



**NATIONAL  
QUALITY FORUM**

Driving measurable health  
improvements together

<http://www.qualityforum.org>

# Scientific Methods Panel Fall 2020 Meeting

*October 28 – 29, 2020*

*Funded by the Centers for Medicare & Medicaid Services under contract  
HHSM-500-2017-00060I –75FCMC19F0007.*

# **Day 1: Welcome, Introductions, and Disclosures of Interest**



## Meeting and Webinar 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



## NQF Scientific Methods Panel Team

- Senior Leads
  - ▣ Sai Ma, PhD
- Project Management
  - ▣ Mike DiVecchia, MBA, PMP
  - ▣ Hannah Ingber, MPH
  - ▣ Caitlin Flouton, MS
- CMS Contracting Office Representative
  - ▣ Michael Brea, MBA, MAS

## 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
Daniel Deutscher, PT, PhD	Sean O'Brien, PhD
Lacy Fabian, PhD	Jennifer Perloff, PhD
Marybeth Farquhar, PhD, MSN, RN	Patrick Romano, MD, MPH
Jeffrey Geppert, EdM, JD	Sam Simon, PhD
Laurent Glance, MD	Alex Sox-Harris, PhD, MS
Joseph Hyder, MD	Christie Teigland, PhD, Co-Chair
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



## Meeting Agenda

- Day 1
  - ▣ Evaluation Updates
  - ▣ Process Overview and Evaluation Reminders
  - ▣ Measure Evaluations
  - ▣ Opportunity for Public Comment
- Day 2
  - ▣ Discussion of Criteria and Evaluation Guidance
  - ▣ Opportunity for Public Comment
  - ▣ Next Steps



## Meeting Agenda: Day 1

- Welcome, Introductions, and Disclosures of Interest
- Evaluation Updates
- Process Overview and Evaluation Reminders
- Fall 2020 Measure Evaluations
- Adjourn



## 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, 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
  - Measures are included 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
- Background Materials
  - [2011 Testing Task Force Report](#)
  - [2019 NQF Measure Evaluation Criteria and Guidance](#)
  - [SMP Measure Evaluation Guidance](#)

# Evaluation Updates

# Spring 2020 Updates



## Spring 2020 Evaluation Cycle Statistics

- 21 measures evaluated by SMP
- 7 measures discussed at meeting (33% of total)
- Final results
  - ▣ 16 passed SMP, evaluated by SCs (76%)
  - ▣ 2 consensus not reached, evaluated by SCs (10%)
  - ▣ 3 did not pass (14%)
  - ▣ Standing Committees revoted on 7 measures



## Status Update on Spring 2020 Measures

NQF ID	Measure Title	SMP Decision	SC Decision	Status
3563	Medicare Spending Per Beneficiary – Post Acute Care Measure for Skilled Nursing Facilities	R: Pass V: Pass	R: Pass V: <b>CNR</b>	Not Recommended for Endorsement
3564	Medicare Spending Per Beneficiary – Post Acute Care Measure for Home Health Agencies	R: Pass V: Pass	R: Pass V: <b>CNR</b>	Not Recommended for Endorsement
3574	Medicare Spending Per Beneficiary (MSPB) Clinician	R: Pass V: Pass	R: <b>CNR</b> V: <b>Not Pass</b>	Not Recommended for Endorsement
2496	Standardized Readmission Ratio (SRR) for dialysis facilities	R: CNR V: Not Pass	R: <b>Pass</b> V: Not Pass	Not Recommended for Endorsement
3566	Standardized Ratio of Emergency Department Encounters Occurring Within 30 Days of Hospital Discharge (ED30) for Dialysis Facilities	R: Not Pass V: Pass	R: <b>Pass</b> V: Pass	Recommended for Endorsement

