NATIONAL QUALITY FORUM

Primary Care and Chronic Illness

Measure Evaluation Tutorial Web Meeting

John Bernot, MD
Suzanne Theberge, MPH
Kathryn Goodwin, MS
Hiral Dudhwala, RN, MSN/MPH
Madison Jung

May 9, 2018

Welcome

Project Team – Primary Care and Chronic Illness

John Bernot, MD
VP, Quality Measurement Initiatives

Kathryn Goodwin, MS
Senior Project Manager

Suzanne Theberge, MPH
Senior Project Manager

Hiral Dudhwala, RN, MSN, MPH
Project Manager

Madison Jung
Project Manager

Agenda

- Overview of the Consensus Development Process (CDP)
- Overview of the Role of the Standing Committee, Expert Reviewers, and Scientific Methods Panel
- Project Activities and Timeline
- Overview of Measure Evaluation Criteria
- Overview of SharePoint
- Next steps

Primary Care and Chronic Illness
Spring 2018 Cycle Standing Committee

- Dale Bratzler, DO, MPH *(Co-Chair)*
- Anne Leddy, MD, FACE
- Adam Thompson, BA *(Co-Chair)*
- Grace Lee, MD
- Thiru Annaswamy, MD
- Anna McCollister-Slipp
- Robert Bailey, MD
- Janice Miller, DNP, CRNP, CDE
- Lindsay Botsford, MD
- Andrew Schachat, MD
- Roger Chou, MD
- Steven Strode, MD, Med, MPH, FAAFP
- William Curry, MD, MS
- William Taylor, MD
- Jim Daniels, BSN
- Kimberly Templeton, MD
- Woody Eisenberg, MD
- John Ventura, DC
- Kim Elliott, PhD
- V. Katherine Gray, PhD
- Ann Kearns,MD, PhD

Primary Care and Chronic Illness
Expert Reviewers

- Emily Aaronson, MD
- Amesh Adalja, MD
- Esther Babady, PhD, D(ABMM)
- Carlos Bagley, MD, FAANS
- Gerene Bauldoff, PhD, RN, FAAN
- Kenneth Benson
- Tamala Bradham, DHA, PhD, CCC-A
- Kathleen Brady, MD, MSCE
- Steven Brotman, MD, JD
- Craig Butler, MD, MBA, CPE
- Laura Evans, MD, MSc
- Scott Friedman, MD
- Piero Garzaro, MD
- William Glomb, MD, FCCP, FAAP
- Donald Goldmann, MD
- Stephen Grossbart, PhD
- James Mitchell Harris, PhD
- Marci Harris Hayes, PT, DPT, MSCI, OCS
- Jeffrey Hart, MS
- Starlin Haydon-Greatting, MS, BS, Pharm, FAPhA
- Mark Jarrett, MD, MBA
- Ella Kazerooni, MD, MS

Primary Care and Chronic Illness
Expert Reviewers

- Michael Lane, MD, MSc, MPHS, CPPS
- David Lang, MD
- Jeffrey Lewis, BA
- Jason Matuszak, MD, FAAFP, CAQSM, RMSK
- Richard Madonna, OD, MA, FAAO
- John McClay, MD
- Daniel Merenstein, MD
- Richard Murray, MD
- Melinda Neuhauser, PharmD, MPH, FCCP, FASHP
- Rocco Orlando, MD, FACS
- Susan Pollart, MD
- Todd Rambasek, MD, FAAAAI
- Crystal Riley, PharmD, MHA, MBA, CPHQ, CHPIT
- Catherine Roberts, MD
- James Rosenzweig, MD
- Christine Schindler, PhD, RN, CPNP-AC/PC, WCC
- Michael Stewart, MD, MPH
- Christopher Visco, MD
- Chana West, RN, MSN
- Kathleen Yaremchuk, MD, MSA
- Jacquelyn Youde, AuD, CCC-A

