NATIONAL QUALITY FORUM

National Consensus Standards for All-Cause Admissions and Readmissions
Standing Committee Orientation

Erin O'Rourke
Taroon Amin
Kate McQueston
Miranda Kuwahara

January 29, 2018

Welcome

Agenda for the Call

- Standing Committee roll call
- CDP redesign overview
- Changes to NQF evaluation criteria
- Overview of NQF's portfolio of readmissions measures
- Overview of social risk
- Overview of eMeasure Approval for Trial Use
- Public comment
- Next steps

NQF Staff

- Project staff
  - *Erin O'Rourke, Senior Director*
  - *Kate McQueston, Senior Project Manager*
  - *Miranda Kuwahara, Project Manager*
  - *Taroon Amin, Consultant*
- NQF Quality Measurement leadership staff
  - *Elisa Munthali, Vice President*

Standing Committee

- John Bulger, DO, MBA (co-chair)
- Cristie Travis, MSHA (co-chair)
- Katherine Auger, MD, MSc
- Frank Briggs, PharmD, MPH
- Jo Ann Brooks, PhD, RN
- Mae Centeno, DNP, RN, CCRN, CCNS, ACNS-BC
- Helen Chen, MD
- Susan Craft, RN
- William Wesley Fields, MD, FACEP
- Steven Fishbane, MD
- Paula Minton Foltz, RN, MSN
- Brian Foy, MHA
- Laurent Glance, MD
- Anthony Grigonis, PhD
- Bruce Hall, MD, PhD, MBA
- Leslie Kelly Hall
- Paul Heidenreich, MD, MS, FACC, FAHA
- Karen Joynt Maddox, MD, MPH
- Sherrie Kaplan, PhD
- Keith Lind, JD, MS, BSN
- Paulette Niewczyk, PhD, MPH
- Carol Raphael, MPA
- Mathew Reidhead, MA
- Pamela Roberts, PhD, MSHA, ORT/L, SCFES, FAOTA, CPHQ
- Derek Robinson, MD, MBA, FACEP, CHCQM
- Thomas Smith, MD, FAPA

Overview of CDP Redesign

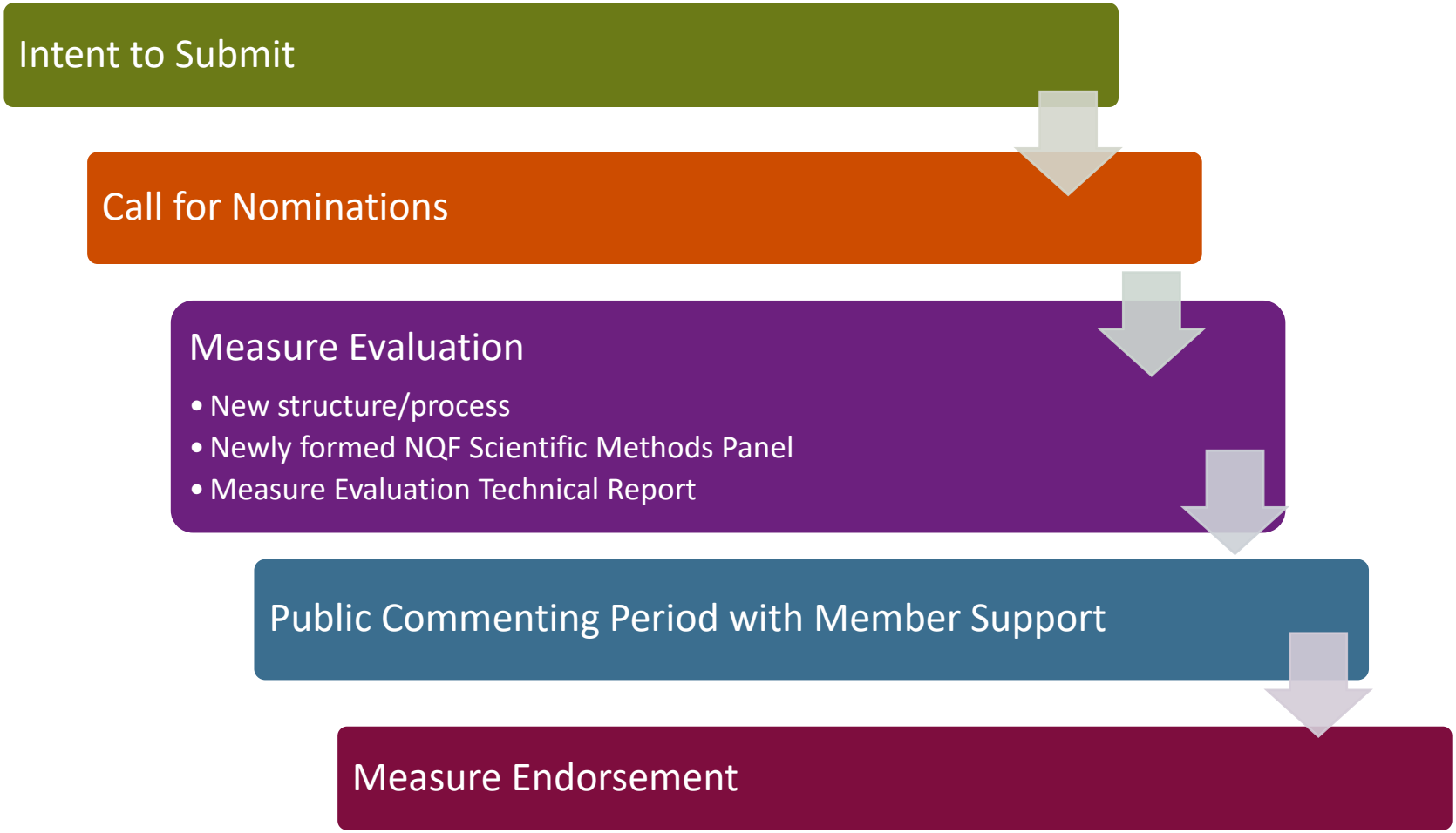
The National Quality Forum: A Unique Role

Established in 1999, NQF is a nonprofit, nonpartisan, membership-based organization that brings together public and private sector stakeholders to reach consensus on healthcare performance measurement. The goal is to make healthcare in the U.S. better, safer, and more affordable.

