



**NATIONAL
QUALITY FORUM**
Driving measurable health
improvements together

<http://www.qualityforum.org>

Scientific Methods Panel Fall 2021 Evaluation Meeting

October 26-27, 2021

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

Housekeeping Reminders – Day 1

- This is a Webex meeting with audio and video capabilities:
 - ▣ **Meeting link:** <https://nqf.webex.com/nqf/j.php?MTID=m78e188f77a6b6215e1d47936417f6b80>
 - ▣ **Meeting number:** 2347 849 8731
 - ▣ **Password:** QMEvent
- Optional: Dial **1-844-621-3956** and enter passcode [2347 849 8731]
- Please place yourself on mute when you are not speaking
- 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 Scientific Methods Panel roll call once the meeting begins

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

Housekeeping Reminders

- Meeting breaks
- Voting Quorum
- Chat feature
- Raising hand
- Muting and unmuting
- If possible, do not speak on speaker phone
- Introduce yourself; we are transcribing the discussion
- Technical support

Day 1: Welcome, Introductions, and Disclosures of Interest

Welcome

NQF Scientific Methods Panel (SMP) Team

- Tricia Elliott, MBA, CPHQ, FNAHQ, Senior Managing Director
- Mike DiVecchia, MBA, PMP, Senior Project Manager
- Hannah Ingber, MPH, Senior Analyst
- Gabby Kyle-Lion, MPH, Coordinator
- Sharon Hibay, DNP, BS, RN, Senior Consultant
- Elisa Munthali, MPH, Senior Consultant
- Jill Herndon, PhD, Consultant

Scientific Methods Panel Members

- **David Nerenz, PhD, Co-chair**
- **Christie Teigland, PhD, Co-chair**
- J. Matt Austin, PhD
- Bijan Borah, MSc, PhD
- John Bott, MBA, MSSW
- Daniel Deutscher, PT, PhD
- Lacy Fabian, 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, MPH
- 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

Meeting Overview

Meeting Agenda: Day 1

- Welcome, Introductions, and Disclosures of Interest
- Evaluation Updates (Spring 2021 and Fall 2021 cycles)
- Process Overview and Evaluation Reminders
- Break 12:00 – 12:30PM EST
- Fall 2021 Measure Evaluations
- Opportunity for public comment
- Adjourn

Meeting Materials

■ Discussion Guide

- » A synopsis document of scientific acceptability content (i.e., reliability and validity requirement) for all complex measures in a measure cycle evaluated by the SMP members.
 - *Each measure includes pertinent information from the submission, SMP reviewer feedback, related developer responses, and identification of measures that are pulled for SMP discussion.*
 - *Goal is to summarize and highlight priority information for SMP discussion, reduce developer burden from multiple submission materials requests, and target critical scientific acceptability questions/concerns*
- » Appendix B: Additional information provided by measure developers

■ Background Materials

- » [2011 Testing Task Force Report](#)
- » [2021 NQF Measure Evaluation Criteria and Guidance](#)
- » [SMP Measure Evaluation Guidance](#)

Spring 2021 Evaluation Updates

Spring 2021 SMP Measure Evaluation Cycle Statistics

- 29 measures were evaluated by the SMP
 - » 13 measures were discussed at the meeting (45% of total)
- Final results
 - » 23 of 29 measures passed SMP and were evaluated by the Standing Committees (79%)
 - *2 measures were consensus not reached (CNR) by the SMP, evaluated by Standing Committees (7%)*
 - *2 did not pass (7%)*
 - *2 withdrawn mid-cycle (7%)*
 - » Standing Committees revoted on Scientific Acceptability for 2 of the 29 measures

Spring 2021 SMP Measures Revoted on by the Standing Committee

NQF ID	Measure Title	SMP Decision	Standing Committee and Decision	Current Status
3621	Composite weighted average for 3 CT Exam Types: Overall Percent of CT exams for which Dose Length Product is at or below the size-specific diagnostic reference level (for CT Abdomen-pelvis with contrast/single phase scan, CT Chest without contrast/single)	Reliability: Pass Validity: CNR Composite: Pass	Patient Safety Reliability: Pass Validity: Pass Composite: Pass	Recommended for Endorsement
3622	National Core Indicators for Intellectual and Developmental Disabilities (ID/DD) Home- and Community-Based Services (HCBS) Measures	Reliability: Pass Validity: CNR	Patient Experience and Function Reliability: Pass Validity: Pass	Recommended for Endorsement



