NATIONAL QUALITY FORUM

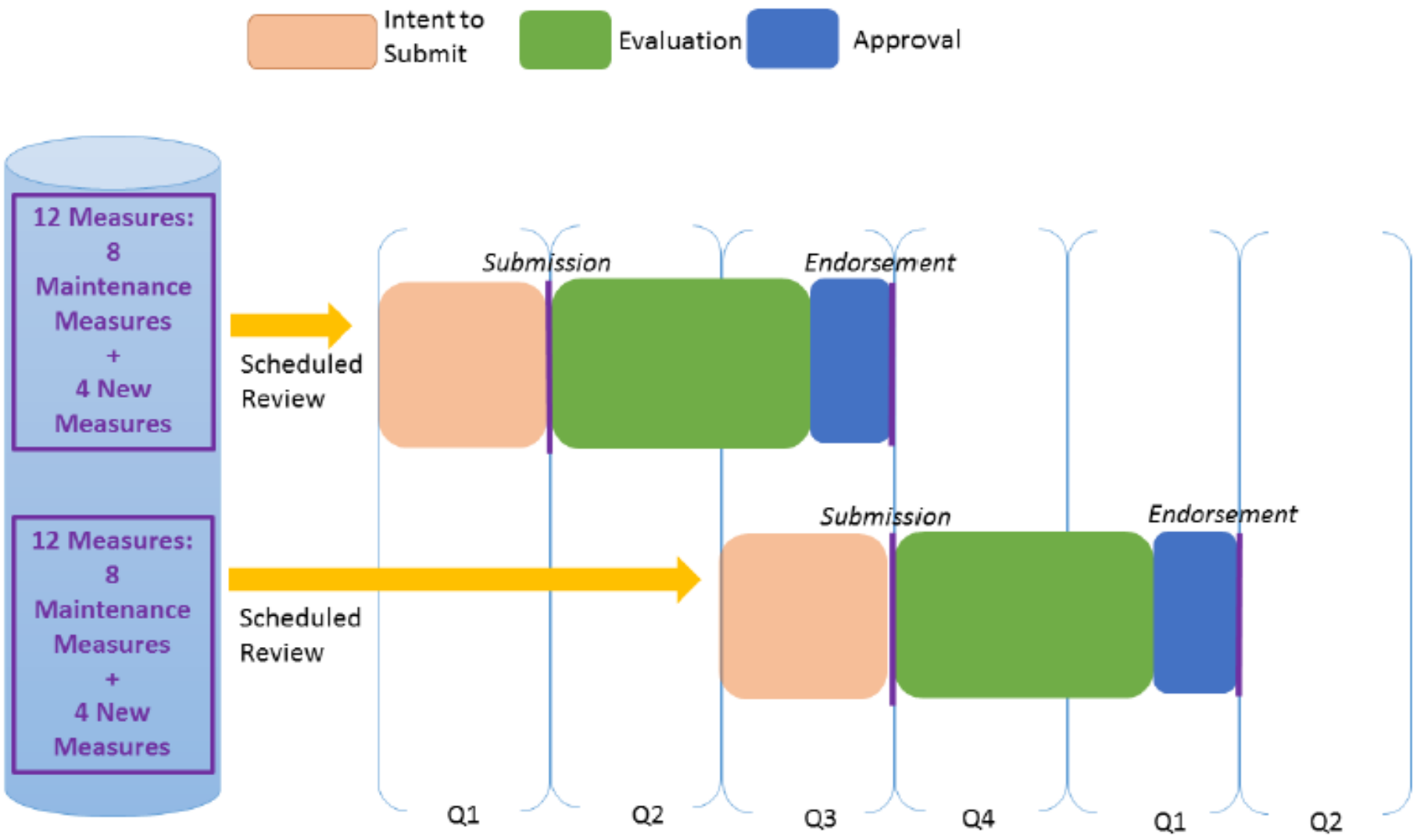
2017 CDP Redesign Highlights

Intent to Submit

Measure stewards/developers notify NQF of their intent to submit a measure at least three months prior to their selected cycle.

