



**NATIONAL
QUALITY FORUM**

Driving measurable health
improvements together

<http://www.qualityforum.org>

Scientific Methods Panel Spring 2021 Evaluation Meeting

March 30-31, 2021

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

Day 1: Welcome, Introductions, and Disclosures of Interest



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

Welcome



NQF Scientific Methods Panel (SMP) Team

- Senior Lead
 - ▣ Sai Ma, PhD, Managing Director & Senior Technical Expert
- Project Management
 - ▣ Mike DiVecchia, MBA, PMP, Senior Project Manager
 - ▣ Hannah Ingber, MPH, Senior Analyst
 - ▣ Caitlin Flouton, MS, Senior Analyst

Scientific Methods Panel Members

SMP 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

Meeting Agenda: Day 1

- Welcome, Introductions, and Disclosures of Interest
- Evaluation Updates (Fall 2020 and Spring 2021 cycles)
- Process Overview and Evaluation Reminders
- Break 12:00 – 12:30PM EST
- Spring 2021 Measure Evaluations
 - ▣ Afternoon break 2:15 – 2:30PM EST
- Opportunity for public comment
- Adjourn



Meeting Materials

- Discussion Guide
 - ▣ Includes pertinent information from the submission, panelist reviews, and developer responses
 - » Goal is to minimize need for back-and-forth with submission materials and to guide discussion so that we address critical questions/concerns
 - ▣ Appendix B: Additional information provided by measure developers
- Background Materials
 - ▣ [2011 Testing Task Force Report](#)
 - ▣ [2019 NQF Measure Evaluation Criteria and Guidance](#)
 - ▣ [SMP Measure Evaluation Guidance](#)

Fall 2020 Evaluation Updates

Fall 2020 Evaluation Cycle Statistics

- 25 measures were evaluated by SMP
 - ▣ 8 measures were discussed at the meeting (32% of total)
- Final results
 - ▣ 20 passed SMP, evaluated by Standing Committees (80%)
 - ▣ 2 consensus not reached, evaluated by Standing Committees (8%)
 - ▣ 2 did not pass (8%)
 - ▣ 1 withdrawn mid-cycle (4%)
 - ▣ Standing Committees revoted on Scientific Acceptability for 2 measures



Fall 2020 SMP Measure Revoted on by the Standing Committee

NQF ID	Measure Title	SMP Decision	Standing Committee Decision	Current Status
0505	Hospital 30-day all-cause risk-standardized readmission rate (RSRR) following acute myocardial infarction (AMI) hospitalization.	R: CNR V: Pass	R: Pass V: Pass	Recommended for Endorsement

Performance Metrics

Metrics	Fall 2017	Spring 2018	Fall 2018	Spring 2019	Fall 2019	Spring 2020	Fall 2020	Spring 2021*
Total number of complex measures submitted for evaluation by the SMP	8	21	39	47	22	21	25	29
Total Passed	4	7	25	30	17	16	20	19
Total Not Passed	4	13	10	11	4	3	3	5
Consensus Not Reached**	0	1	4	6	1	2	2	5
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%)	TBD

TBD: to be determined

*Data for the Spring 2021 cycle are preliminary

**These measures were sent to the Standing Committees

Spring 2021 Cycle Overview



Spring 2021 Evaluation Cycle Statistics

- 29 complex measures assigned to the SMP
 - ▣ 9 new measures
- 3 subgroups of 8-9 SMP members with 9-10 measures
 - ▣ 19 passed reliability AND validity
 - ▣ 7 consensus not reached (CNR) on reliability or validity
 - ▣ 0 did not pass on reliability
 - ▣ 5 did not pass validity
 - ▣ 2 withdrawn after preliminary review
 - ▣ 13 slated for discussion
- Measure Types
 - ▣ 11 outcome
 - ▣ 8 cost/resource use
 - ▣ 3 composite
 - ▣ 2 outcome: intermediate clinical outcome
 - ▣ 2 PRO-PM
 - ▣ 2 process
 - ▣ 1 structure

Measures Slated for Discussion

Day 1

- Subgroup 1
- *All-Cause Admissions and Readmissions*
 - ▣ #2880
 - ▣ #2881
 - ▣ #2882
 - ▣ #3612
 - ▣ #3188
- *Patient Experience and Function*
 - ▣ #3622

Day 2

- Subgroup 2
- *Neurology*
 - ▣ #3614
- Subgroup 3
- *Patient Safety*
 - ▣ #3621
 - ▣ #0500
 - ▣ #0674
 - ▣ #0679
- Subgroup 1
- *Renal*
 - ▣ #3615
 - ▣ #3616