Performance Metrics

Metrics	Fall 2017	Spring 2018	Fall 2018	Spring 2019	Fall 2019	Spring 2020	Fall 2020	Spring 2021	Fall 2021*
Total number of complex measures submitted for evaluation by the SMP	8	21	39	47	22	21	25	29	12
Total Passed	4	7	25	30	17	16	20	23	8
Total Not Passed	4	13	10	11	4	3	3	2	2
Consensus Not Reached**	0	1	4	6	1	2	2	2	2
Percent agreement with Standing Committee ratings and SMP recommendations	6/8 (75%)	100%	23/29 (79%)	35/47 (74%)	16/18 (89%)	13/18 (72%)	21/22 (95%)	19/22 (86%)	TBD

TBD: to be determined

*Data for the Fall 2021 cycle are preliminary

**These measures were sent to the Standing Committees

Fall 2021 Cycle Overview

Fall 2021 Evaluation Cycle Statistics

- 12 complex measures were assigned to the SMP
 - » 11 were new measures
- 2 subgroups of 12-13 SMP members were each assigned 6 measures
 - » 8 measures passed reliability AND validity
 - » 2 measures were consensus not reached (CNR) on reliability OR validity
 - » 1 measure did not pass on reliability
 - » 2 measures did not pass validity
 - » 3 measures were withdrawn prior to SMP preliminary review
 - » 7 slated for discussion
- Reviewed Measures by Types
 - » 4 outcome
 - » 0 cost/resource use
 - » 0 composite
 - » 3 outcome: intermediate clinical outcome
 - » 4 PRO-PM
 - » 1 process
 - » 0 structure



Fall 2021 Measures Slated for Discussion

- Subgroup 1

- » *Surgery*

- #3649e
 - #3650e
 - #3652e
 - #3638
 - #3639

- » *Primary Care and Chronic Illness*

- #3667

- Subgroup 2

- » *Patient Safety*

- #0689

Process Overview and Evaluation Reminders

Overall Ratings

■ High (H)

- » Score-level testing is required
- » A measure may be eligible for “HIGH,” but the sampling method/results may warrant a “MODERATE” rating

■ Moderate (M)

- » 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 warrant a “LOW” rating

■ Low (L)

- » Used primarily if testing results are not satisfactory or an inappropriate methodology was applied

■ Insufficient (I)

- » Use when the reviewer does not have sufficient information to assign a “HIGH,” “MODERATE,” or “LOW” rating
 - *Examples: unclear specifications; unclear testing methodology, not conducting criteria required testing*

Meeting Quorum and Achieving Consensus

- A meeting quorum is met with 66% of active SMP Members in attendance
- Achieving consensus is calculated from the percent of quorum members during a vote
- SMP scientific acceptability (i.e., reliability and validity criteria) evaluation results
 - » Pass/Recommended: Greater than 60% “Yes” of quorum votes (i.e., high + moderate ratings)
 - » Consensus not reached (CNR): 40-60% “Yes” of quorum votes (inclusive of 40% and 60%)
 - » Does not pass/Not recommended: Less than 40% “Yes” of quorum votes

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** patient or encounter level (previously known as data element level) testing **OR** accountability entity level (previously known as measure score level) testing for new measures
 - *Both testing types are preferred, yet not currently required*
 - *Impacts rating, as described previously*
 - *Exception: face validity testing of the “computed measure score” for new measures is accepted at the accountable entity level*
 - » If patient-/encounter-level (i.e., data element level) validity testing is provided, we **do not** require additional reliability testing
 - *In this case, use the rating you give for validity as the rating for reliability*
 - *Submissions that accept patient-/encounter-level validity testing for patient-/encounter-level reliability testing is occurring less frequently in recent measure cycles*

Differences in Testing Requirements by Measure Type – *Instrument-based Measures (including PRO-PMs)*

- For reliability and validity, testing is required at **both** patient/encounter (i.e., data element) and accountable entity (i.e., measure score) levels for initial endorsement evaluation
 - » Patient-/Encounter level testing must be conducted for reliability AND validity of the multi-item scales at the patient level
 - » Accountable entity level testing must be conducted for reliability AND validity testing of the actual performance measure at the level of analysis as defined in the measure specifications
 - *Face validity testing of the “computed measure score” is accepted at initial endorsement evaluation in lieu of empirical accountable entity level validity testing*

