NATIONAL QUALITY FORUM

Scientific Methods Panel Monthly Call Meeting

Karen Johnson, MS
May Nacion, MPH
Miranda Kuwahara, MPH

October 25, 2017

Welcome, Roll Call, and Review of Meeting Objectives

Scientific Methods Panel Members

Co-Chairs	David Cella, PhD
	Karen Joynt Maddox, MD, MPH

Panel Members	
J. Matt Austin, PhD	Paul Kurlansky, MD
Bijan Borah, MSc, PhD	Zhenqiu Lin, PhD
John Bott, MBA, MSSW	Jack Needleman, PhD
Lacy Fabian, PhD	David Nerenz, PhD
Marybeth Farquhar, PhD, MSN, RN	Eugene Nuccio, PhD
Jeffrey Geppert, EdM, JD	Jennifer Perloff, PhD
Paul Gerrard, BS, MD	Sam Simon, PhD
Laurent Glance, MD	Michael Stoto, PhD
Stephen Horner, RN, BSN, MBA	Christie Teigland, PhD
Sherrie Kaplan, PhD, MPH	Ronald Walters, MD, MBA, MHA, MS
Joseph Kunisch, PhD, RN-BC, CPHQ	Susan White, PhD, RHIA, CHDA

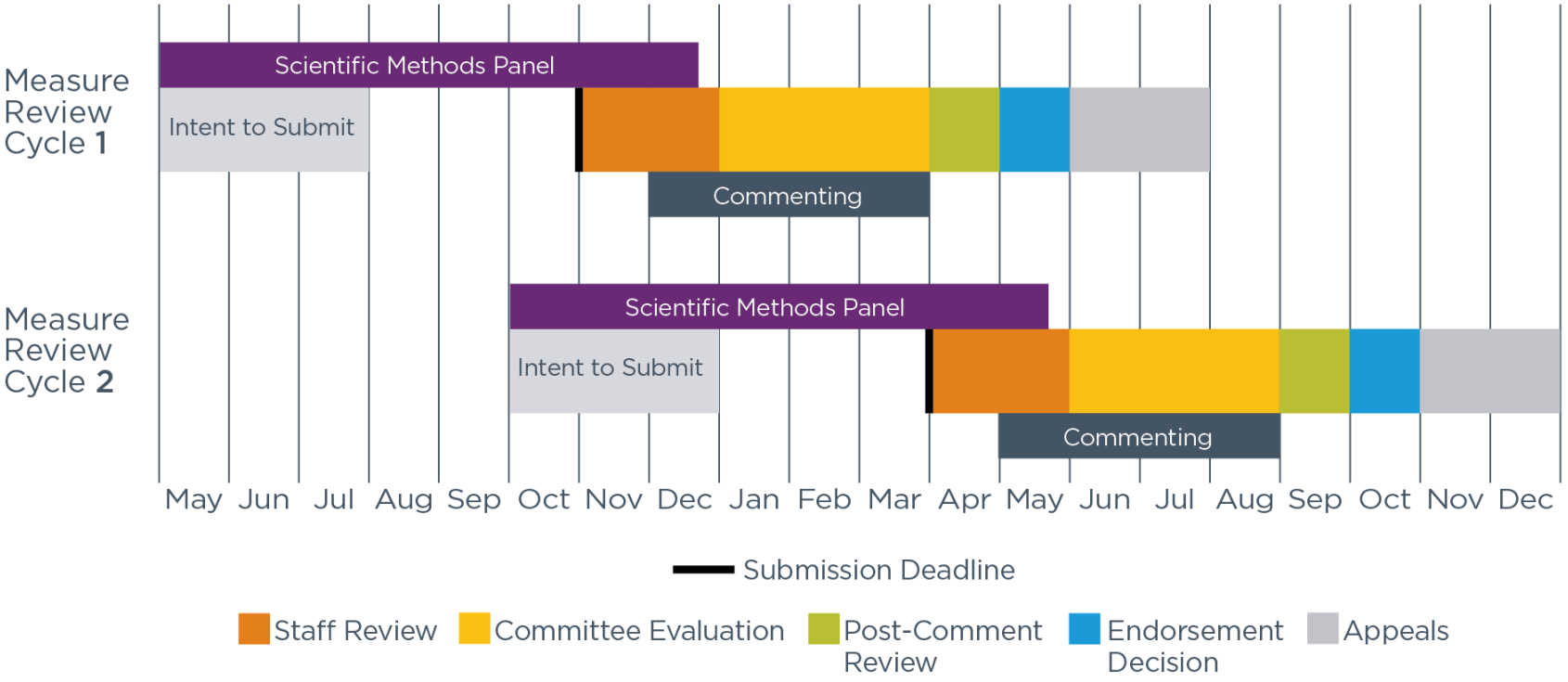
Meeting Objectives

- Review Cycle 1 timeline and evaluation process
- Review measure evaluation key points
- Q&A Session