Overview of the CDP Process

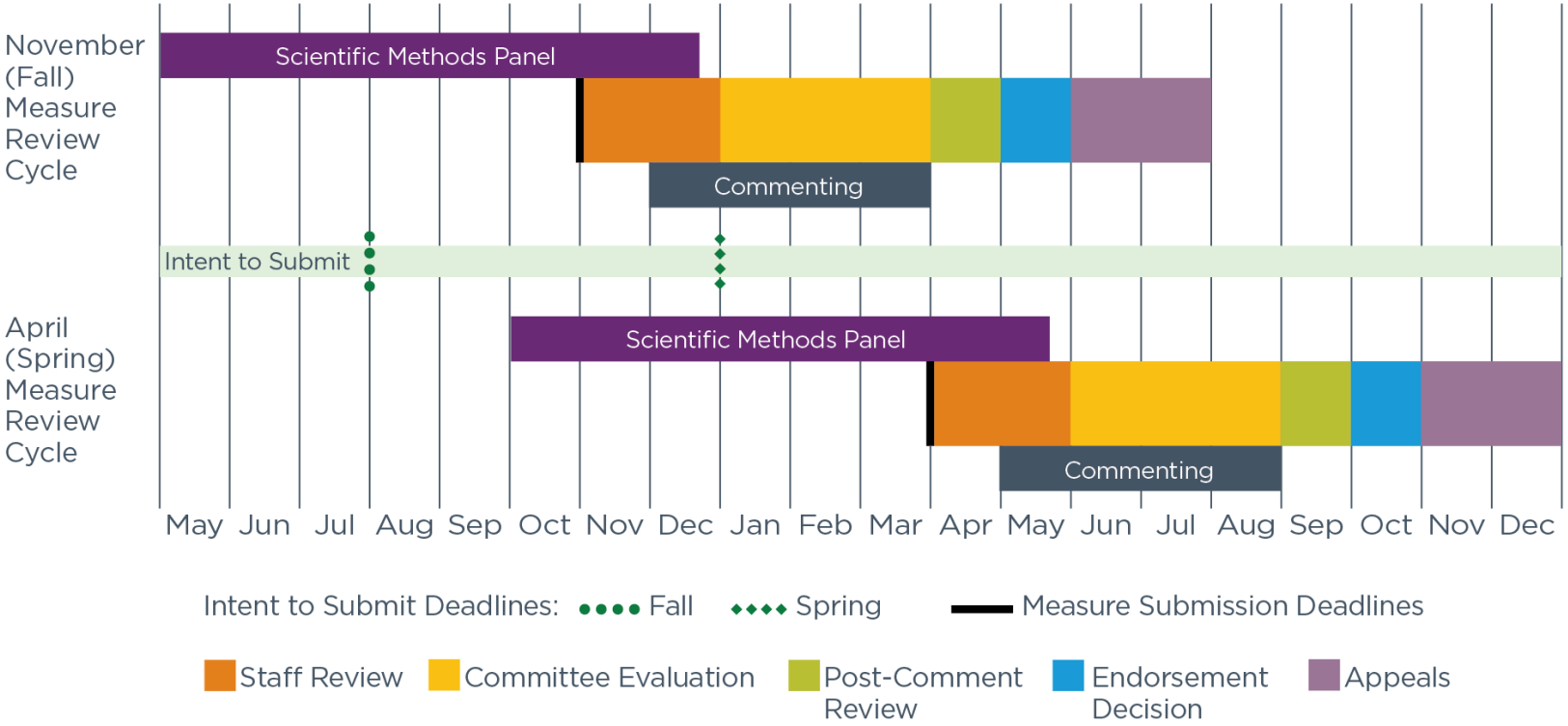
NQF Consensus Development Process (CDP) 6 Steps for Measure Endorsement

- Intent to Submit
- Call for Nominations
- Measure Evaluation
  - New structure/process
  - Newly formed NQF Scientific Methods Panel
  - Measure Evaluation Technical Report
- Public Commenting Period with Member Support
- Measure Endorsement
- Measure Appeals