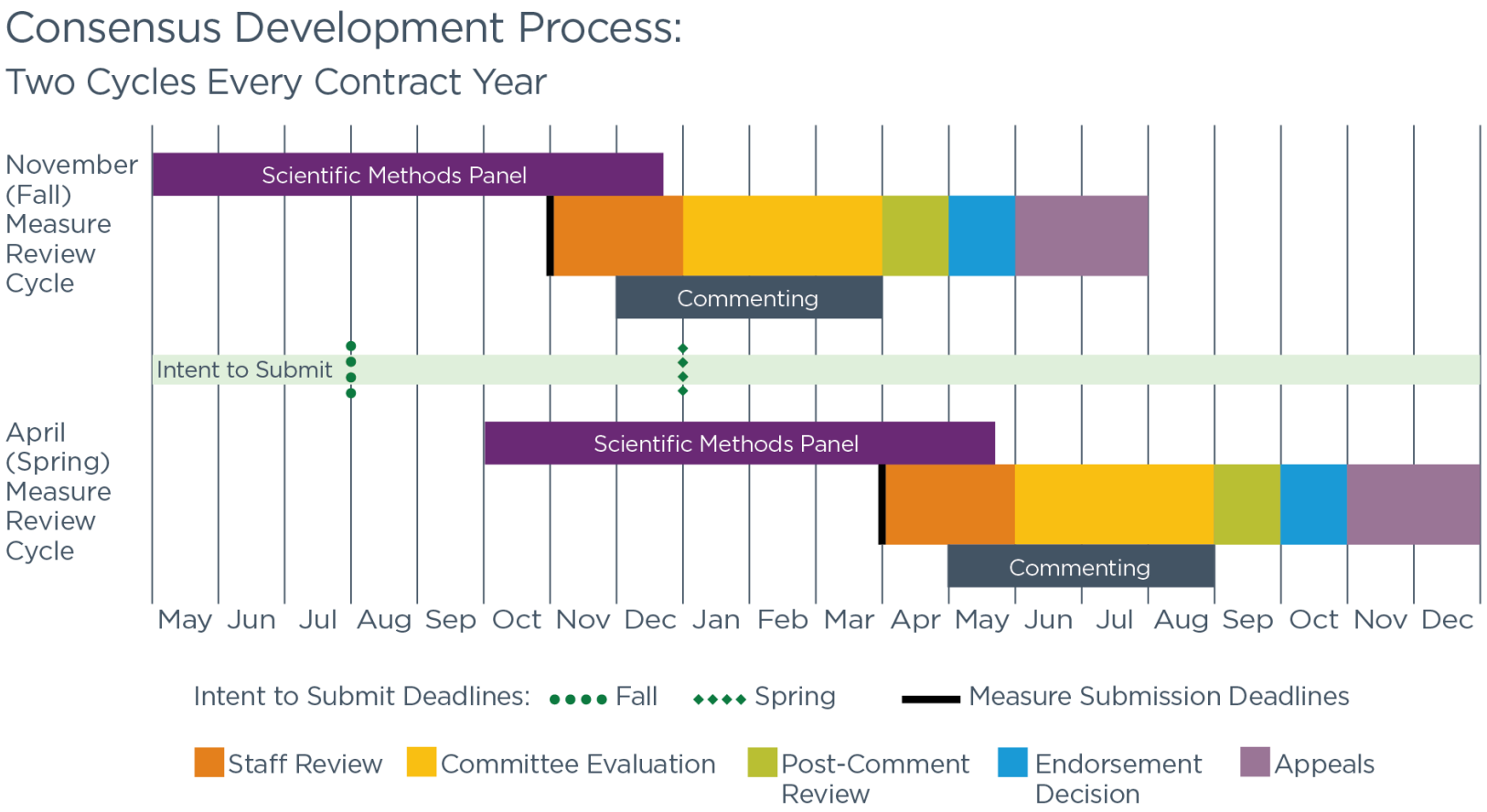
Mission: To lead national collaboration to improve health and healthcare quality through measurement

- An Essential Forum
- Gold Standard for Quality Measurement
- Leadership in Quality

NQF Consensus Development Process (CDP): 6 Steps for Measure Endorsement



Measure Review: Two Cycles Per Year



15 New Measure Review Topical Areas

	All Cause Admission/ Readmissions	Behavioral Health	
Cancer	Cardiovascular	Care Coordination	Infectious Disease
Cost and Resource Use	Endocrine	Eyes, Ears, Nose and Throat Conditions	Palliative and End-of Life Care
Gastrointestinal	Genitourinary	Health and Well Being	Musculoskeletal
Neurology	Patient Safety	Pediatrics	Perinatal
Person and Family-Centered Care	Pulmonary and Critical Care	Renal	Surgery

All Cause Admission/ Readmissions	Behavioral Health & Substance Use	Cancer
Cardiovascular	Cost and Efficiency[A]	Geriatric and Palliative Care[B]
Neurology	Patient Experience & Function	Patient Safety[C]
Pediatrics	Perinatal and Women's Health	Prevention and Population Health[D]
Primary Care and Chronic Illness	Renal	Surgery

☐ Denotes expanded topic area

[A] Cost & Efficiency will include efficiency-focused measures from other domains
[B] Geriatric & Palliative Care includes pain-focused measures from other domains
[C] Patient Safety will include acute infectious disease and critical measures
[D] Prevention and Population Health is formerly Health and Well Being

Measure Review

- NEW!! Scientific Methods Panel
  - *Evaluate Scientific Acceptability of Complex Measures*
  - *Serve in Advisory Capacity to NQF*

- Methods Review by Staff
  - *NQF will continue to provide preliminary analysis, review for noncomplex measures*

Role of Methods Panel

- Scientific Methods Panel created to ensure higher-level and more consistent reviews of the scientific acceptability of measures
- The Methods Panel is charged with:
  - *Conducting evaluation of complex measures for the Scientific Acceptability criterion, with a focus on reliability and validity analyses and results*
  - *Serve in advisory capacity to NQF on methodologic issues, including those related to measure testing, risk adjustment, and measurement approaches.*
- The method panel review will help inform the standing committee's endorsement decision. The panel will not render endorsement recommendations.

Measure Review

Complex Measures

- Outcome measures, including intermediate clinical outcomes
- Instrument-based measures (e.g., PRO-PMs)
- Cost/resource use measures
- Efficiency measures (those combining concepts of resource use and quality)
- Composite measures

Noncomplex Measures

- Process measures
- Structural measures
- Previously endorsed complex measures with no changes/updates to the specifications or testing

Role of the Expert Reviewers

- In 2017, NQF executed a CDP redesign that resulted in restructuring and reducing the number of topical areas as well as a bi-annual measure review process.

- Given these changes, there is a need for diverse yet specific expertise to support longer and continuous engagement from standing committees.