Differences in Testing Requirements by Measure Type – *Composite Measures*

- NQF provides specific guidance and definitions for “composite” measures
 - » Components of the composite measure should have their own properties of reliability and validity
 - » NQF does NOT consider multi-item scales in surveys/questionnaires as composites
 - » NQF does NOT consider multiple component measures without a single performance rate and multiple component performance rates as composites
- Accountability entity (i.e., measure score) level reliability testing of the composite is required
- Demonstrating reliability of individual components alone is not sufficient to pass the criterion
- Accountability entity (i.e., measure score) level validity testing is not required until maintenance
- Additional scientific acceptability subcriterion is required for composite measures
 - » Empirical analyses supporting the composite construction including the value of the components to the composite and the component aggregation and weighting consistency to composite quality construct

Testing and Evaluation Reminders

- All testing must align with specifications
 - » This is not a new requirement, yet NQF is more rigorously in upholding the requirement, particularly for level of analysis testing and minimum sample sizes
 - *If multiple levels of analysis are specified, each must be tested separately*
 - » NQF's requirements permit passing some or all levels of analysis for a measure
- Occasionally there are several performance measures included under one NQF number
 - » Each measure must be evaluated separately; some measures may pass and others may not pass

Additional Reminders

- Consideration for risk-adjustment is required for all outcome, resource use, and some process measures
 - » Inclusion (or exclusion) of certain factors in the risk-adjustment approach **should not** be a reason for not passing a measure
 - » Concerns with discrimination, calibration, or overall method of adjustment are grounds for not passing a measure
 - » In the absence of a risk adjustment for outcome, resource use, and some process measures, a strong rationale/data for excluding must be provided
- For all measures
 - » Incomplete or ambiguous specifications are grounds for not passing a measure—but remember that there is an option to get clarifications, although this must be done early on
- Empirical validity testing is required at time of maintenance evaluation
 - » If not possible, a strong justification is required and must be accepted by the Standing Committee

Additional Reminders (continued)

- The SMP articulated additional **guidance** for submissions
 1. Provide greater detail when describing testing methodologies and results
 2. Provide more than one overall statistic when conducting signal-to-noise reliability testing
 3. Provide greater detail in description of construct validation describing:
 - » Hypothesized relationships
 - » Why examining hypothesized relationships would validate the measure
 - » Expected direction and strength of the association
 - » Specific statistical tests used, results, results interpretation, how the results related to hypothesis, and whether the results assist to validate the measure
- Lack of #2 and #3 should not be grounds for not passing a measure

Standing Committee Complex Measure Evaluation

- All measures reviewed by the SMP can be **discussed** by the Standing Committees
 - » Standing Committees will evaluate and make recommendations for endorsement for:
 - *Measures that pass SMP review*
 - *Measures where the SMP did not reach consensus (i.e., CNR)*
 - » Measures that do not pass the SMP may be pulled by a Standing Committee member for further discussion and revote if it is an **eligible** measure

Committee Consideration of Measures that Do Not Pass the SMP

- Eligibility will be determined by NQF Staff and SMP co-chairs
 - » Measures that did not pass the SMP due to the following will not be eligible for revote:
 - *Inappropriately applied methodology or testing approach to demonstrate reliability or validity*
 - *Incorrect calculations or formulas used for testing*
 - *Description of testing approach, results, or data is insufficient for SMP to apply the criteria*
 - *Appropriate levels of testing not provided or otherwise did not meet NQF's minimum evaluation requirements*

Fall 2021 Measure Evaluation: Day 1

Measure Discussion Process

- Measures discussed by the SMP are determined during the SMP measure review activities
- Staff will briefly introduce the measure
- SMP member lead discussants will summarize key concerns
- Other SMP subgroup members are invited to comment
- Developers are given 2-3 minutes for an initial response, and may respond to SMP questions
- Discussions are opened to the full SMP and proceed by individual criterion
- **Recused members cannot discuss measures where conflicts are identified**

The Voting Process

- Voting is conducted synchronously, virtually, and confidentially via Poll Everywhere
- Voting occurs following each criterion discussion
- SMP subgroup members only vote on measures they were assigned
- **Recused SMP members cannot vote for measures where conflicts are identified**
- Subgroup voting results taken during the meeting are the official SMP vote
- Measures that are not pulled for discussion will pass in a consent calendar vote

Voting Test

#0689 Percent of Residents Who Lose Too Much Weight (Long-Stay)

- Subgroup 2
- Preliminary Voting Result:
 - ▣ Reliability: H-3, M-5, L-3, I-0 *Pass*
 - ▣ Validity: H-1, M-5, L-5, I-0 *CNR*
- Lead Discussant: Jeffrey Geppert
- Measure Developer: Acumen/Centers for Medicare & Medicaid Services
- Measure Steward: Centers for Medicare & Medicaid Services
- Discussion Guide page 19
- For SMP discussion:
 - ▣ The SMP needs to discuss and re-vote on the validity concerns, including the correlation testing results and the lack of risk adjustment.