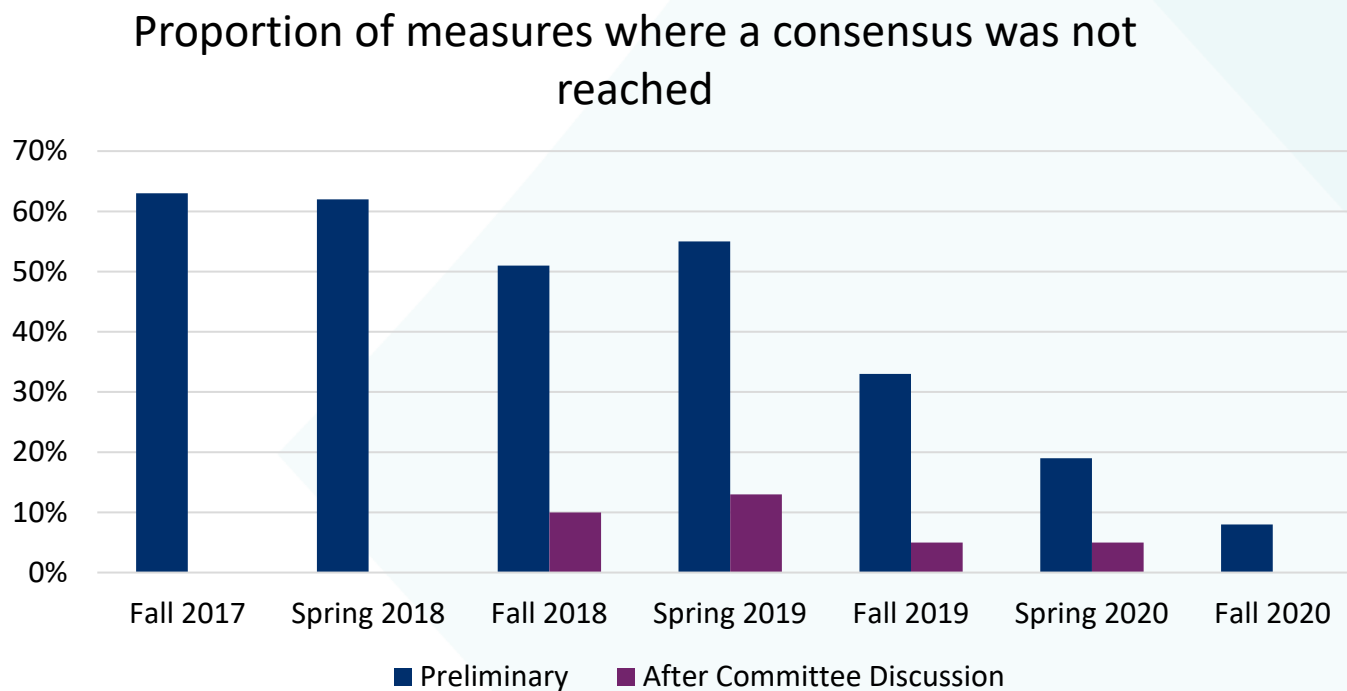
## Performance Metrics

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

\*Data for the Fall 2020 cycle are preliminary

\*\*These measures were sent to the Standing Committees

## Consensus Not Reached Statistics



# Fall 2020 Cycle Overview



## Fall 2020 Evaluation Cycle Statistics

- 25 complex measures assigned to the SMP
  - ▣ 6 new measures
- 3 subgroups of 8-9 Panel members with 8-9 measures
  - ▣ 20 passed reliability AND validity
  - ▣ 2 consensus not reached (CNR) on reliability or validity
  - ▣ 3 did not pass validity
  - ▣ 1 withdrawn after preliminary review
  - ▣ 8 slated for discussion
- Measure Types
  - ▣ Outcome: 19
  - ▣ Intermediate Clinical Outcome: 1
  - ▣ Composite: 3
  - ▣ PRO-PMs: 1
  - ▣ Cost/Resource Use: 1

# Process Overview and Evaluation Reminders



## 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 or an inappropriate methodology was applied

### ■ Insufficient

- ▣ 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 Panel 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** data element testing **OR** score-level testing
    - » We 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 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
  - ▣ “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 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 (including PRO-PMs)