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 make you choose “MODERATE” instead

■ 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 make you choose “LOW” instead

■ Low (L)

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

■ Insufficient (I)

- Use when you don’t have sufficient information to assign a “HIGH,” “MODERATE,” or “LOW” rating
 - » Example: unclear specifications; unclear testing methodology



Achieving Consensus

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



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 score-level) testing for new measures
 - » We prefer both, but currently do not require both
 - » Impacts rating, as described previously
 - » Exception: face validity for new measures accepted
 - ▣ If patient or encounter level (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
 - » 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
 - ▣ Components of the composite measure should have own properties of reliability and validity
 - ▣ Does NOT include multi-item scales in surveys/questionnaires
- Require reliability testing of the composite measure score (i.e., accountability entity level)
- Can also show reliability testing of the components, but this is not sufficient to pass the criterion
- Score-level validity testing is 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, 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)



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 levels of analysis are specified, each must be tested separately
 - ▣ It is possible for you to “pass” part of the measure
- Occasionally there are several performance measures included under one NQF number
 - ▣ Each must be evaluated separately; some may pass and others may not pass



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 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



Additional Reminders (continued)

- The SMP articulated additional **guidance** for submissions
 1. Desire for more detail when describing methodology
 2. Requirement for more than one overall statistic if reporting on signal-to-noise reliability
 3. Desire for detail in description of construct validation (e.g., 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; or 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



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
 - ▣ Measures that do not pass the SMP can 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:
 - » Inappropriate methodology or testing approach applied 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

Break

Will resume at 12:30 pm EST

Spring 2021 Measure Evaluation: Day 1



Measure Discussion Process

- Staff will introduce the measure
- Lead discussants 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
 - ▣ Recused members cannot discuss or vote
 - ▣ Developers can respond to questions from panelists
- Final vote



The Voting Process

- Only the subgroup votes
 - ▣ Done via Poll Everywhere
 - ▣ Results from this vote will be the official vote of the SMP
- Measures not pulled for discussion: Pass with consent calendar



NATIONAL
QUALITY FORUM

Voting Test



#2880: Excess Days in Acute Care (EDAC) after Hospitalization for Heart Failure (HF)

- Subgroup 1
- Preliminary Voting Result:
 - ▣ Reliability: H-0, M-8, L-1, I-0 *Pass*
 - ▣ Validity: H-1, M-3, L-2, I-3 *Consensus Not Reached*
- Lead Discussant: Dave Nerenz
- Measure Developer: Yale Center for Outcomes Research and Evaluation (CORE)
- Measure Steward: Centers for Medicare & Medicaid Services (CMS)
- Discussion Guide page 7
- For SMP discussion:
 - ▣ The SMP should discuss the appropriateness of the risk adjustment model and revote on validity.

#2882: Excess Days in Acute Care (EDAC) after Hospitalization for Pneumonia

- Subgroup 1
- Preliminary Voting Result:
 - ▣ Reliability: H-1, M-8, L-0, I-0 *Pass*
 - ▣ Validity: H-0, M-3, L-3, I-3 *Does Not Pass*
- Lead Discussant: Sherrie Kaplan
- Measure Developer: Yale CORE
- Measure Steward: CMS
- Discussion Guide page 10
- For SMP discussion:
 - ▣ The SMP should review and discuss the issues raised, including the developer's responses, related to the appropriateness of the risk adjustment model and revote on validity.



#2881: Excess Days in Acute Care (EDAC) after Hospitalization for Acute Myocardial Infarction (AMI)

- Subgroup 1
- Preliminary Voting Result:
 - ▣ Reliability: H-0, M-4, L-5, I-0 *Consensus Not Reached*
 - ▣ Validity: V: H-0, M-3, L-3, I-3 *Does Not Pass*
- Lead Discussants: Eric Weinhandl, Larry Glance
- Measure Developer: Yale CORE
- Measure Steward: CMS
- Discussion Guide page 8
- For SMP discussion:
 - ▣ The SMP should review and discuss the issues raised, including the developer's responses related to reliability and revote on reliability.
 - ▣ The SMP should review and discuss the issues raised, including the developer's responses, related to the appropriateness of the risk adjustment model and revote on validity.