Break

Will resume at 1:00 pm EST

#3649e: Risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA) electronic clinical quality measure (eCQM)

- Subgroup 1
- Preliminary Voting Result:
 - ▣ Reliability: H-1, M-3, L-6, I-0 *CNR*
 - ▣ Validity: H-0, M-8, L-1, I-1 *Pass*
- Lead Discussant: Sherri Kaplan
- Measure Developer: Brigham and Women's Hospital
- Measure Steward: Brigham and Women's Hospital
- Discussion Guide page 5
- For SMP discussion:
 - ▣ How does the SMP view the results of reliability and validity testing?
 - ▣ How does the SMP view the validity and methods for building the risk adjustment model?



#3650e Risk-standardized inpatient respiratory depression (IRD) rate following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA) eCQM

- Subgroup 1
- Preliminary Voting Result:
 - ▣ Reliability: H-0, M-7, L-4, I-0 *Pass*
 - ▣ Validity: H-0, M-3, L-8, I-0 *No Pass*
- Lead Discussant: David Nerenz
- Measure Developer: Brigham and Women's Hospital
- Measure Steward: Brigham and Women's Hospital
- Discussion Guide page 8
- For SMP discussion:
 - ▣ Does the SMP have concerns related to the small testing samples?
 - ▣ For reliability, how does the SMP interpret the low Spearman correlation coefficient in light of the ICC score?
 - ▣ How does the SMP interpret the low TEP (3/7, 43%) face validity testing results?

#3652e Risk-standardized prolonged opioid prescribing rate following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA) eCQM

- Subgroup 1
- Preliminary Voting Result:
 - ▣ Reliability: H-0, M-7, L-3, I-1 *Pass*
 - ▣ Validity: H-2, M-5, L-4, I-0 *Pass*
- Lead Discussant: Jennifer Perloff
- Measure Developer: Brigham and Women's Hospital
- Measure Steward: Brigham and Women's Hospital
- Discussion Guide page 10
- For SMP discussion:
 - ▣ How does the SMP interpret the Spearman Correlation Coefficient and ICC, in terms of demonstrating accountable entity reliability?
 - ▣ Does the expanded timeline for the test-retest sample effect the reliability of the measure?
 - ▣ How does the SMP interpret the validity testing results in light of missing data?
 - ▣ Is the risk adjustment strategy appropriate for this process measure?

Break

Will resume at 3:00PM EST

#3638 Care Goal Achievement Following a Total Hip Arthroplasty (THA) or Total Knee Arthroplasty (TKA)

- Subgroup 1
- Preliminary Voting Result:
 - ▣ Reliability: H-0, M-1, L-5, I-3 *No Pass*
 - ▣ Validity: H-0, M-3, L-3, I-3 *No Pass*
- Lead Discussant: Jack Needleman
- Measure Developer: Brigham and Women's Hospital
- Measure Steward: Brigham and Women's Hospital
- Discussion Guide page 12
- For SMP discussion:
 - ▣ Is there any new information that the SMP would like to revisit related to the reliability vote, specifically expressed concerns about the reliability of the PROM?
 - ▣ The SMP should discuss the effect that small sample sizes have on measure's reliability and validity.
 - ▣ How does the SMP interpret the validity testing results?



#3639: Clinician-Level and Clinician Group-Level Total Hip Arthroplasty and/or Total Knee Arthroplasty (THA and TKA) Patient-Reported Outcome-Based Performance Measure (PRO-PM)

- Subgroup 1
- Preliminary Voting Result:
 - ▣ Reliability: H-3, M-3, L-1, I-2 *Pass*
 - ▣ Validity: H-0, M-7, L-1, I-1 *Pass*
- Lead Discussant: Daniel Deutscher
- Measure Developer: Yale CORE
- Measure Steward: Centers for Medicare & Medicaid Services
- Discussion Guide page 14
- For SMP discussion:
 - ▣ How does the SMP view the results of reliability testing for this PRO-PM?
 - ▣ Does the testing sample adequately represent diverse populations to be generalized among and between all groups?
 - ▣ How do the large non-response bias volumes and missing risk factors, especially in social risk populations, impact validity?