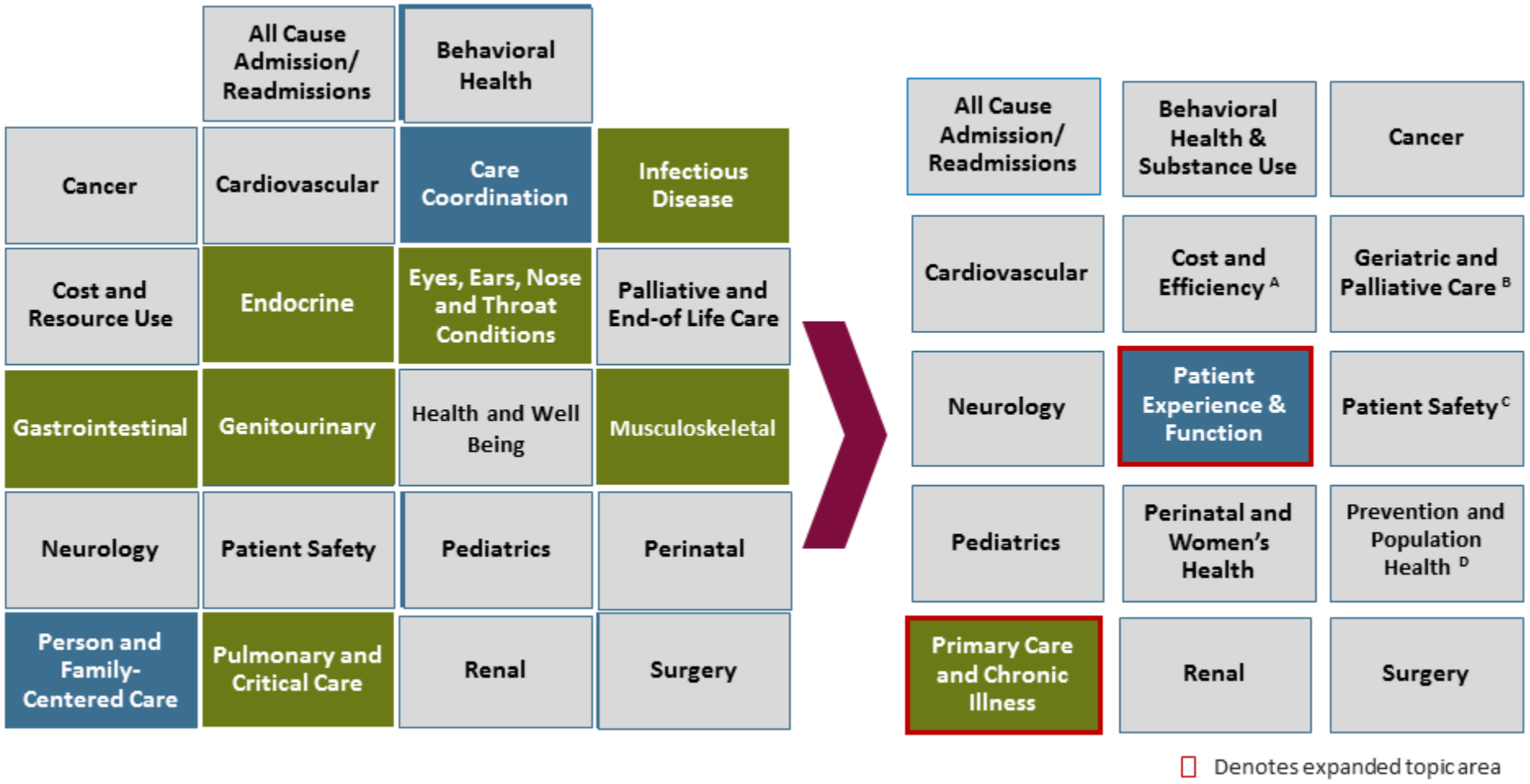
Measure Review: Two Cycles Per Year

Consensus Development Process:

Two Cycles Every Contract Year



15 New Measure Review Topical Areas



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

Roles of Standing Committee, Expert Reviewers, and Scientific Methods Panel

Role of the Standing Committee
General Duties

- Act as a proxy for the NQF multistakeholder membership
- Serve 2-year or 3-year terms
- Work with NQF staff to achieve the goals of the project
- Evaluate candidate measures against the measure evaluation criteria
- Respond to comments submitted during the review period
- Respond to any directions from the CSAC