NQF Consensus Development Process (CDP)

Consensus Development Process:

Two Cycles Every Contract Year



November 2017

Sun	Mon	Tue	Wed	Thu	Fri	Sat
29	30	31 Halloween	1 Measure Submission Group 1 (All Projects)	2 Staff will determine complex and noncomplex measures	3 Staff will perform a completeness check	4 ○ Full Moon
5	6 Staff assigns measures to Panel	7 Staff notifies Panel of final assignments	8 Group 1 measures provided to Panel for Review Measure Submission Group 2 (All Projects)	9	10 Veterans Day observed ◑ 3rd Quarter	11 Veterans Day
12	13	14	15 Group 2 measures provided to Panel for Review Measure Submission Group 3 (All Projects)	16	17	18 ● New Moon
19	20	21	22 Group 1 Review Due Group 3 measures provided to Panel for Review	23 Thanksgiving Day	24 Group 1 CNR measures reviewed by co-chairs	25
26 ◐ 1st Quarter Group 1 CNR measures reviewed by co-chairs	27 Staff sends Group 1 PA to project teams	28 Group 1 full preliminary analysis shared with the developer	29 Group 2 Review Due	30	1	2

December 2017

Sun	Mon	Tue	Wed	Thu	Fri	Sat
26 ◐ 1st Quarter	27	28	29	30	1	2
					Group 2 CNR measures reviewed by co-chairs	
3 ○ Full Moon	4	5 Group 2 full preliminary analysis shared with the developer	6 Group 3 Review Due	7	8	9
Group 2 CNR measures reviewed by co-chairs					Group 3 CNR measures reviewed by co-chairs	
10 ◑ 3rd Quarter	11	12 Group 3 full preliminary analysis shared with the developer	13	14	15	16
Group 3 CNR measures reviewed by co-chairs						
17	18 ● New Moon	19	20	21	22	23
24 Christmas Eve	25 Christmas Day	26 ◐ 1st Quarter	27	28	29	30
31 New Year's Eve	1 New Year's Day ○ Full Moon	2	3	4	5	6

Evaluation of the Scientific Acceptability Criterion

- Provide evaluation and ratings for reliability and validity subcriteria
  - *This information will help to inform the standing committee's endorsement decision*
  - *The Scientific Methods Panel will not render endorsement recommendations*
  - *Standing committees may raise concerns with the specifications of the measure or with potential threats to validity (e.g., selection of variables for risk adjustment model) and can overturn the Scientific Methods Panel ratings*

Workflow for Evaluations

- A minimum of three panel members will independently evaluate each measure
- The majority recommendation from the three evaluations will serve as the overall assessment of reliability and validity
- If there is substantial disagreement in the ratings between the three reviewers, the panel co-chairs will evaluate the measure and determine the overall recommendation
- NQF staff will compile the Methods Panel's ratings, evaluation, and commentary on reliability and validity and provide it to NQF's standing committees

A Few More Details…

- 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*
- Workload ~15-20 measures per year (per panel member)
  - *We will try to match you based on expertise, availability, and need for recusal*

Questions?

Key Points for Measure Evaluation

Available Resources

- 2017 Criteria and Guidance Document (located on the Methods Panel SharePoint page)
- Standing Committee Guidebook (section 7)
- "Key Points" guidance document (draft available on ShP page, but we may be updating a little before Nov 1)
- Methods Panel staff (for questions about the criteria)

- Potentially, clarification from developers
  - *Have to be requested very early on*
  - *NQF staff are the go-betweens*
  - *Will give developers 48 hours to respond to questions/make changes to their forms*

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

Evaluating Scientific Acceptability – Key Points (page 41)