Role of the Expert Reviewers

- The expert reviewer pool serves as an adjunct to NQF standing committees to ensure broad representation and provide technical expertise when needed
- Expert reviewers will provide expertise as needed to review measures submitted for endorsement consideration by:
  - *Replacing an inactive committee member;*
  - *Replacing a committee members whose term has ended; or*
  - *Providing expertise that is not currently represented on the committee.*
- Expert reviewers may also:
  - *Provide comments and feedback on measures throughout the measure review process*
  - *Participate in strategic discussions in the event no measures are submitted for endorsement consideration*

Measure Evaluation Report

"Old" Technical Report	New Technical Report
Executive Summary	Executive Summary
Introduction	Measure Evaluation
NQF Portfolio of Performance Measures	Details of Measure Evaluation (Appendix)
Measure Evaluation	Use in Federal Programs (Appendix)
Details of Measure Evaluation (Appendix)	Standing Committee and NQF Staff (Appendix)
Use in Federal Programs (Appendix)	Measure Specifications (Appendix)
Standing Committee and NQF Staff (Appendix)	Related and Competing Measures (Side-by-Side Table) (Appendix)
Measure Specifications (Appendix)	
Related and Competing Measures (Side-by-Side Table) (Appendix)	

Public Commenting Period with Member Support

- Extended opportunity for public and NQF member commenting
- 16+ week commenting period
  - » Comments can be submitted at any time throughout this period

Consensus Standards Approval Committee (CSAC)

- NQF Board-approved advisory committee's role remains the same
  - *Provide guidance to NQF leadership regarding enhancements to the CDP*
  - *Maintains Measure Evaluation Criteria*
  - *Renders Final Endorsement Decision*

Measure Appeals

- 30-Day appeals period remains the same
- Any interested party may file an appeal on an endorsed measure during this period
- The Appeals Board will review all appeals submitted to NQF
  - *The five-member Appeals Board is composed of NQF Board members and former CSAC and/or committee members*
  - *The Appeals Board adjudicates appeals to measure endorsement decisions without a review by the CSAC - the decision will be final*

Enhanced Training and Education

All Stakeholders

- Public
- NQF Members
- Measure Developers
- Standing Committee Members & Co-Chairs
- NQF Staff

Questions?

2017 Changes to NQF Evaluation Criteria and Guidance

NQF Measure Evaluation Criteria for Endorsement

NQF endorses measures for accountability applications (public reporting, payment programs, accreditation, etc.) as well as quality improvement.

- Standardized evaluation criteria
- Criteria have evolved over time in response to stakeholder feedback
- The quality measurement enterprise is constantly growing and evolving—greater experience, lessons learned, expanding demands for measures—the criteria evolve to reflect the ongoing needs of stakeholders

Major Endorsement Criteria (page 28)

- **Importance to measure and report**: Goal is to measure those aspects with greatest potential of driving improvements; if not important, the other criteria are less meaningful (must-pass)
- **Reliability and Validity-scientific acceptability of measure properties**: Goal is to make valid conclusions about quality; if not reliable and valid, there is risk of improper interpretation (must-pass)
- **Feasibility**: Goal is to, ideally, cause as little burden as possible; if not feasible, consider alternative approaches
- **Usability and Use**: Goal is to use for decisions related to accountability and improvement; if not useful, probably do not care if feasible
- **Comparison** to related or competing measures

Evidence (subcriterion 1a): Strengthen requirements for outcome measures

- **Revised criterion**
  - *For all outcomes:* **Empirical data** *demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service.* **If not available, wide variation in performance** *can be used as evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.*
  - **For measures derived from patient report**, *evidence should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.*
    - » Additional guidance: Examples of such evidence include, but are not limited to, patient input in the development of the instrument, survey, or tool; focus group input regarding the value of the performance measure derived from the instrument/survey/tool.

Evidence (subcriterion 1a): Additional guidance for instrument-based measures

- Current requirements for structure and process measures (i.e., a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence that the measured structure/process leads to a desired health outcome) **also apply to patient-reported structure/process measures**.

Evidence (subcriterion 1a): Additional guidance for thresholds and timeframes

- Evidence for specific timeframes or thresholds included in a measure **should be presented**. If evidence is limited, then literature regarding standard norms would be **considered**.

Performance Gap (subcriterion 1b): Additional guidance

- For **maintenance measures**
  - *Measure stewards are expected to provide current performance data. If limited data are available (e.g., use is voluntary),* **data from the literature** *can be considered.*

Reliability (subcriterion 2a): Potential for additional guidance

- Establishing thresholds for testing results
  - *NQF will ask our newly-formed Scientific Methods Panel for input on norms and/or rules of thumb*

Validity (subcriterion 2b): Remove "evidence aligns with specifications"

- Subcriterion 2b.1 now removed
  - *The measure specifications are consistent with the evidence presented to support the focus of measurement under criterion 1a. The measure is specified to capture the most inclusive target population indicated by the evidence, and exclusions are supported by the evidence.*

- Evidence now considered as part of subcriterion 1a

Validity (subcriterion 2b): Strengthen guidance for face validity

- **Revised guidance**
  - **Empirical validity testing is expected at time of maintenance review**; *if not possible, justification is required.*
  - *Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and 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.**

Validity (subcriterion 2b): Exclusions criterion re-worded

- **Revised criterion**
  - *Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure*
    - *Previous wording:* Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion
- Potential for updated guidance
  - *Will ask NQF's newly-formed Scientific Methods Panel for input on what might be sufficient frequency and how to handle non-uniformity of frequency across providers*

Validity (subcriterion 2b): Missing data requirement (2b.6) applicable to all measures

- Revised criterion
  - *Analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.*

    - *Previous criterion*: For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Usability and Use: Now partly must-pass for maintenance measures

- Use: Change to must-pass for **maintenance** measures
  - *In use in accountability program within 3 years and publicly reported within 6 years*
  - *Measure has been vetted by those being measured or others*
- Usability*: still not must-pass
  - *Demonstrated improvement*
  - *Benefits outweigh evidence of unintended negative consequences to patients*

* Information for these two subcriteria may be obtained via literature, feedback to NQF, and from developers during the submission process.

