NATIONAL QUALITY FORUM

CDP Redesign: Scientific Methods Panel's 2018 In-Person Meeting Update

June 4, 2018

Methods Panel Charge

- Conduct evaluation of complex measures for the criterion of Scientific Acceptability, with a focus on reliability and validity analyses and results

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

Scientific Methods Panel: Objectives for the May 16, 2018 In-Person Meeting

- Review current Methods Panel (MP) processes and discuss potential improvements
- Discuss conceptual definitions: Reliability and Validity
- Discuss potential changes to NQF measure evaluation criteria and guidance
- Discuss next steps for the panel

The Process

Current Process

- A minimum of three panel members independently evaluate each measure
  - *Assignments based on expertise, availability, need for recusal, other assigned measures*
  - *NQF provides a standard evaluation form that mirrors the rating algorithms*
- The majority recommendation from the three evaluations serve as the overall assessment of reliability and validity

Current Process

- If there is substantial disagreement in the ratings between the three reviewers, the Panel co-chairs evaluate the measure and determine the overall recommendation
  - *Requires substantial NQF staff time*
  - *More than expected need for co-chair evaluation*
  - *Measures with LOW/INSUFFICIENT rating sent back to developers, not to SCs*
- NQF staff compile the Methods Panel's ratings, evaluation, and commentary on reliability and validity and provide it to NQF's standing committees
  - *Meant to inform SCs endorsement decision*
  - *SCs can overturn the Scientific Methods Panel ratings*

Lessons Learned and Course Corrections to Date

- More information needed for evaluation
  - *For maintenance measures, we now provide a summary of the last evaluation*
  - *Now providing Feasibility Scorecard (for eCQMs)*
  - *Will provide full measure specifications (fully implemented by Fall 2018)*
- Difficulties with the evaluation form
  - *MP members have had trouble navigating the form*
    - » Some revisions made between Fall and Spring cycles (revised directions; implemented continuous numbering; reordered questions)
  - *Desire (by many) for more, not less, MP feedback provided as part of the evaluation*

Lessons Learned and Course Corrections to Date

The evaluation process

- Completely independent evaluations not working as desired
  - *Allow for informal discussions between evaluators (phone or e-mail), but still require separate evaluations*
- Need for extensive review by NQF staff to ensure consistency
  - *Incorporating phone calls as needed*
- Some additional guidance needed
  - *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 still 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

Year One Performance Metrics

Metrics	Fall 2017	Spring 2018
Total number of complex measures submitted for evaluation by the Scientific Methods Panel (SMP)	8 (7 new)	21 (9 new)
Unanimous "pass"	*2*	*TBD*
Unanimous "did not pass"	*1*	*TBD*
Split decision: co-chairs arbitrated	*5 (63%)*	*13 (62%)*
Total number of complex measures that received "low" or "insufficient" ratings from the SMP (i.e., did not go to SC)	4 (50%)	13 (62%)
Percent of time the standing committees were in agreement with the Scientific Methods Panel's recommendations	75%	TBD
Percent of time the Scientific Methods Panel's recommendations were overturned by standing committees	25%	TBD
Average turnover rate of SMP membership	0%	0%

After 2 Cycles: Key Challenges with the Process

- Lack of consensus on measures
  - *Excessive burden on reviewers, co-chairs, and staff → delays*
  - *Confusing handoff to Standing Committee*
  - *Proposed solutions:*
    - » Option #1: Keep process as is, but with relatively minor changes
    - » Option #2: Shift to group discussion/decision

- Dissatisfaction with the evaluation form
  - *MP members don't necessarily like the form*
  - *Desire (by many) for more, not less, MP feedback*
  - *Proposed solutions:*
    - » Option #1: Keep the form as is with minor changes
    - » Option #2: Go to mostly free-text (what staff used to do)
    - » Option #3: Something in the middle (mostly free-text, but some check-boxes)

Panel Recommendations to Improve Process

- For the process: Suggested "Option 1.5"
  - *Learning from each other is a key "ask"*
  - *Consensus from a larger group would be better*
  - *HOWEVER, impossible to evaluate up to 50 measures in a 2-3 day in-person meeting*
    - » Might work if they discuss only those measures where there is disagreement
    - » BUT there are "cons" to doing it this way
  - *Regardless, they agreed we need earlier resolution*
  - *Wanted more info from us about final ratings, points of disagreement*

- For the evaluation form: Preferred option #3

Reliability and Validity

Reliability

- Conceptual definitions (i.e., repeatability and precision) still okay—but may need some wordsmithing on formal definitions
- Importance of repeatability of the measure score
  - *But not completely sure if this is "have to have" or "nice to have"*
- Recommendation to require score-level testing for all measure types
- Feedback about submissions
  - *Quality is highly variable*
  - *More detail about methods needed*

Validity

- Conceptual definition mostly okay
  - *May need to add more about whether the measure actually measures "what it intends to measure"*
    - » Recommended submissions gather rationale/quality construct for all measures
    - » MP should give insight but clinical/content expertise also needed
- Need for insight into the extent to which a higher score on the measure actually reflects higher quality
  - *Discussed meaningful differences (not just statistical differences)*
  - *MP should give insight but clinical/content expertise also needed*
  - *Recommended asking for power calculations*
- MP can still reject based on statistical results

A Few Parking Lot Issues

- Data elements for multi-item scales: another level of testing needed?
- High (or low) prevalence: how does this affect reliability?
- How should we think about small n's and reliability?
- Differentiating between quality improvement and accountability

Evaluation Criteria Discussion

Panel Input

- Should validity be considered before reliability?
  - *No preference – both are needed*
- Okay with minor updates to the algorithms

Next Steps for the Panel

Next Steps

- Methods "Toolkit"
  - *Definitions of important terms*
  - *Descriptions of methods for demonstrating reliability and validity*
  - *Guidance on best methods for different measure types*
  - *"Thresholds" or acceptable results (or maybe rules of thumb)*
- "Toolkit" article in peer-reviewed journal
- Maybe an earlier "Perspective" article

Discussion Questions

CSAC Discussion Questions

- Do you have any feedback on the overall implementation of the MP to date?
- Do you have any input regarding potential changes to the MP process?
- What do you think about the recommendation for requiring score-level reliability testing for all measures?
- What do you think about potentially asking for data element validity testing for all measures?
  - *Could allow submission of justification of why it's not possible*
- What do you think about submitting to a peer-reviewed journal?
  - *Methods only? A perspective piece? Both? Neither?*