Role of the Standing Committee
Measure Evaluation Duties

- All members evaluate ALL measures
- Evaluate measures against each criterion
  - *Indicate the extent to which each criterion is met and rationale for the rating*
- Make recommendations for endorsement
- Oversee Primary Care and Chronic Illness portfolio of measures
  - *Promote alignment and harmonization*
  - *Identify gaps*

Role of the Standing Committee Co-Chairs

- Co-facilitate Standing Committee (SC) meetings
- Work with NQF staff to achieve the goals of the project
- Assist NQF in anticipating questions and identifying additional information that may be useful to the SC
- Keep SC on track to meet goals of the project without hindering critical discussion/input
- Represent the SC at CSAC meetings
- Participate as a SC member

Role of NQF Staff

- NQF project staff works with SC to achieve the goals of the project and ensure adherence to the consensus development process:
  - Organize and staff SC meetings and conference calls
  - Guide the SC through the steps of the CDP and advise on NQF policy and procedures
  - Review measure submissions and prepare materials for Committee review
  - Draft and edit reports for SC review
  - Ensure communication among all project participants (including SC and measure developers)
  - Facilitate necessary communication and collaboration between different NQF projects

Role of NQF Staff
Communication

- Respond to NQF member or public queries about the project
- Maintain documentation of project activities
- Post project information to NQF's website
- Work with measure developers to provide necessary information and communication for the SC to fairly and adequately evaluate measures for endorsement
- Post final project report

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.

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 to retain a diverse, yet specific, expertise within an “expert reviewer pool” 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
- 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.*

NQF Consensus Development Process (CDP) Measure Evaluation

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**

Non-Complex Measures

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

Questions?

Primary Care and Chronic Illness Spring 2018 Cycle Activities

Spring 2018 Measures

- Eight Maintenance Measures (Steward NCQA):
  - NQF#0037: Osteoporosis Testing in Older Women (OTO)
  - NQF#0046: Screening for Osteoporosis for Women 65-85 Years of Age
  - NQF#0053: Osteoporosis Management in Women Who Had a Fracture
  - NQF#0055: Comprehensive Diabetes Care: Eye Exam (retinal) performed
  - NQF#0056: Comprehensive Diabetes Care: Foot Exam
  - NQF#0057: Comprehensive Diabetes Care: Hemoglobin A1c (HbA1c) Testing
  - NQF#0062: Comprehensive Diabetes Care: Medical Attention for Nephropathy
  - NQF#0575: Comprehensive Diabetes Care: Hemoglobin A1c (HbA1c) Control (<8.0%) (Methods Panel Reviewed)

Activities and Timeline - Spring Cycle 2018

All times ET

Meeting	Date/Time
Measure Submission Deadline	Monday, April 9, 2018
Commenting Period	May 1 - June 12, 2018
Committee Measure Evaluation Tutorial Web Meeting	Wednesday, May 9, 2018, 2:00-4:00 pm
Committee In-Person Meeting (1 day)	June 21, 2018
Committee Post-Measure Evaluation Web Meeting (2 hours)	Tuesday, June 26, 2018, 2:00-4:00 pm
Draft Report Comment Period (30 days)	TBD (July/August, 2018)
Committee Post-Comment Web Meeting	Wednesday, September 19, 2018, 1:00-3:00pm
CSAC Review	October 15 - November 2, 2018
Appeals Period (30 days)	November 6 - December 5, 2018

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
(Committee Guidebook 2017, 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