- 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)



## 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 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

- The SMP previously 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 did not pass the SMP can be pulled by a standing committee member for further discussion and re-vote if it is an ***eligible*** measure



## Committee Consideration of Measures that Do Not Pass the SMP

- Any measure pulled by a Standing Committee member will be discussed
- Some measures may be eligible for re-vote by the Standing Committee
- 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 re-vote:
    - » 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 12:15 EST**

# Fall 2020 Measure Evaluation



## 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 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



# Voting Test



## #0141: Patient Fall Rate

- Subgroup 3
- Preliminary Voting Result:
  - ▣ Reliability: H-0; M-7; L-0; I-1 [*Pass*]
  - ▣ Validity: H-0; M-2; L-6; I-0 [*Does Not Pass*]
- Lead Discussants: Sean O'Brien, Terri Warholak
- Measure Developer: American Nurses Association
- Discussion Guide page 16
- For SMP discussion:
  - ▣ What concerns to you have regarding the lack of analysis for the exclusion of certain nursing unit types?
  - ▣ What concerns do you have regarding the negative correlations found?
  - ▣ What concerns do you have regarding the measure's ability to identify meaningful differences?
  - ▣ Considerations for the Standing Committee
    - » Are there concerns regarding the social risk factor assessment?



## #0202: Falls with Injury

- Subgroup 3
- Preliminary Voting Result:
  - ▣ Reliability: H-0; M-7; L-0; I-1 [*Pass*]
  - ▣ Validity: H-0; M-1; L-5; I-2 [*Does Not Pass*]
- Lead Discussants: John Bott, Sherrie Kaplan
- Measure Developer: American Nurses Association
- Discussion Guide page 18
- For SMP discussion:
  - ▣ What concerns do you have regarding the lack of analysis for the exclusion of certain nursing unit types?
  - ▣ What concerns do you have regarding the measure's ability to identify meaningful differences?
  - ▣ Considerations for the Standing Committee
    - » Are there concerns regarding the social risk factor assessment?



## #0505 Hospital 30-Day, All-Cause, Risk-Standardized Readmission Rate Following Acute Myocardial Infarction Hospitalization

- Subgroup 1
- Preliminary Voting Result:
  - ▣ Reliability: H-1; M-4; L-4; I-0 [*Consensus Not Reached*]
  - ▣ Validity: H-0; M-8; L-1; I-0 [*Pass*]
- Lead Discussants: Bijan Borah, Christie Teigland
- Measure Developer: Centers for Medicare & Medicaid Services and Yale CORE
- Discussion Guide page 6
- For SMP discussion:
  - ▣ What information was the primary influencer for your reliability vote?
  - ▣ How do the reliability statistics for NQF 0505 compare to other similar measures in the All-Cause Admissions and Readmissions portfolio?



## #3599: Pediatric Asthma Emergency Department Use

- Subgroup 2
- Preliminary Voting Result:
  - ▣ Reliability: H-2; M-5; L-0; I-1 [*Pass*]
  - ▣ Validity: H-0; M-4; L-3; I-1 [*Consensus Not Reached*]
- Lead Discussants: Daniel Deutscher, Susan White
- Measure Developer: University of California San Francisco
- Discussion Guide page 10
- For SMP discussion:
  - ▣ What are your concerns regarding missing data?
  - ▣ Do you have any concerns with the construct validity values provided?
  - ▣ Do you have any concerns regarding the risk adjustment method selected?
  - ▣ Considerations for the Standing Committee
    - » Are there concerns regarding the selection of social risk factors?