Updated guidance for measures that use ICD-10 coding: Fall 2017 and 2018

- Gap can be based on literature and/or data based on ICD-9 or ICD-10 coding
- Submit updated ICD-10 reliability testing if available; if not, testing based on ICD-9 coding will suffice
- Submit updated validity testing
  - *Submit updated empirical validity testing on the ICD-10 specified measure,* **if available**
  - **OR** *face validity of the ICD-10 coding scheme* **plus face validity** *of the measure score as an indicator of quality*
  - **OR** *face validity of the ICD-10 coding scheme* **plus score-level** *empirical validity testing based on ICD-9 coding*
  - **OR** *face validity of the ICD-10 coding scheme* **plus data element** *level validity testing based on ICD-9 coding, with face validity of the measure score as an indicator of quality due at* **annual update**

Best practices for ICD-10 coding

- Use team of **clinical and coding experts** to identify specific areas where questions of clinical comparability exist, evaluate consistency of clinical concepts, and ensure appropriate conversion
- Determine **intent**
- If desired, use appropriate conversion tool
  - *Not required, but also* **not sufficient by itself**
  - *If using conversion tool, consider both* **forward and backward mapping**

Best practices for ICD-10 coding (continued)

- Assess for material change, if possible
  - *Assess extent to which the population identified with the new code set overlaps with that identified in the old code set*
  - *Assess whether the conversion results in rates that are similar within defined tolerances; options include:*
    - » Test using dual-coded data if possible **OR**
    - » Face validity (using the above code-conversion process, including use of clinical/coding experts) **OR**
    - » Criterion validity (if dual-coded data not available) **OR**
    - » Consistency across time (pre/post conversion)
- Solicit stakeholder comments

eMeasures

- "Legacy" eMeasures
  - *Beginning September 30, 2017 all respecified measure submissions for use in federal programs will be required to the same evaluation criteria as respecified measures – the "**BONNIE testing only**"* **option will no longer meet endorsement criteria**

- For all eMeasures: Reliance on data from structured data fields is expected; otherwise, unstructured data must be shown to be both reliable and valid

NQF All-Cause Admissions and Readmissions Portfolio: All Cause/All Condition Population Based Measures

Measure Number	Measure Title
1768	Plan All-Cause Readmissions [NCQA]
2504	30-Day Rehospitalizations Per 1000 Medicare Fee-for-Service (FFS) Beneficiaries [CMS]
2503	Hospitalizations Per 1000 Medicare Fee-for-Service (FFS) Beneficiaries [Colorado Foundation for Medical Care]
2888	Risk-Standardized Acute Admission Rates for Patients with Multiple Chronic Conditions [Yale/CORE]

NQF All-Cause Admissions and Readmissions Portfolio: Condition-Specific Admissions Measures

Measure Number	Measure Title
0272	Diabetes Short-Term Complications Admission Rate (PQI 1) [AHRQ]
0273	Perforated Appendix Admission Rate (PQI 2) [AHRQ]
0274	Diabetes Long-Term Complications Admission Rate (PQI 3) [AHRQ]
0277	Heart Failure Admission Rate (PQI 8) [AHRQ]
0279	Bacterial Pneumonia Admission Rate (PQI 11) [AHRQ]
0280	Dehydration Admission Rate (PQI 10) [AHRQ]
0281	Urinary Tract Infection Admission Rate (PQI 12) [AHRQ]
0283	Asthma in Younger Adults Admission Rate (PQI 15) [AHRQ]
0638	Uncontrolled Diabetes Admission Rate (PQI 14) [AHRQ]
0727	Gastroenteritis Admission Rate (pediatric) [AHRQ]
0728	Asthma Admission Rate (Pediatric) [AHRQ]
2886	Risk-Standardized Acute Admission Rates for Patients with Heart Failure [Yale/CORE]
2887	Risk-Standardized Acute Admission Rates for Patients with Diabetes [Yale-CORE]

NQF All-Cause Admissions and Readmissions Portfolio: Hospital All-Cause/All-Condition Readmission Measures

Measure Number	Measure Title
0335	PICU Unplanned Readmission Rate [Virtual PICU Systems, LLC]
1789	Hospital-Wide All-Cause Unplanned Readmission Measure (HWR) [CMS]
2393	Pediatric All-Condition Readmission Measure [Center of Excellence for Pediatric Quality Measurement]
2879	Hybrid Hospital-Wide Readmission Measure with Claims and Electronic Health Record Data [Yale/CORE]

NQF All-Cause Admissions and Readmissions Portfolio: Cardiovascular Condition-Specific Hospital Readmission Measures

Measure Number	Measure Title
0330	Hospital 30-Day, All-Cause, Risk-Standardized Readmission Rate Following Heart Failure Hospitalization for Patients 18 and Older [CMS]
0505	Thirty-Day All-Cause Risk Standardized Readmission Rate Following Acute Myocardial Infarction (AMI) Hospitalization [CMS]
0695	Hospital 30-Day Risk-Standardized Readmission Rates following Percutaneous Coronary Intervention (PCI) [American College of Cardiology]
2514	Risk-Adjusted Coronary Artery Bypass Graft (CABG) Readmission Rate [STS]
2515	Hospital 30-Day, All-Cause, Unplanned, Risk-Standardized Readmission Rate (RSRR) Following Coronary Artery Bypass Graft (CABG) Surgery [CMS]
2880	Excess Days in Acute Care (EDAC) After Hospitalization for Heart Failure [Yale/CORE]
2881	Excess Days in Acute Care (EDAC) After Hospitalization for Acute Myocardial Infarction (AMI) [Yale/CORE]

NQF All-Cause Admissions and Readmissions Portfolio: Pulmonary Condition-Specific Hospital Readmission Measures

Measure Number	Measure Title
0506	Thirty-Day All-Cause Risk Standardized Readmission Rate Following Pneumonia Hospitalization. [CMS]
1891	Hospital 30-Day, All-Cause, Risk-Standardized Readmission Rate (RSRR) following Chronic Obstructive Pulmonary Disease (COPD) Hospitalization [CMS]
2414	Pediatric Lower Respiratory Infection Readmission Measure [Center of Excellence for Pediatric Quality Measurement]
2882	Excess Days in Acute Care (EDAC) After Hospitalization for Pneumonia

NQF All-Cause Admissions and Readmissions Portfolio: Surgical Condition-Specific Hospital Readmission Measures

Measure Number	Measure Title
2513	Hospital 30-Day All-Cause Risk-Standardized Readmission Rate (RSRR) following Vascular Procedures [CMS]
1551	Hospital-Level 30-Day, All-Cause Risk-Standardized Readmission Rate (RSRR) Following Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) [CMS]

NQF All-Cause Admissions and Readmissions Portfolio: Setting-Specific Readmission Measures