(Committee Guidebook 2017, 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

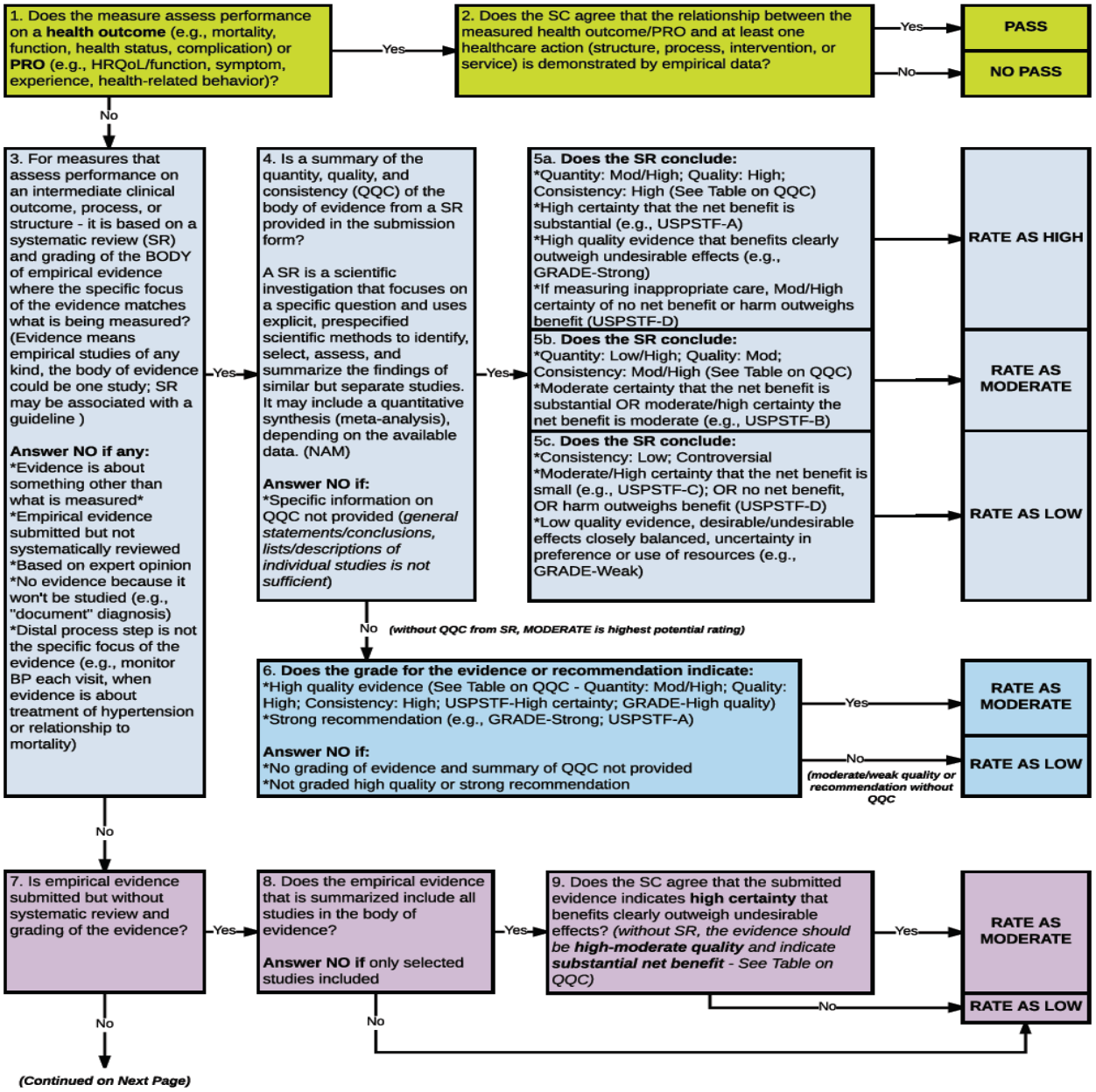
1b. Opportunity for Improvement: demonstration of quality problems and opportunity for improvement, i.e., data demonstrating considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or disparities in care across population groups

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

Subcriteron 1a: Evidence
(Committee Guidebook 2017, 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 43*



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 from last evaluation; Standing Committee to affirm no change in evidence IF evidence has changed, 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

(Committee Guidebook 2017, 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)**