# Opportunity for Public Comment

# Adjourn

# Day 2: Welcome, Review of Agenda

## 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
Daniel Deutscher, PT, PhD	Sean O'Brien, PhD
Lacy Fabian, PhD	Jennifer Perloff, PhD
Marybeth Farquhar, PhD, MSN, RN	Patrick Romano, MD, MPH
Jeffrey Geppert, EdM, JD	Sam Simon, PhD
Laurent Glance, MD	Alex Sox-Harris, PhD, MS
Joseph Hyder, MD	Christie Teigland, PhD, Co-Chair
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



## Agenda for Day 2

- Welcome
- Discussion of Criteria and Evaluation Guidance
- Opportunity for Public Comment
- Next Steps
- Adjourn

# Criteria Recommendations and Evaluation Guidance



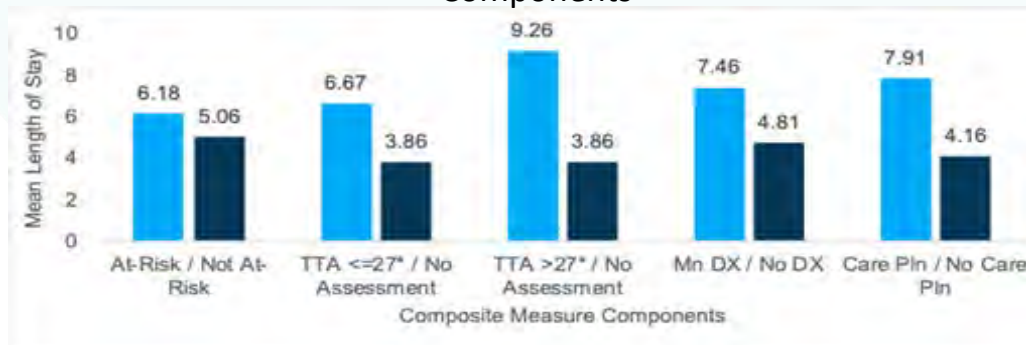
## Key Topics Identified During the Fall 2020 Review Cycle

- Validity
  - ▣ Correlations used to claim validity
- Reliability
  - ▣ Acceptable thresholds for reliability, and comparison of reliability statistics

## Correlations Used to Claim Validity

- Process-outcome correlations are used to establish validity
- Construction of the composite was associated with worse outcomes, not better outcomes
  - ▣ This is relatively uncommon
- What guidance does the panel have for measures that use a correlation to establish validity?
- Example: 3592

Length of Stay Predictability of Malnutrition Composite Measure Components



\*TTA = Time to Assessment; The timing was tested at the median split for all hospitals included in the testing dataset.

# Acceptable Thresholds for Reliability

- Acceptable Thresholds
  - ▣ Differing threshold values within the literature (Landis, Adams, others)
  - ▣ What is the appropriate threshold? How would the evaluation ratings be assigned based on the threshold?

Landis <sup>1</sup>	Adams <sup>2</sup>
<p>&lt; 0 – Less than chance agreement;            0 – 0.2 Slight agreement;            0.21 – 0.39 Fair agreement;            0.4 – 0.59 Moderate agreement;            0.6 – 0.79 Substantial agreement;            0.8 – 0.99 Almost Perfect agreement;            and            1 Perfect agreement</p>	<p>0.5- difficult to detect differences between physicians            0.7-start to see differences between some physicians and the mean            0.9-start to see significant differences between pairs of physicians.</p>

Reference:

1. Landis J, Koch G. The measurement of observer agreement for categorical data, *Biometrics* 1977;33:159-174.
2. Estimating Reliability and Misclassification in Physician Profiling. John L. Adams, Ateev Mehrotra, Elizabeth A. McGlynn, RAND 2010
3. Koo & Li. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research *J Chiropr Med.* 2016



## Comparison of Reliability Statistics

- Reliability statistics for measures can be similar across similar measures
  - ▣ Split Sample Reliability
  - ▣ Median SNR
- How do panel members view their rationale behind a high/moderate vote or a low/insufficient vote for similar reliability statistics?
- Examples for Reference: 1891, 0505, 2515, 0230