- Empirical analysis is expected
  - *Face validity of the measure score is allowed for new measures, but not for maintenance measures unless there is justification*
- NQF is not prescriptive about how empirical measure testing is done
- NQF has not set minimum thresholds for reliability or validity testing results
- Reliability and Validity are “must-pass”
- There is an *extra criterion* for composite measures (empirical analyses support the composite construction approach)
- There may be different/additional testing requirements depending on measure type

Criterion #2a: Reliability

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

2a1. Precise specifications including exclusions

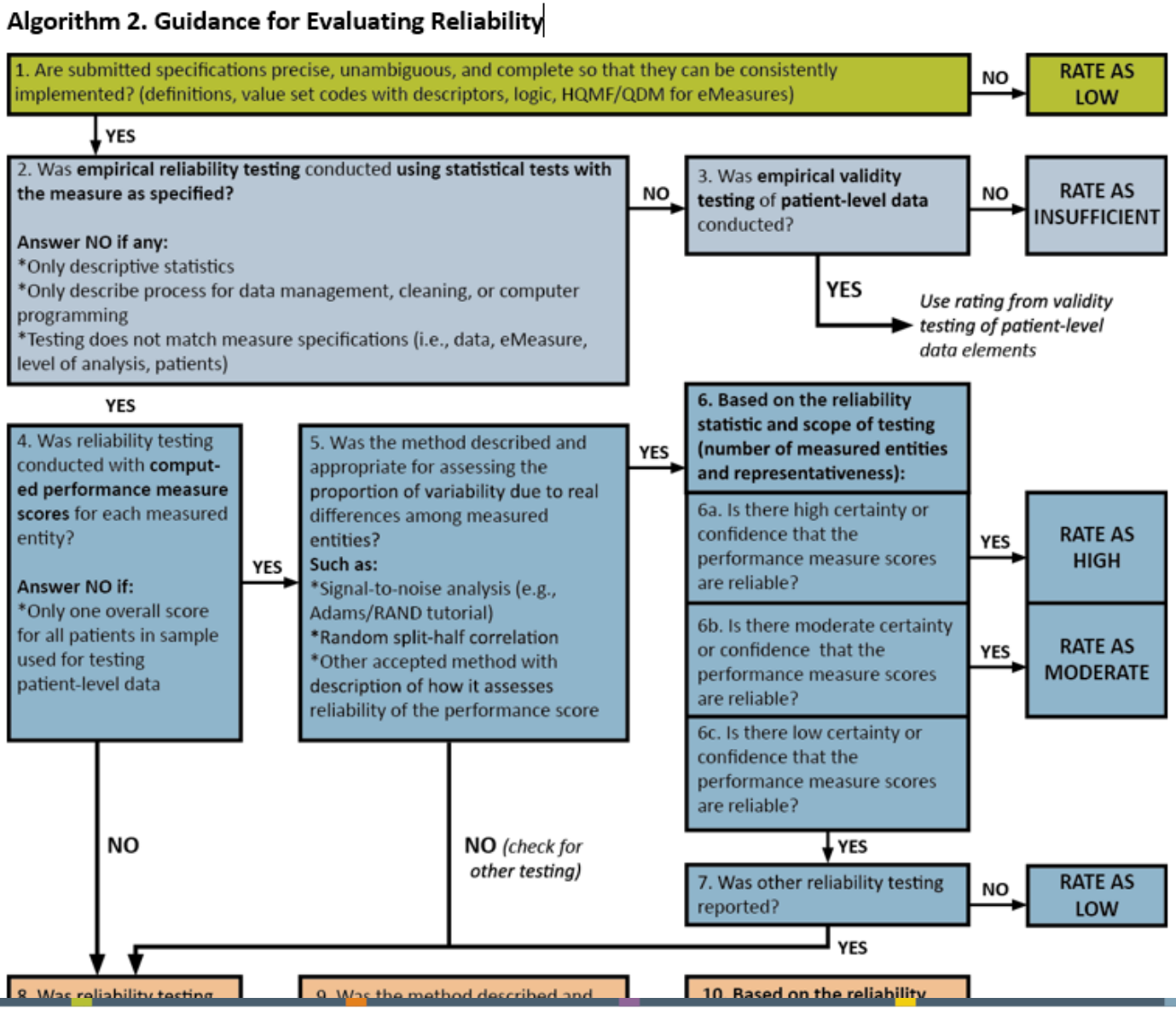
2a2. Reliability testing—data elements or measure score