#3667 Days at Home for Patients with Complex, Chronic Conditions

- Subgroup 1
- Preliminary Voting Result:
 - ▣ Reliability: H-5, M-6, L-0, I-0 *Pass*
 - ▣ Validity: H-2, M-7, L-2, I-0 *Pass*
- Lead Discussant: Patrick Romano
- Measure Developer: Yale CORE
- Measure Steward: Centers for Medicare & Medicaid Services
- Discussion Guide page 16
- For SMP discussion:
 - ▣ Was the measure assessed and tested for an ACO and/or provider group level?
 - ▣ Is the risk adjustment approach sound and methodologically appropriate?
 - ▣ Is the exclusions list comprehensive?
 - ▣ Should the testing sample stretch beyond one year?

Opportunity for Public Comment

Next Steps

Next Steps Explained

- Tomorrow's (10/27) meeting will be from 3:00PM-5:00PM
- Topics for discussion
 - ▣ CSAC Update
 - ▣ Reliability Thresholds – Patient/Encounter level testing
 - ▣ Maintenance Reliability testing – Accountable Entity Level
 - ▣ Maintenance Validity Testing – Accountable Entity Level

Adjourn

Day 2: Welcome, Review of Agenda

Housekeeping Reminders – Day 2

- This is a Webex meeting with audio and video capabilities:
 - ▣ **Meeting link:** <https://nqf.webex.com/nqf/j.php?MTID=m7c1ac5b4ba6cb0a616b54deae2b6c23b>
 - ▣ **Meeting number:** 2339 920 4941
 - ▣ **Password:** QMEvent
- Optional: Dial **1-844-621-3956** and enter passcode [2339 920 4941]
- Please place yourself on mute when you are not speaking
- 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 Scientific Methods Panel roll call once the meeting begins

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

Scientific Methods Panel Members(Cont.)

- **David Nerenz, PhD, Co-chair**
- **Christie Teigland, PhD, Co-chair**
- J. Matt Austin, PhD
- Bijan Borah, MSc, PhD
- John Bott, MBA, MSSW
- Daniel Deutscher, PT, PhD
- Lacy Fabian, 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, MPH
- 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

Meeting Agenda: Day 2

- Welcome
- Advisory discussion
 - ▣ CSAC Update
 - ▣ Reliability Thresholds – Patient/Encounter level testing
 - ▣ Maintenance Reliability testing – Accountable Entity Level
 - ▣ Maintenance Validity Testing – Accountable Entity Level
- Opportunity for Public Comments
- Next Steps
- Adjourn

SMP Topics Discussed with CSAC

Topic	Issue	Status
Reliability Threshold	Clarity on acceptable minimum reliability thresholds for patient-/encounter-level and accountable entity level testing, including considerations of 0.7 for specific accountable entity testing, as appropriate.	Under review at SMP
Reliability Testing – Accountable Entity Level	SMP Members recommend that empirical accountable-entity level reliability testing should be required for all measures by the time of the first maintenance cycle.	Proposed from SMP
Validity Testing – Accountable Entity Level	SMP Members recommend that empirical accountable-entity level validity testing should be required for all measures by the time of the first maintenance cycle.	Proposed from SMP

Reliability Testing Thresholds - Questions from the CSAC

- The CSAC reviewed the reliability table in full on October 12
- Regarding the threshold values, many CSAC members felt that 0.4 for Kappa should be higher
- If a measure does not meet the threshold, what is the result?
 - » The measure will not be automatically rejected by NQF Staff at initial submission
 - » The measure will be reviewed by SMP, but thresholds would be used by the SMP to justify their ratings
 - » Standing Committees will still be able to pull these measures for discussion and potential re-vote, if they are eligible for re-vote according to the criteria
- In the future, Standing Committees may find it useful to have a write-up of what each test is demonstrating about reliability for reference
- The CSAC expressed widespread appreciation and thanks for the work of the SMP



Person-/Encounter-Level Reliability Testing (i.e., data element 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	0.4
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

Accountable Reporting Entity Level Reliability Testing (i.e., performance measure score 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

Recommendations for Reliability Thresholds Timeline

While a single reliability threshold is not possible, evidence-based minimum reliability thresholds for both testing levels include testing methods and interpretation standards.