Measure Number	Measure Title
0171	Acute Care Hospitalization During the First 60 Days of Home Health (Risk-Adjusted) [CMS]
0173	Emergency Department Use without Hospitalization During the First 60 Days of Home Health (Risk Adjusted)
1463	Standardized Hospitalization Ratio for Dialysis Facilities [CMS]
2375	PointRight OnPoint-30 Skilled Nursing Facility Rehospitalizations [AHCA]
2510	Skilled Nursing Facility 30-Day All-Cause Readmission Measure (SNFRM) [RTI]
2380	Rehospitalization During the First 30 Days of Home Health [CMS]
2505	Emergency Department Use without Hospital Readmission During the First 30 Days of Home Health [CMS]
2512	All-Cause Unplanned Readmission Measure for 30 Days Post Discharge from Long-Term Care Hospitals (LTCHs) [CMS]
2502	All-Cause Unplanned Readmission Measure for 30 Days Post Discharge from Inpatient Rehabilitation Facilities [CMS]
2496	Standardized Readmission Ratio (SRR) for dialysis facilities [CMS]
2539	Facility 7-Day Risk-Standardized Hospital Visit Rate after Outpatient Colonoscopy [CMS]
2827	PointRight® Pro Long Stay(TM) Hospitalization Measure (PointRight)
2858	Discharge to Community [ACHA]
2860	Thirty-Day All-Cause Unplanned Readmission Following Psychiatric Hospitalization in an Inpatient Psychiatric Facility (IPF)

Social Risk Overview

Background

- NQF conducted a two-year trial period from 2015-2017. During this time, adjustment of measures for social risk factors **was no longer prohibited**
- The NQF Board of Directors reviewed the results of the trial period and determined there was a need to launch a new social risk initiative
- As part of the Equity Program, NQF will continue to explore the need to adjust for social risk
- Each measure must be assessed individually to determine if SDS adjustment is appropriate (included as part of **validity** subcriterion)
- The Standing Committee will continue to evaluate the measure as a whole, including the appropriateness of the risk adjustment approach used by the measure developer
- Efforts to implement SDS adjustment may be constrained by data limitations and data collection burden

Standing Committee Evaluation

- The Standing Committee will be asked to consider the following questions:
  - *Is there a conceptual relationship between the SDS factor and the measure focus?*
  - *What are the patient-level sociodemographic variables that were available and analyzed during measure development?*
  - *Does empirical analysis (as provided by the measure developer) show that the SDS factor has a significant and unique effect on the outcome in question?*
  - *Does the reliability and validity testing match the final measure specifications?*

Questions?

eMeasure Approval for Trial Use

eMeasure Approval for Trial Use

Requirements

- eMeasure submissions only
  - *HQMF specified, use QDM, use value sets published in the VSAC, as verified by staff review*
- Meet NQF criteria, except testing criteria
  - *Important to measure*
  - *Feasibility*
    - » specifically eMeasure Feasibility Criteria which gauges "implementation readiness"
  - *Plan for Use*
  - *Harmonization*
- **Approval for Trial Use is <u>not</u> NQF endorsement**
  - *Approval for further testing*
  - *3-year window to bring back testing for endorsement*

Questions?

Public Comment

Next Steps

Next Steps

- In-Person/Web Meetings
  - *[NQF] Readmissions Off Cycle Webinar #2: Feedback on SES Annual Update for Readmissions Measures, SES Trial 2.0, and Introduction to the Equity Program*
    - » Tuesday February 6, 2018

Project Contact Info

- Email: readmissions@qualityforum.org
- NQF Phone: 202-783-1300
- Project page: http://www.qualityforum.org/All-Cause_Admissions_and_Readmissions_2017.aspx
- SharePoint site: http://share.qualityforum.org/Projects/admissions_readmissions/SitePages/Home.aspx

Questions?

THANK YOU

Measure Evaluation Criteria Overview

NQF Measure Evaluation Criteria for Endorsement

NQF endorses measures for accountability applications (public reporting, payment programs, accreditation, etc.) as well as quality improvement.

- Standardized evaluation criteria
- Criteria have evolved over time in response to stakeholder feedback
- The quality measurement enterprise is constantly growing and evolving – greater experience, lessons learned, expanding demands for measures – the criteria evolve to reflect the ongoing needs of stakeholders

Major Endorsement Criteria *(page 28)*

- **Importance to measure and report**: Goal is to measure those aspects with greatest potential of driving improvements; if not important, the other criteria are less meaningful (**must-pass**)
- **Reliability and Validity-scientific acceptability of measure properties** : Goal is to make valid conclusions about quality; if not reliable and valid, there is risk of improper interpretation (**must-pass**)
- **Feasibility**: Goal is to, ideally, cause as little burden as possible; if not feasible, consider alternative approaches
- **Usability and Use**: Goal is to use for decisions related to accountability and improvement; if not useful, probably do not care if feasible
- **Comparison to related or competing measures**

Criterion #1: Importance to Measure and Report (page 30-39)

1. Importance to measure and report - Extent to which the specific measure focus is evidence-based and important to making significant gains in healthcare quality where there is variation in or overall less-than-optimal performance.

1a. Evidence: *the measure focus is evidence-based*

1b. Opportunity for Improvement: *demonstration of quality problems and opportunity for improvement, i.e., data demonstrating considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or disparities in care across population groups*

1c. Quality construct and rationale *(composite measures only)*

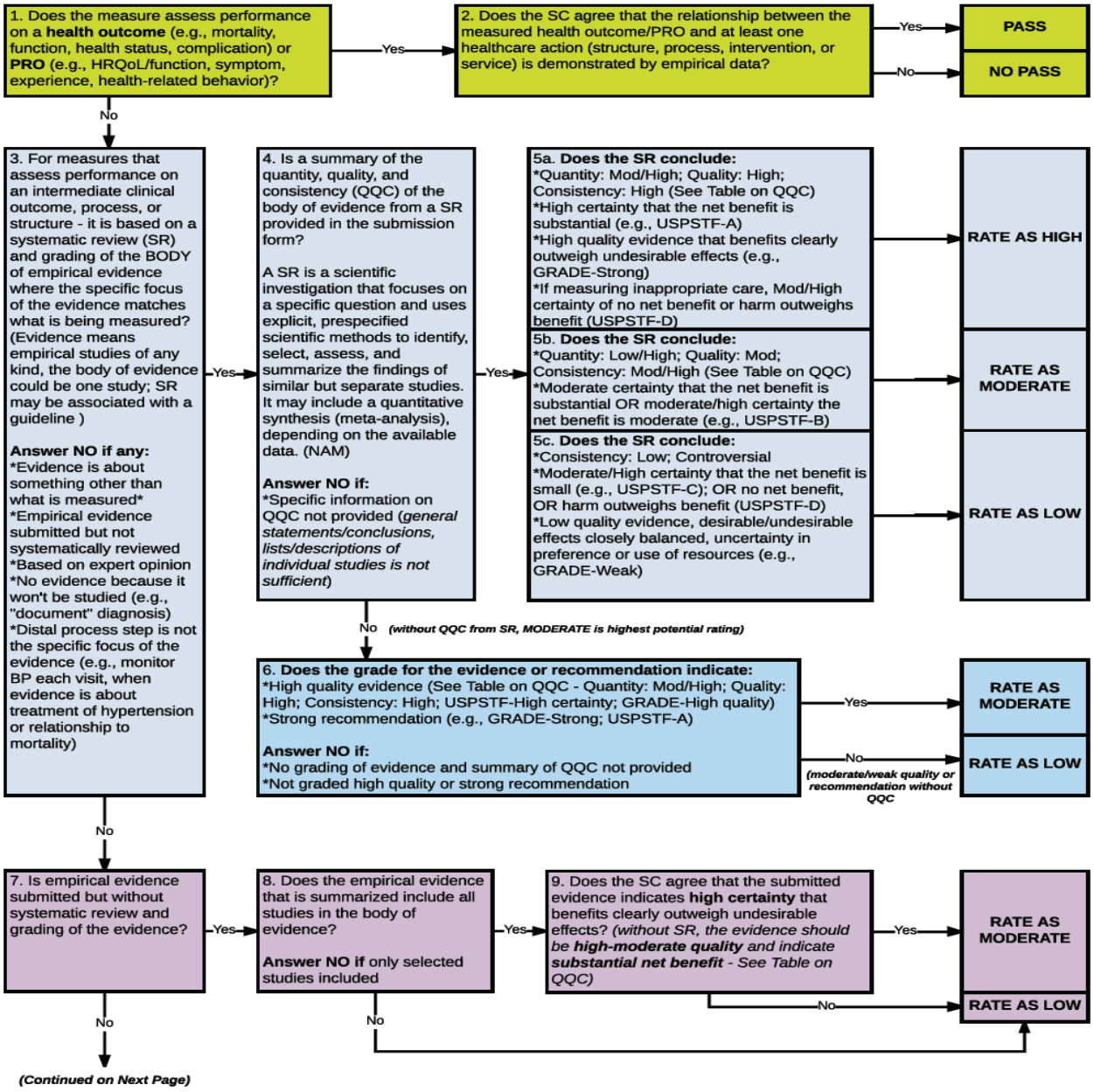
Subcriteron 1a: Evidence (page 31-37)