## Comparison of Reliability Statistics

NQF ID	Measure Title	Split Sample Reliability	25th percentile SNR	Median SNR	Group Reliability Vote
3597	Clinician-Group Risk-Standardized Acute Hospital Admission Rate for Patients with Multiple Chronic Conditions under the Merit-based Incentive Payment System	-	0.24 (0.683 when apply patient threshold)	0.58 (0.809 when apply patient threshold)	7/8=87.5%
1891	Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization	0.406	0.25	0.43	6/9=66.7%
0505	Hospital 30-day all-cause risk-standardized readmission rate (RSRR) following acute myocardial infarction (AMI) hospitalization.	0.424	0.33	0.51	5/9=55.6%
0230	Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute myocardial infarction (AMI) hospitalization	0.428	0.41	0.59	5/8=62.5%
2515	Hospital 30-day, all-cause, unplanned, risk-standardized readmission rate (RSRR) following coronary artery bypass graft (CABG) surgery	0.436	0.45	0.60	8/9=88.9%

Green text used to note where the measure passed. Grey text used to note where consensus was not reached

## Comparison of Reliability Statistics

NQF ID	Measure Title	Split Sample Reliability	25th percentile SNR	Median SNR	Group Reliability Vote
1551	Hospital-level 30-day risk-standardized readmission rate (RSRR) following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA)	0.454	0.58	0.77	7/8=87.5%
1893	Hospital 30-Day, all-cause, risk-standardized mortality rate (RSMR) following chronic obstructive pulmonary disease (COPD) hospitalization	0.477	0.54	0.72	6/7=85.7%
1623	Bereaved Family Survey	0.520	-	-	8/8=100%
1550	Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA)	0.524	0.74	0.87	8/8=100%
0506	Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization	0.544	0.34	0.56	8/9=88.9%
0330	Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization	0.587	0.31	0.57	7/8=87.5%

Green text used to note where the measure passed. Grey text used to note where consensus was not reached

## Comparison of Reliability Statistics

NQF ID	Measure Title	Split Sample Reliability	25th percentile SNR	Median SNR	Group Reliability Vote
2888	Risk-Standardized Acute Admission Rates for Patients with Multiple Chronic Conditions	-	0.94	0.96	8/8=100%
2158	Medicare Spending Per Beneficiary (MSPB) - Hospital	0.83 (2018 split-sample) 0.79 (2017 & 2018 split-sample)	0.91	0.96	7/7=100%
3596	Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute ischemic stroke hospitalization with claims-based risk adjustment for stroke severity	-	0.59	0.75	8/8=100%
0229	Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following heart failure (HF) hospitalization	0.632	0.58	0.79	8/8=100%
0468	Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following pneumonia hospitalization	0.668	0.59	0.78	8/8=100%
3599	Pediatric Asthma Emergency Department Use	0.72 (MA) 0.86 (CA)	-	-	7/8=87.5%

Green text used to note where the measure passed. Grey text used to note where consensus was not reached

## Comparison of Reliability Statistics

NQF ID	Measure Title	Split Sample Reliability	25th percentile SNR	Median SNR	Group Reliability Vote
0531	Patient Safety and Adverse Events Composite	0.74	-	-	7/8=87.5%
3235	Hospice and Palliative Care Composite Process Measure—Comprehensive Assessment at Admission	0.86	-	3.55	8/8=100%

# Opportunity for Public Comment

# Next Steps

## Next Steps

- Measure submission deadlines: November 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 January-February timeframe
  - ▣ CSAC decisions expected in June 2021
- Next Intent to Submit deadline: January 5, 2021

## 2020 SMP Meetings

Meeting Date	Tentative Topic/Activity
December 8 - 1-3PM ET	Discussion of Updating Guidance on Scientific Acceptability Criteria



## Project Contact Info

- Email: [MethodsPanel@qualityforum.org](mailto: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 (*new*):  
<https://share.qualityforum.org/portfolio/ScientificMethodsPanel/SitePages/Home.aspx>

# Adjourn

**THANK YOU.**

**NATIONAL QUALITY FORUM**

<http://www.qualityforum.org>