be available May 10, 2021
- Develop Technical Guidance (TG) for measure developers that includes emerging best practices on when and how to adjust for functional and social risk factors in measure development
  - SMP input and counsel on TG to align with SMP members' expectations
- An option year may extend the project to enhance/update the Environmental Scan and TG with key informant interviews

NATIONAL QUALITY FORUM

The Technical Guidance Will Cover the Following

- Conceptualizing a Model
  - Describing the Rationale for Risk Adjustment or Not
- Identifying and Selecting Potential Data Sources and Variables
- Empirically Testing Risk Factors
  - Assessing the variation in prevalence of the factor across measured entities
  - Empirically testing the association between the factor with the outcome
  - Empirically testing the contribution of unique variation in the outcome in a multivariable model
  - Assessing the between-unit effects versus within-unit effects
  - Determining the impact of adjusting for risk (or not) on accountable entities in the tails of the performance distribution
- Empirically Testing the Adequacy of the Risk Model
  - Model discrimination
  - Model calibration
- Considerations for Determining the Final Risk Adjustment Model

NATIONAL QUALITY FORUM

Considerations of a Standard Risk Adjustment Framework

- A statistical rational for a standard framework provides an opportunity for more consistency and credibility of risk adjustment models
- Standards cannot be too rigid to allow for differences in:
  - Data availability
  - Patient populations
  - Quality constructs, diseases, and outcomes
- A standard should always encourage consideration of how all risk (i.e., clinical, demographic, and social) impacts the measured outcome from the beginning, not only in the end.
- Any standard that is developed will acknowledge present vs. future state. NQF will plan to revisit the framework regularly and update it.
- Different uses will necessitate some provider accountability for factors that are traditionally considered applicable to risk adjustment

NATIONAL QUALITY FORUM

Questions for the Scientific Methods Panel

1. Do you have any concerns about the development and use of a standard risk adjustment framework?
2. Agree or disagree: Better clinical risk adjustment can obviate the need for additional social risk adjustment
   - *What positions does the SMP have on ordering of factors?*
3. Agree or disagree: TG should encourage an analysis of risk adjustment in different groups (e.g., calibration in subgroups).
   - *When and why are differences expected and how will this impact adjustment?*
4. What about the clinical/practical significance of a factor? What if the association is not statistically significant, is that grounds to not include the factor in the final model? What if it is statistically significant?

Risk Adjustment Project Next Steps

- Members of the SMP are invited to join the next risk adjustment project web meeting to continue this conversation
- Objectives of the meeting include:
  - Obtain TEP input on the first draft of the Technical Guidance report
  - Continue discussion on minimum standards, requirements, and good examples/practices
- SMP members will ultimately review measures that follow the TG
- The risk adjustment TEP is looking to confirm their thoughts on the TG with SMP members
- Next web meeting will take place on **May 13, 2021, 1:00-3:00pm ET.** Please visit the Risk Adjustment project page to see meeting materials and learn more

Next Steps

NATIONAL QUALITY FORUM

Concepts for Future Advisory Guidance

- Review NQF's evaluation guidance to evaluate measures with:
  - Non-clinical conceptual framework (e.g., home- and community-based services (HCBS))
  - Diagnostic accuracy
- Validity Testing Technical Support to Developers
  - Developing a validity testing Constructing adequate tests of face validity testing
  - Identifying measure correlates or comparators for construct validity
  - Using predictive validity to assess appropriateness of outcome measure to evaluate performance
- Modifications to guidance testing algorithms
- Identifying the utility of collapsing the high and moderate reliability or validity voting scores into a single pass score
- Availability of measure evidence and conceptual framework for validity evaluations

NATIONAL QUALITY FORUM

What's Next

- Meeting summary review by the SMP members; will be posted on NQF website
- Important upcoming dates:
  - Risk Adjustment TEP Meeting May 13, 1:00-3:00pm ET
  - NQF 2021 Annual Conference July 20 – 22
- Next meeting: To Be Determined
  - July 20 SMP meeting will be rescheduled
  - A survey with proposed meeting times for rescheduling will be sent to SMP members

THANK YOU.

NATIONAL QUALITY FORUM

http://www.qualityforum.org