- Outcome measures
  - Empirical data *demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service. If not available, wide variation in performance can be used as evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.*

- Structure, process, intermediate outcome measures
  - *The quantity, quality, and consistency of the body of evidence underlying the measure should demonstrate that the measure focuses on those aspects of care known to influence desired patient outcomes*
    - » Empirical studies (expert opinion is not evidence)
    - » Systematic review and grading of evidence
      - *Clinical Practice Guidelines – variable in approach to evidence review*

- For measures derived from patient (or family/parent/etc.) report
  - *Evidence should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.*
  - *Current requirements for structure and process measures* also apply to patient-reported structure/process measures.

Rating Evidence: Algorithm #1 – page 34



Criterion #1: Importance to measure and report Criteria emphasis is different for new vs. maintenance measures

New measures	Maintenance measures
• Evidence – Quantity, quality, consistency (QQC) • Established link for process measures with outcomes	**DECREASED EMPHASIS**: Require measure developer to attest evidence is unchanged evidence from last evaluation; Standing Committee to affirm no change in evidence IF changes in evidence, the Committee will evaluate as for new measures
• Gap – opportunity for improvement, variation, quality of care across providers	**INCREASED EMPHASIS**: data on current performance, gap in care and variation

Criterion #2: Reliability and Validity–Scientific Acceptability of Measure Properties (page 39 -48)

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of health care delivery

2a. **Reliability (must-pass)**

2a1. Precise specifications including exclusions

2a2. Reliability testing—data elements or measure score

2b. **Validity (must-pass)**

2b1. Validity testing—data elements or measure score

2b2. Justification of exclusions—relates to evidence

2b3. Risk adjustment—typically for outcome/cost/resource use

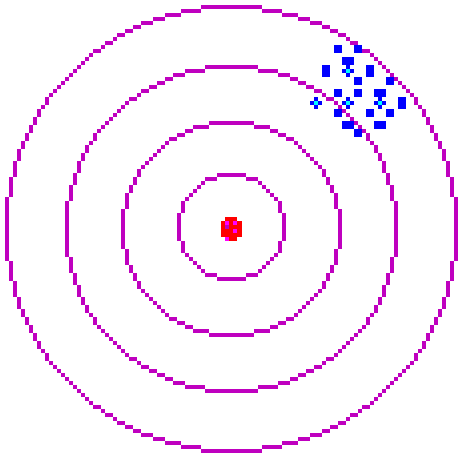
2b4. Identification of differences in performance

2b5. Comparability of data sources/methods

2b6. Missing data

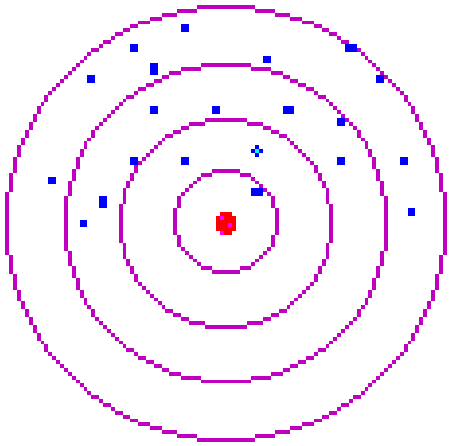
Reliability and Validity (page 40)