#3612: Risk-Standardized Acute Cardiovascular-Related Hospital Admission Rates for Patients with Heart Failure under the Merit-based Incentive Payment System

- Subgroup 1
- Preliminary Voting Result:
 - ▣ Reliability: H-0, M-5, L-4, I-0 *Consensus Not Reached*
 - ▣ Validity: V: H-1, M-4, L-2, I-2 *Consensus Not Reached*
- Lead Discussants: John Bott, Terri Warholak
- Measure Developer: Yale CORE
- Measure Steward: CMS
- Discussion Guide page 12
- For SMP discussion:
 - ▣ The SMP should review and discuss the issues raised regarding reliability (unit of analysis, result, etc.) and revote on reliability.
 - ▣ The SMP should review and discuss the issues raised validity and revote on validity.

Break

Will resume at 2:30PM EST

#3188: 30-Day Unplanned Readmissions for Cancer Patients

- Subgroup 1
- Preliminary Voting Result:
 - ▣ Reliability: H-0, M-7, L-2, I-0 *Pass*
 - ▣ Validity: H-0, M-3, L-4, I-2 *Does Not Pass*
- Lead Discussant: Patrick Romano
- Measure Developer/Steward: Alliance of Dedicated Cancer Centers
- Discussion Guide page 11
- For SMP discussion:
 - ▣ The SMP should review and discuss the issues raised regarding validity and revote on validity.



#3622: National Core Indicators for Intellectual and Developmental Disabilities (ID/DD) Home- and Community-Based Services (HCBS) Measures

- Subgroup 1
- Preliminary Voting Result:
 - ▣ Reliability: H-3, M-3, L-2, I-1 *Pass*
 - ▣ Validity: H-0, M-2, L-3, I-4 *Does Not Pass*
- Lead Discussant: Dave Nerenz
- Measure Developer/Steward: Human Services Research Institute
- Discussion Guide page 19
- For SMP discussion:
 - ▣ The SMP should review and discuss the issues raised regarding validity and revote on validity.

Opportunity for Public Comment

Adjourn

Day 2: Welcome, Review of Agenda



Scientific Methods Panel Members

SMP 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 Agenda: Day 2

- Welcome
- Spring 2021 Measure Evaluations
 - ▣ Break 1:00 – 1:30PM EST
- Opportunity for Public Comment
- Next Steps
- Adjourn

Spring 2021 Measure Evaluation: Day 2

#3614: Hospitalization After Release with Missed Dizzy Stroke (H.A.R.M Dizzy-Stroke)

- Subgroup 2
- Preliminary Voting Result:
 - ▣ Reliability: H-0, M-5, L-1, I-2 *Pass*
 - ▣ Validity: V: H-2, M-2, L-3, I-1 *Consensus Not Reached*
- Lead Discussants: Sam Simon, Susan White
- Measure Developer/Steward: Armstrong Institute for Patient Safety and Quality at Johns Hopkins University
- Discussion Guide page 23
- For SMP discussion:
 - ▣ The SMP should review and discuss the issues raised regarding validity and revote on validity.



#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 Phase Scan and CT Head/Brain Without Contrast/Single Phase Scan)

- Subgroup 3
- Preliminary Voting Result:
 - ▣ Reliability: H-5, M-2, L-0, I-1 *Pass*
 - ▣ Validity: H-0, M-4, L-0, I-4 *Consensus Not Reached*
 - ▣ Composite Construction: H-2, M-3, L-0, I-1 *Pass*
- Lead Discussants: Marybeth Farquhar, Matt Austin
- Measure Developer/Steward: American College of Radiology
- Discussion Guide page 30
- For SMP discussion:
 - ▣ The SMP should review and discuss the issues raised regarding validity and revote on validity.



#0500: Severe Sepsis and Septic Shock: Management Bundle

- Subgroup 3
- Preliminary Voting Result:
 - ▣ Reliability: H-5; M-1; L-0; I-2 *Pass*
 - ▣ Validity: H-3; M-2; L-1; I-2 *Pass*
 - ▣ Composite Construction: H-2; M-3; L-0; I-1 *Pass*
- Lead Discussant: Joseph Kunisch
- Measure Developer/Steward: Henry Ford Hospital
- Discussion Guide page 25
- For SMP discussion:
 - ▣ The SMP should discuss issues raised regarding reliability and validity testing, both of which were performed using only 2018 Q3-Q4 data without changes to the measure specifications.