Date	Proposed Activity
10/12/2021	CSAC Input garnered
10/27/2021	<i>SMP - Patient Encounter Table 3rd Review/Approval*</i>
11/2021	Public Comment
11/2021	Measure Developer Advisory Panel Review
11/2021	Standing Committee Advisory Group Review
12/14/2021	SMP Review/Status
12/16/2021	Measure Developer Webinar
Q1 2022	SMP Review/Status/Approval
Q1 2022	<i>CSAC Review/Status/Approval</i>
Q3 2022	Implementation in Guidance documents

* Note: If SMP consensus not reached, public comment will be delayed

Reliability Testing at the Accountable Entity Level for Maintenance Measures – CSAC Responses

- In May 2021, the SMP made a recommendation to require accountable entity level reliability testing for all maintenance measures.
 - » NQF's current policy allows for measures to undergo maintenance evaluation with only patient-/encounter-level testing, when performance data should be available for reliability testing.
 - » Requiring accountable entity-level reliability testing for all maintenance measures tests the consistency of specifications implemented in practice.
- CSAC members agreed with requiring reliability testing at the accountable entity level at maintenance
- This policy will move forward to public comment

Reliability Testing at the Accountable Entity Level for Maintenance Measures

In May SMP members also agreed that further discussion was needed regarding the following item:

- (1) prioritizing accountable entity-level reliability testing when patient/encounter-level validity testing is also present. NQF's current policy allows for measures with moderate patient/encounter-level testing to pass, despite having very poor accountable entity-level reliability testing.

Question for the SMP:

- Should accountable entity-level reliability testing results be prioritized over patient/encounter-level validity testing at maintenance?

Reliability Testing at the Accountable Entity Level for Maintenance Measures – Proposed Next Steps

Date	Proposed Activity
10/12/2021	CSAC Input garnered
10/27/2021	<i>SMP – Further Discussion</i>
11/2021	Public Comment
11/2021	Measure Developer Advisory Panel Review
11/2021	Standing Committee Advisory Group Review
12/14/2021	SMP Review/Status
12/16/2021	Measure Developer Webinar
Q1 2022	SMP Review/Status/Approval
Q1 2022	<i>CSAC Review/Status/Approval</i>
Q3 2022	Implementation in Guidance documents

Validity Testing at the Accountable Entity Level for Maintenance Measures - CSAC Responses

In July 2021, the SMP made two recommendations for validity testing in maintenance evaluations:

1. Validity testing at the accountable-entity level should be required for all maintenance measures.
 - *If measure developers are unable to meet this requirement, NQF should require, and developers should provide, a strong rationale supporting this rare instance.*
2. Measures submitted for maintenance evaluations with face validity testing should include other level validity testing (i.e., accountable-entity validity testing or, in rare instances, patient-/encounter-level validity testing with a strong rationale for not performing accountable-entity validity testing).
 - CSAC members largely agreed, but raised additional questions for SMP consideration

Questions for the SMP:

- What rational(s) would justify a lack of accountable-entity validity testing at maintenance?
- What rational(s) would justify a lack of accountable-entity empirical validity testing at maintenance?

Validity Testing at the Accountable Entity Level for Maintenance Measures – Proposed Next Steps

Date	Proposed Activity
10/12/2021	CSAC Input garnered
10/27/2021	<i>SMP – Further Discussion</i>
11/2021	Public Comment
11/2021	Measure Developer Advisory Panel Review
11/2021	Standing Committee Advisory Group Review
12/14/2021	SMP Review/Status
12/16/2021	Measure Developer Webinar
Q1 2022	SMP Review/Status/Approval
Q1 2022	<i>CSAC Review/Status/Approval</i>
Q3 2022	Implementation in Guidance documents

Opportunity for Public Comment

Next Steps

Next Steps and Reminders

- Full measure submission deadlines: November 1, 8, 15
- NQF staff will summarize the relevant measure information and discussions of the SMP, and provide to the various standing committees
 - » These Standing Committees will evaluate measures in February
 - » CSAC will review Spring 2021 measures on November 30 – December 1
 - » CSAC will review Fall 2021 measures in the June-July timeframe
- Next Intent to Submit deadline (Spring 2022): January 5, 2022

2021 SMP Meetings

Meeting Date	Tentative Topic/Activity
Tuesday December 14 from 12:00 – 2:00 pm ET	<ul style="list-style-type: none">- Reliability threshold table status update and discussion- Testing requirement policies for maintenance measures

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>

Adjourn

THANK YOU.

NATIONAL QUALITY FORUM

<http://www.qualityforum.org>