Assume the center of the target is the true score…



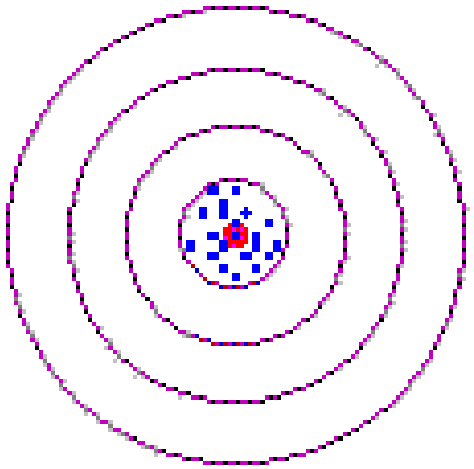
Reliable Not Valid

Consistent, but wrong



Neither Reliable Nor Valid

Inconsistent & wrong



Both Reliable And Valid

Consistent & correct

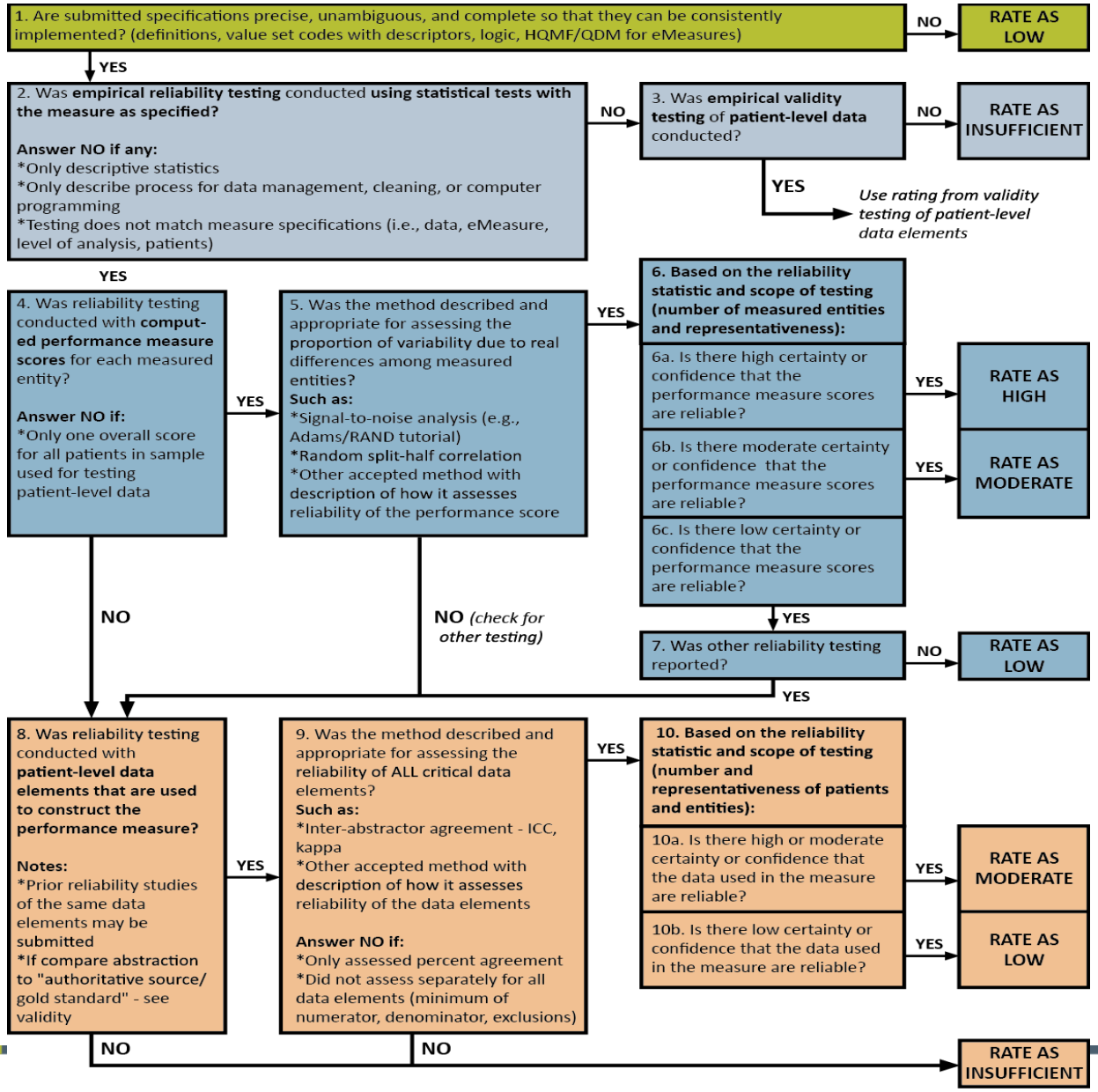
Evaluating Scientific Acceptability – Key Points (page 41)

Empirical analysis to demonstrate the reliability and validity of the measure as specified, including analysis of issues that pose threats to the validity of conclusions about quality of care such as exclusions, risk adjustment/stratification for outcome and resource use measures, methods to identify differences in performance, and comparability of data sources/methods.

Reliability Testing
Key points - page 42

- Reliability of the **measure score** refers to the proportion of variation in the performance scores due to systematic differences across the measured entities in relation to random variation or noise (i.e., the precision of the measure).
  - *Example - Statistical analysis of sources of variation in performance measure scores (signal-to-noise analysis)*
- Reliability of the **data elements** refers to the repeatability/ reproducibility of the data and uses patient-level data
  - *Example –inter-rater reliability*
- Consider whether testing used an appropriate method and included adequate representation of providers and patients and whether results are within acceptable norms
- Algorithm #2

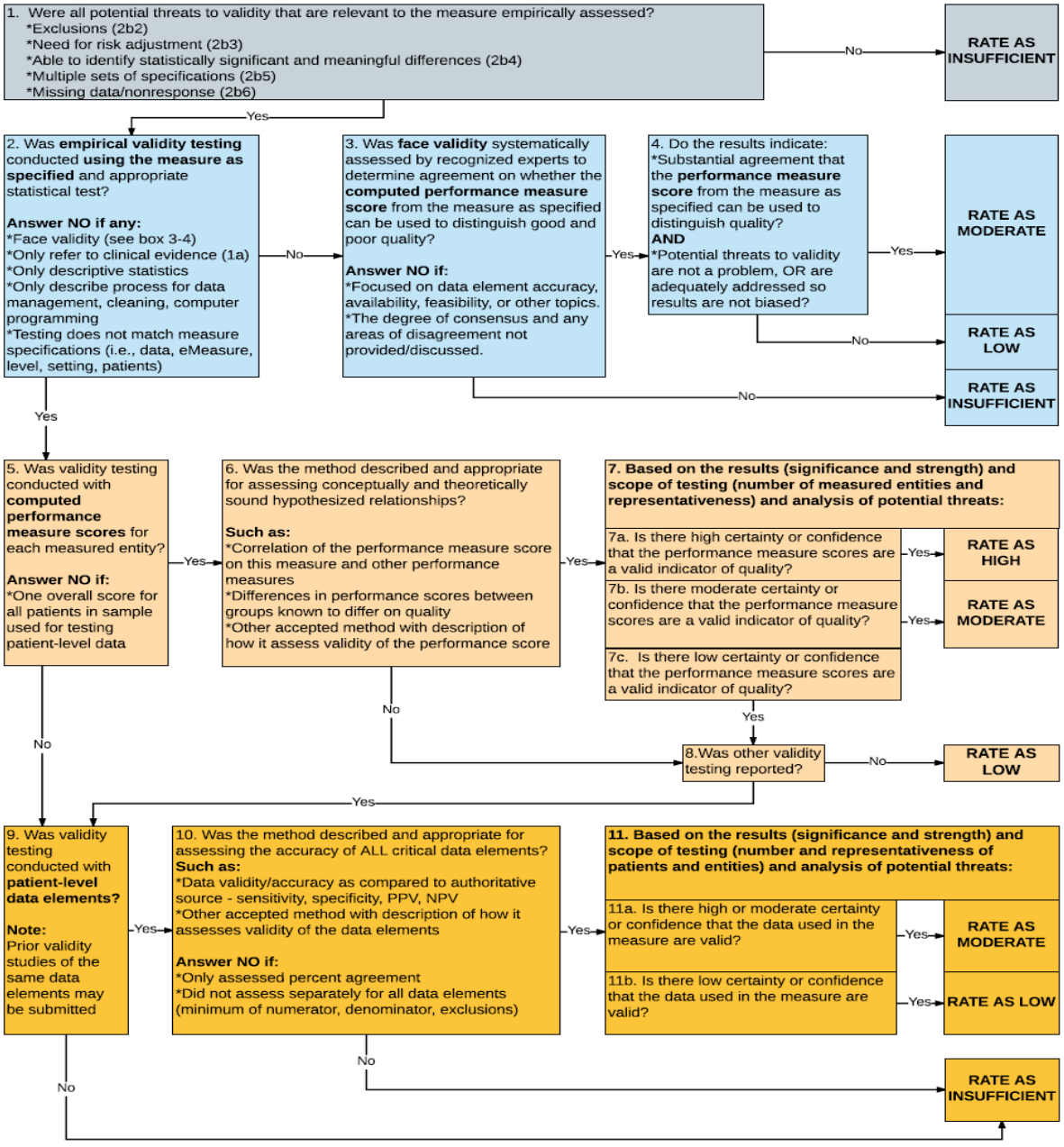
Rating Reliability: Algorithm #2 – page 43