- Specifications: Precise, unambiguous, complete
  - *eMeasure (eCQM) logic evaluated by NQF staff*
- Testing: either data element level OR measure score level – doesn't have to be both
  - *If data element validity testing provided, we **do not** require additional reliability testing*
  - *For eMeasures (eCQMs): reliability testing not required **if** based on data from structured data fields*

Reliability Testing - Key points (page 42)

- Reliability of the ***measure score***: proportion of variation in the performance scores due to systematic differences across the measured entities in relation to random variation or noise
  - *Can you differentiate between providers?*
  - *Example - 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*
  - *At minimum, for numerator, denominator, exclusions*
- Consider whether testing used an appropriate method and included adequate representation of providers and patients and whether results are within acceptable norms
- Algorithm #2 – page 43

Rating Reliability: Algorithm #2 – page 43



Validity testing (pages 44)
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
  - *We don't worry too much about labels such as concurrent predictive, etc.*
- Data element – assesses the correctness of the data elements compared to a "gold standard"
  - *We want sensitivity/specificity; have allowed less (kappa values showing agreement with gold standard)*

Validity testing (pages 44)
Key points – page 47

Face validity

- Subjective determination by experts that the measure appears to reflect quality of care
  - *Systematic and transparent process*
  - *Explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality*
  - *Degree of consensus and any areas of disagreement must be provided/discussed*

Validity testing (pages 44)
Key points – page 47

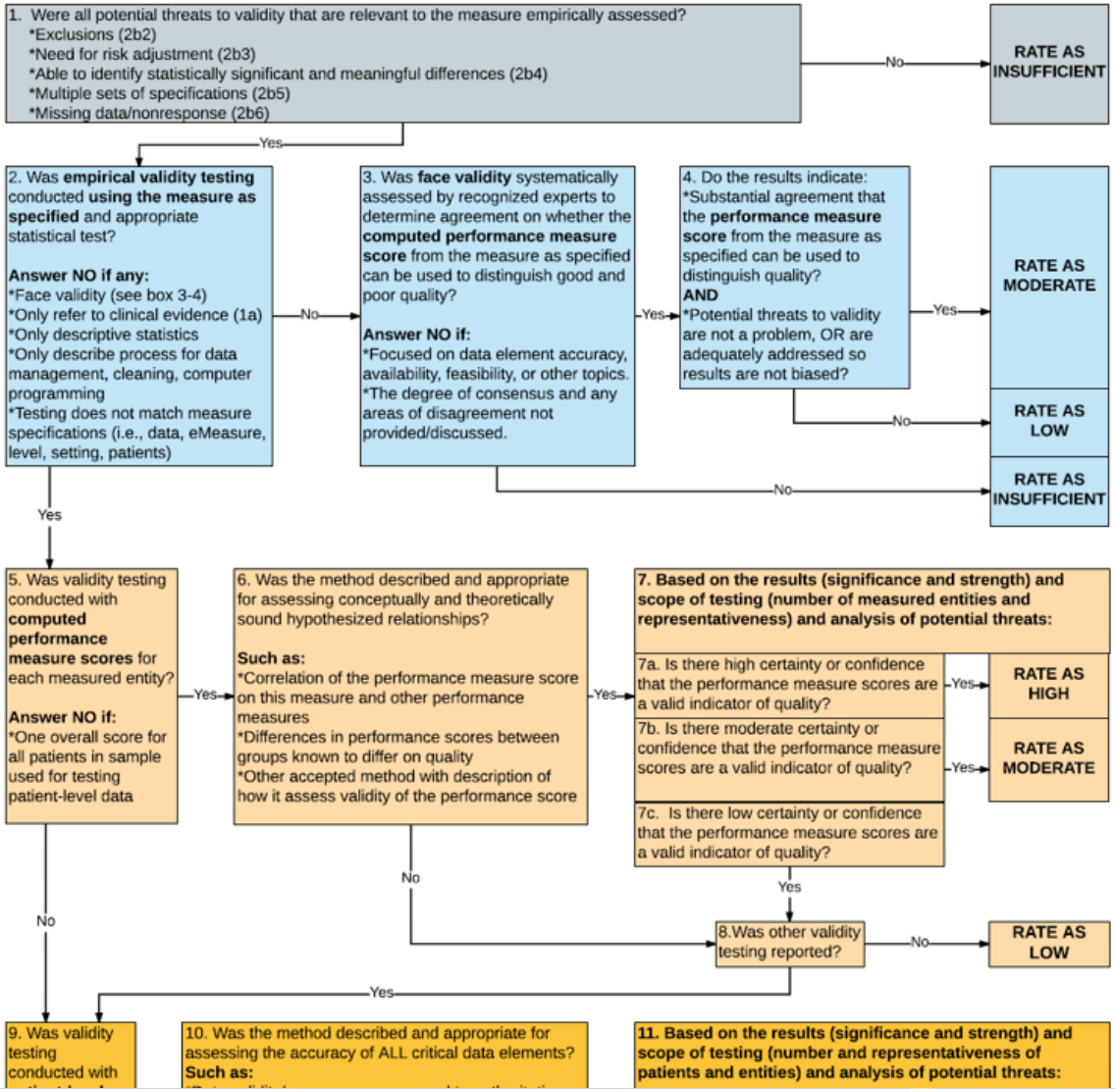
- **If measure converted from ICD-9 to ICD-10**
  - *Submit updated empirical validity testing on the ICD-10 specified measure, if available*
  - ***OR** face validity of the ICD-10 coding scheme plus face validity of the measure score as an indicator of quality*
  - ***OR** face validity of the ICD-10 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 (eCQMs)
  - *Testing in >1 EHR system*