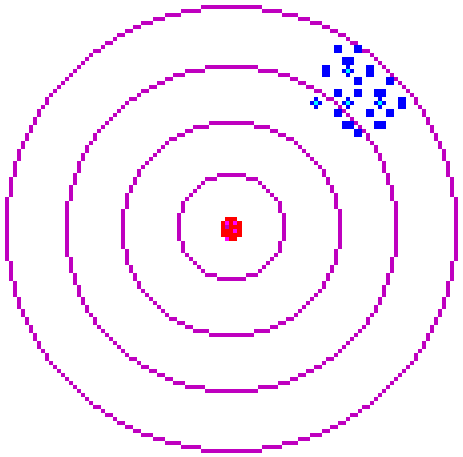
- 2a1. Precise specifications including exclusions
- 2a2. Reliability testing—data elements or measure score

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

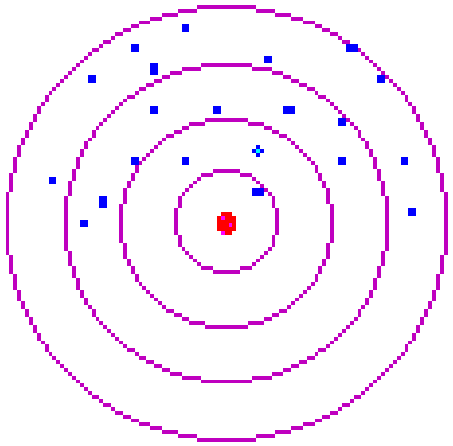
Reliability and Validity
(Committee Guidebook 2017, page 40)