Validity testing (pages 44 - 49)
Key points – page 47

- **Empirical testing**
  - *Measure score* – assesses a hypothesized relationship of the measure results to some other concept; assesses the correctness of conclusions about quality
  - *Data element* – assesses the correctness of the data elements compared to a "gold standard"
- **Face validity**
  - *Subjective determination by experts that the measure appears to reflect quality of care*
    - » Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.
    - » Requires systematic and transparent process, 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.

Rating Validity: Algorithm #3 – page 48



Threats to Validity

- Conceptual
  - *Measure focus is not a relevant outcome of healthcare or not strongly linked to a relevant outcome*
- Unreliability
  - *Generally, an unreliable measure cannot be valid*
- Patients inappropriately excluded from measurement
- Differences in patient mix for outcome and resource use measures
- Measure scores that are generated with multiple data sources/methods
- Systematic missing or “incorrect” data (unintentional or intentional)

Criterion #2: Scientific Acceptability

New measures	Maintenance measures
• Measure specifications are precise with all information needed to implement the measure	NO DIFFERENCE: Require updated specifications
• Reliability • Validity (including risk-adjustment)	**DECREASED EMPHASIS**: If prior testing adequate, no need for additional testing at maintenance with certain exceptions (e.g., change in data source, level of analysis, or setting) Must address the questions regarding use of social risk factors in risk-adjustment approach

Criterion #3: Feasibility (page 49)
Key Points – page 50

Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement.

3a: Clinical data generated during care process

3b: Electronic sources

3c: Data collection strategy can be implemented

Criterion #4: Usability and Use (page 50) Key Points – page 51

Extent to which potential audiences (e.g., consumers, purchasers, providers, policymakers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

Use (4a) Now must-pass for maintenance measures

4a1: Accountability and Transparency: *Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement.*

4a2: Feedback by those being measured or others: *Those being measured have been given results and assistance in interpreting results; those being measured and others have been given opportunity for feedback; the feedback has been considered by developers.*

Usability (4b)

4b1: Improvement: *Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.*

4b2: Benefits outweigh the harms: *The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).*

Criteria #3-4: Feasibility and Usability and Use

Feasibility

New measures	Maintenance measures
• Measure feasible, including eMeasure feasibility assessment	NO DIFFERENCE: Implementation issues may be more prominent

Usability and Use

New measures	Maintenance measures
• Use: used in accountability applications and public reporting	**INCREASED EMPHASIS**: Much greater focus on measure use and usefulness, including both impact and unintended consequences
• Usability: impact and unintended consequences	

Criterion #5: Related or Competing Measures (page 51-52)

If a measure meets the four criteria and there are endorsed/new related measures (same measure focus or same target population) or competing measures (both the same measure focus and same target population), the measures are compared to address harmonization and/or selection of the best measure.

- 5a. The measure specifications are harmonized with related measures **OR** the differences in specifications are justified.
- 5b. The measure is superior to competing measures (e.g., is a more valid or efficient way to measure) **OR** multiple measures are justified.

Updated guidance for measures that use ICD-10 coding: Fall 2017 and 2018

- Gap can be based on literature and/or data based on ICD-9 or ICD-10 coding
- Submit updated ICD-10 reliability testing if available; if not, testing based on ICD-9 coding will suffice
- Submit updated validity testing
  - *Submit updated empirical validity testing on the ICD-10 specified measure,* **if available**
  - **OR** *face validity of the ICD-10 coding scheme* **plus face validity** *of the measure score as an indicator of quality*
  - **OR** *face validity of the ICD-10 coding scheme* **plus score-level** *empirical validity testing based on ICD-9 coding*
  - **OR** *face validity of the ICD-10 coding scheme* **plus data element** *level validity testing based on ICD-9 coding, with face validity of the measure score as an indicator of quality due at* **annual update**

eMeasures

- "Legacy" eMeasures
  - *Beginning September 30, 2017 all respecified measure submissions for use in federal programs will be required to the same evaluation criteria as respecified measures – the "***BONNIE testing only**" **option will no longer meet endorsement criteria**
- For all eMeasures: Reliance on data from structured data fields is expected; otherwise, unstructured data must be shown to be both reliable and valid

Evaluation Process

- **Preliminary analysis (PA)**: To assist the Committee evaluation of each measure against the criteria, NQF staff and Methods Panel (if applicable) will prepare a PA of the measure submission and offer preliminary ratings for each criteria.
  - *The PA will be used as a starting point for the Committee discussion and evaluation*
  - *Methods Panel will complete review of Scientific Acceptability criterion for complex measures*
- **Individual evaluation**: Each Committee member conduct an in-depth evaluation on all measures
  - *Each Committee member will be assigned a subset of measures for which they will serve as lead discussant in the evaluation meeting.*

Evaluation Process

- **Measure evaluation and recommendations at the in-person/web meeting**: The entire Committee will discuss and rate each measure against the evaluation criteria and make recommendations for endorsement.
- **Staff will prepare a draft report** detailing the Committee's discussion and recommendations
  - *This report will be released for a 30-day public and member comment period*
- **Post-comment call**: The Committee will re-convene for a post-comment call to discuss comments submitted
- **Final endorsement decision by the CSAC**
- **Appeals** (if any)

Questions?