- Objective:
  - *Allow NQF to adequately plan for measures*
  - *Provide developers technical assistance prior to submitting measures*

- Intent to submit deadline: 3 months prior to the selected cycle


Intent to Submit
Evaluation
Approval
12 Measures: 8 Maintenance Measures + 4 New Measures
Scheduled Review
12 Measures: 8 Maintenance Measures + 4 New Measures
Scheduled Review
Submission
Endorsement
Submission
Endorsement
Q1
Q2
Q3
Q4
Q1
Q2

Intent to Submit, cont.

- Intent to Submit information needed:
  - **Planned submission date** (cycle and year)
  - Measure name
  - Measure description
  - Measure title
  - Measure type
  - Level of analysis
  - Data source
  - Numerator/Denominator statement
  - Testing information

  - *Testing information is needed for all new measures and maintenance measures that have indicated updated testing data.*

- All information must be submitted by the Intent to Submit deadline

Scientific Methods Panel

Established to allow consumers, patients, purchasers, and other members of NQF standing committees to focus on bringing their expertise to the subject matter under consideration and be more engaged throughout the evaluation process.

- Charge

  1. Conduct evaluation of complex measures for the criterion of Scientific Acceptability; and

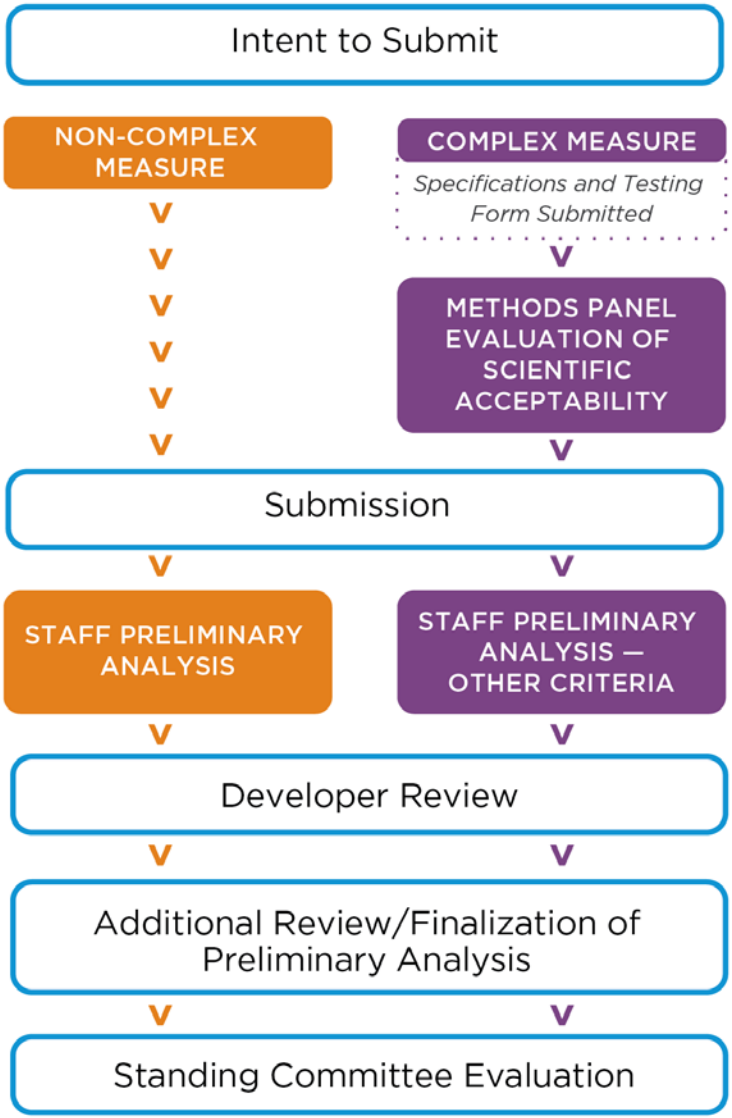
  2. Serve in an advisory capacity to NQF on methodologic issues, including those related to measure testing, risk adjustment, and measurement approaches.

- Webpage:

  https://www.qualityforum.org/Measuring_Performance/Scientific_Methods_Panel.aspx

- Commenting period on proposed roster open until September 20.

Measure Workflow



NATIONAL QUALITY FORUM

2017 Changes to NQF Evaluation Criteria and Guidance

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