Assume the center of the target is the true score…



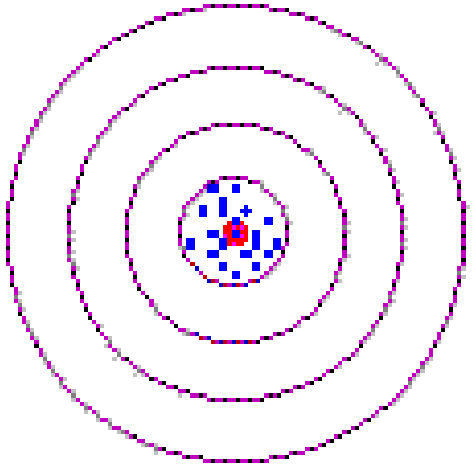
Reliable Not Valid

Consistent, but wrong



Neither Reliable Nor Valid

Inconsistent & wrong



Both Reliable And Valid

Consistent & correct

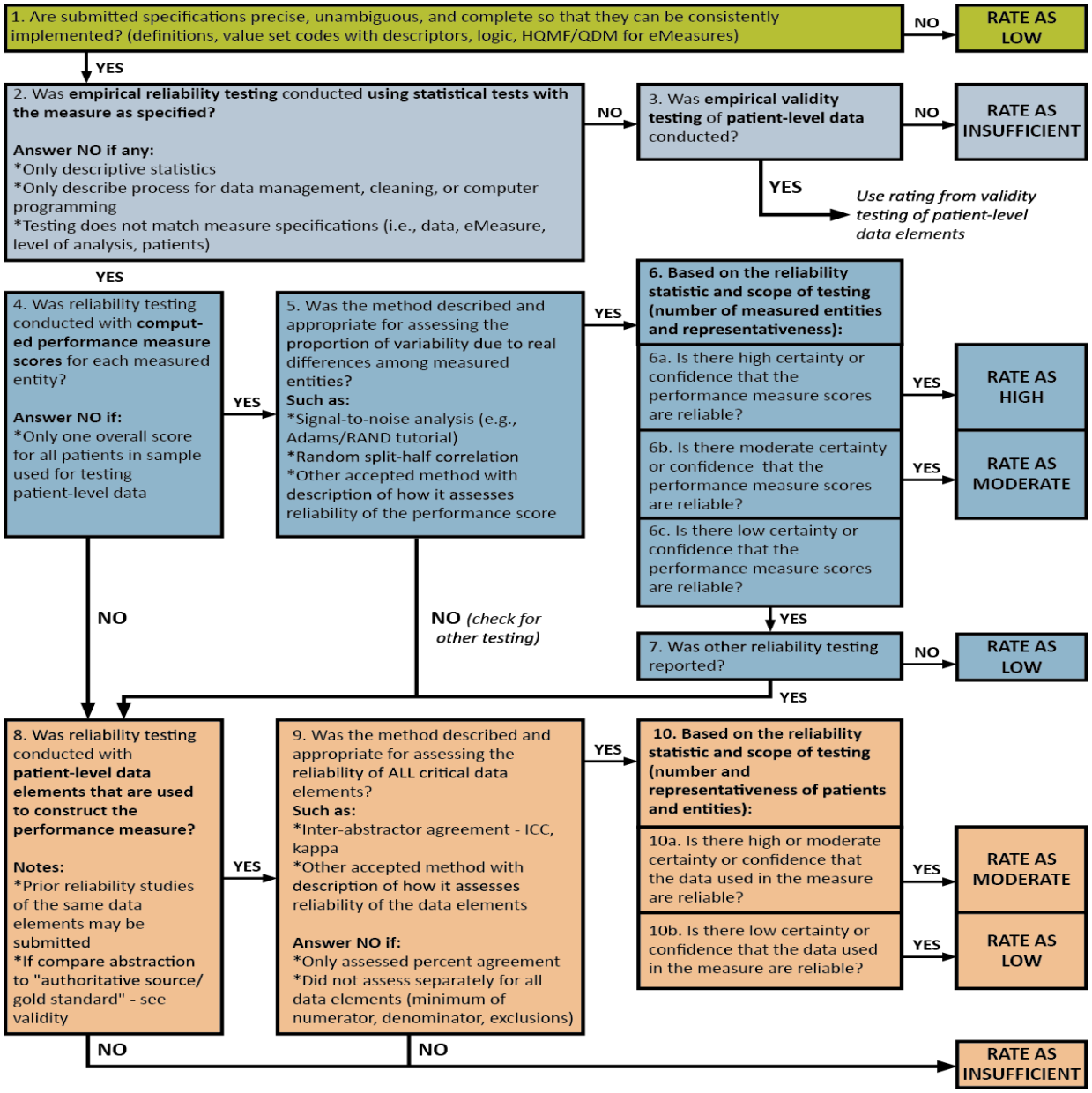
Evaluating Scientific Acceptability – Key Points
(Committee Guidebook 2017, page 41)

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
(Committee Guidebook 2017, 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