Threats to Validity

- Exclusions
  - *Any patients inappropriately excluded from measurement?*
  - *Exclusions consistent with evidence, and of sufficient frequency to warrant inclusion?*
- Risk-adjustment
  - *For outcome and C/RU measures, risk-adjustment is expected – but developer can provide rationale/data to support not adjusting*
  - *Social risk factors can be included, if there is a conceptual rationale*
  - *Conceptual rationale for risk factors should be included in submission materials*
  - *Expect calibration/discrimination statistics, as well as analysis to support inclusion (or not) of social risk factors*
- Meaningful differences in performance
- Comparable results for measure scores that are generated with multiple data sources/methods
- Missing data do not produce biased results

Rating Validity: Algorithm #3 – page 50

Algorithm 3. Guidance for Evaluating Validity



Differences in Testing Requirements by Measure Type

- Composite measures
  - *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
  - *As noted earlier, there is also an extra criterion*
    - » How this is addressed by the developer will depend on the type of composite
- Instrument-based measures
  - *For reliability and validity, require testing at both the data element level (i.e., of the instrument) AND the measure score level (i.e., testing of the actual performance measure)*
- eMeasures (eCQMs)
  - *Usually, only concerned with validity*
  - *As noted earlier, need to test with >1 EHR system*

Questions?

Evaluating Scientific Acceptability

Evaluating Scientific Acceptability

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. **Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion.**

Instructions

- Please complete the following form for each measure you are evaluating and pay close attention to the skip logic directions.
- If your answer indicates you to provide a rationale, please use the space provided to explain your reasoning.
- If you are unable to check a box, please highlight or shade the box for your response.
- The 'overall rating' under each section must be answered.
- TIPSs are provided under each question to help you answer the question.

Name of Reviewer:
Measure Number *(if previously endorsed)*:
Measure Title:

Reliability

1. Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented? *NOTE: NQF staff will conduct a separate, more technical, check of eMeasure (eCQM) specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.*
   TIPS: Consider the following: Are all the data elements clearly defined? Are all appropriate codes included? Is the logic or calculation algorithm clear? Is it likely this measure can be consistently implemented?
   ☐Yes
   ☐No (if no, please explain below)

2. Was empirical reliability testing conducted using statistical tests with the measure as specified?
   TIPS: Check the 2nd "NO" box below if: only descriptive statistics; only describe process for data management/cleaning/computer programming; testing does not match measure specifications (i.e. data, eMeasure, level of analysis, patients)
   ☐Yes (go to Question #4)
   ☐No, there is no reliability testing described at all (go to Question #3)
   ☐No, there is reliability testing information, but not using statistical tests and/or not for the measure as specified (please explain below and rate OVERALL Reliability as INSUFFICIENT and STOP)

Member and Public Comment

Next Steps

- Monthly 1 hour Calls
  - *Every 4th Wednesday of the month?*
  - *For Nov/Dec – Dec 6 at 4 pm ET*
- Measure-Specific DOI
  - *Complete survey link by COB, Friday October 27.*
- Post-Call Exercise
- Contact Information: methodspanel@qualityforum.org

THANK YOU