Break

Will resume at 1:30PM EST



#0674: Percent of Residents Experiencing One or More Falls with Major Injury (Long Stay)

- Subgroup 3
- Preliminary Voting Result:
 - ▣ Reliability: H-0; M-6; L-2; I-0 *Pass*
 - ▣ Validity: H-1; M-6; L-1; I-0 *Pass*
- Lead Discussant: Alex Sox-Harris
- Measure Developer: Acumen
- Measure Steward: CMS
- Discussion Guide page 27
- For SMP discussion:
 - ▣ Is the reliability sufficient to warrant the moderate rating, given the marginal signal-to-noise ratings and the concerns regarding the stability analysis?



#0679: Percent of High Risk Residents with Pressure Ulcers (Long Stay)

- Subgroup 3
- Preliminary Voting Result:
 - ▣ Reliability: H-0; M-6; L-2; I-0 *Pass*
 - ▣ Validity: H-2; M-4; L-2; I-0 *Pass*
- Lead Discussant: Alex Sox-Harris
- Measure Developer: Acumen
- Measure Steward: CMS
- Discussion Guide page 28
- For SMP discussion:
 - ▣ Is the reliability sufficient to warrant the moderate rating?

Comparison of Reliability Statistics

NQF ID	Measure Title	Split Sample Reliability	Median SNR	Subgroup High + Moderate Reliability Votes*
0674	Percent of Residents Experiencing One or More Falls with Major Injury (Long Stay)	0.18	0.45	6/8 = 75% Pass
0679	Percent of High Risk Residents with Pressure Ulcers (Long Stay)	0.33	0.5	6/8 = 75% Pass
2881	Excess Days in Acute Care (EDAC) after Hospitalization for Acute Myocardial Infarction (AMI)	0.230 to 0.628 (depending on # of admissions)	--	4/9 = 44.4% CNR
3612	Risk-Standardized Acute Cardiovascular-Related Hospital Admission Rates for Patients with Heart Failure under the Merit-based Incentive Payment System	--	0.18 (overall) 0.60 (TINs with >20 heart failure patients)	5/9 = 55.6% CNR

*Reliability voting results presented in this table are from the subgroup's preliminary review.



#3615: Unsafe Opioid Prescriptions at the Prescriber Group Level

- Subgroup 1
- Preliminary Voting Result:
 - ▣ Reliability: H-6, M-1, L-1, I-1 *Pass*
 - ▣ Validity: H-2, M-4, L-1, I-2 *Pass*
- Lead Discussant: Patrick Romano
- Measure Developer/Steward: University of Michigan - Kidney Epidemiology and Cost Center (UM-KECC)
- Discussion Guide page 14
- For SMP discussion:
 - ▣ To what extent is the validity analysis confounded by unmeasured case mix, considering that dialysis physicians with sicker patients (such as comorbid cancer) have higher mortality rates, higher hospitalization rates, and higher opioid use?



#3616: Unsafe Opioid Prescriptions at the Dialysis Practitioner Group Level

- Subgroup 1
- Preliminary Voting Result:
 - ▣ Reliability: H-1, M-6, L-1, I-1 *Pass*
 - ▣ Validity: H-1, M-5, L-1, I-2 *Pass*
- Lead Discussant: Patrick Romano
- Measure Developer/Steward: UM-KECC
- Discussion Guide page 16
- For SMP discussion:
 - ▣ To what extent is the validity analysis confounded by unmeasured case mix, considering that dialysis physicians with sicker patients (such as comorbid cancer) have higher mortality rates, higher hospitalization rates, and higher opioid use?

Opportunity for Public Comment

Next Steps



Next Steps and Reminders

- Measure submission deadlines: April 2, 9, 16
- 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 June-July timeframe
 - ▣ CSAC will review Spring 2021 measures on November 30 – December 1
- Next Intent to Submit deadline (Fall 2021): August 2, 2021

2021 SMP Meetings

Meeting Date	Tentative Topic/Activity
Tuesday May 4 from 12:00 – 2:00 pm ET	<ul style="list-style-type: none"> - Complete the reliability guidance - Discuss risk adjustment
Tuesday July 20 from 12:30 – 2:30pm ET	<ul style="list-style-type: none"> - TBD
October 26-27 All day	Measure Evaluation Meeting (Fall 2021)
Tuesday December 14 from 12:00 – 2:00 pm ET	<ul style="list-style-type: none"> - TBD



Project Contact Info

- Email: MethodsPanel@qualityforum.org
- NQF phone: 202-783-1300
- Project page:
[http://www.qualityforum.org/Measuring_Performance/Scientific Methods_Panel.aspx](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>