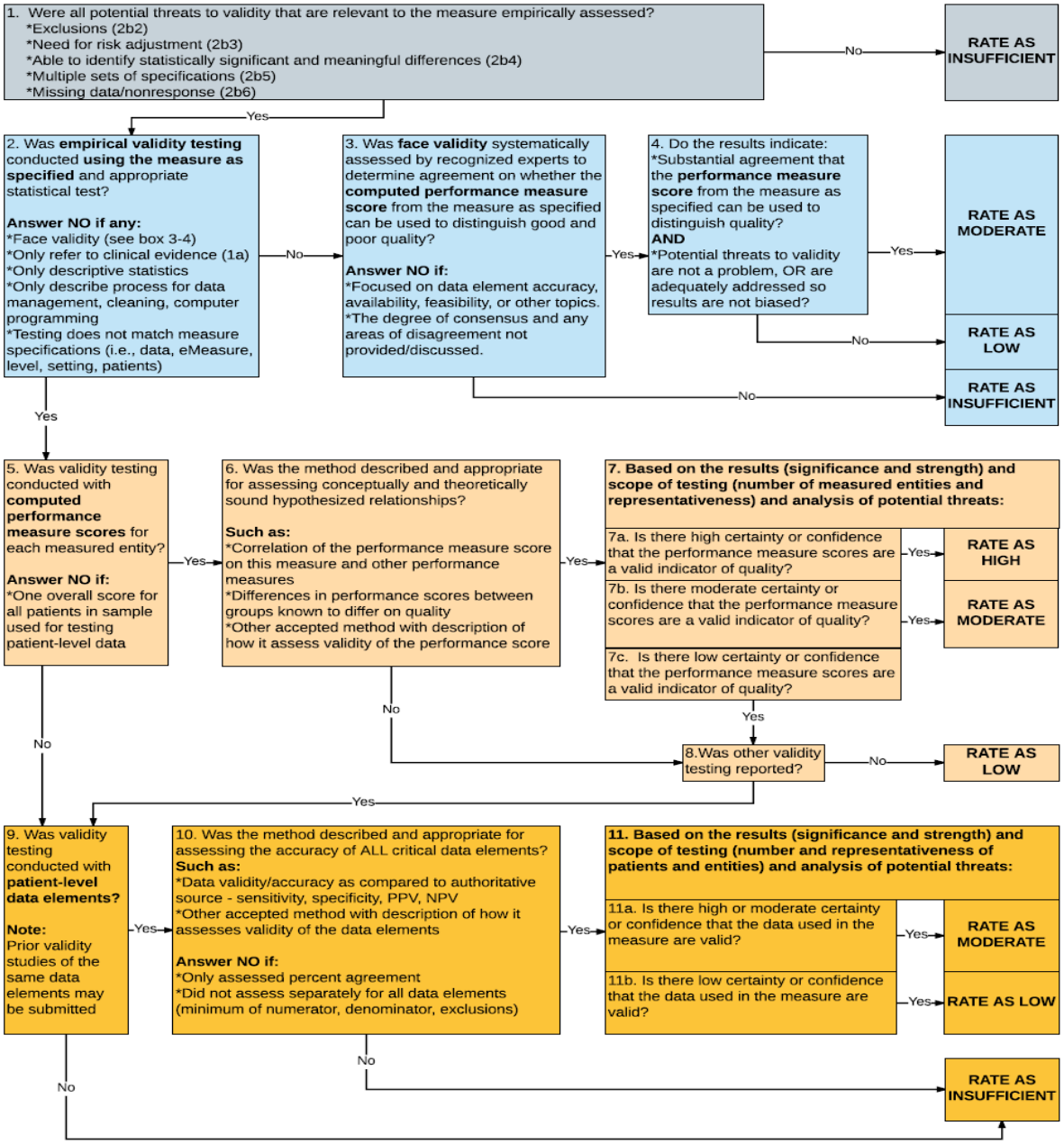
Rating Reliability: Algorithm #2 – *page 43*



Validity testing—Key points
(Committee Guidebook 2017, pages 44 – 49)

- **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 – Key Points
(Committee Guidebook 2017, page 49-50)

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—Key Points
(Committee Guidebook 2017, page 50-51)

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

New measures	Maintenance measures
Feasibility	
• Measure feasible, including eMeasure feasibility assessment	NO DIFFERENCE: Implementation issues may be more prominent
Usability and Use	
• Use: used in accountability applications and public reporting • Usability: impact and unintended consequences	**INCREASED EMPHASIS**: Much greater focus on measure use and usefulness, including both impact and unintended consequences

Criterion #5: Related or Competing Measures
(Committee Guidebook 2017, 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 re-specified measure submissions for use in federal programs will be required to meet the same evaluation criteria as re-specified 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 conducts an in-depth evaluation on all measures (responses collected via SurveyMonkey)
  - 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 submitted comments
- **Final endorsement decision by the CSAC**
- **Appeals** (if any)

Questions?

SharePoint Overview

SharePoint Overview

http://share.qualityforum.org/Projects/Primary%20Care%20and%20Chronic%20Illness/SitePages/Home.aspx

- Accessing SharePoint
- Standing Committee Policy
- Standing Committee Guidebook
- Measure Document Sets
- Meeting and Call Documents
- Committee Roster and Biographies
- Calendar of Meetings

Next Steps

Next Steps

- Measure Submission Deadline, Spring 2018 Cycle
  - Monday, April 9, 2018
  - Committee members should expect to receive measures for review late May/early June, 2018

- In-Person Meeting (at NQF Headquarters in Washington D.C.)
  - Thursday, June 21, 2018, 8:30-5:30 pm EST

Project Contact Info

- Email: primarycare@qualityforum.org
- NQF Phone: 202-783-1300
- Project page:

http://www.qualityforum.org/Primary_Care_and_Chronic_Illness.aspx

- SharePoint site:

http://share.qualityforum.org/Projects/Primary%20Care%20and%20Chronic%20Illness/SitePages/Home.aspx

Questions?

THANK YOU