NATIONAL QUALITY FORUM

Cost and Efficiency Strategic Topical Webinar: *Validity Testing*

March 19, 2020

Matthew Pickering, Senior Director
Janaki Panchal, Project Manager
Funmilayo Idaomi, Project Analyst
Taroon Amin, Consultant
Ashlie Wilbon, Senior Technical Expert

Welcome

Project Team

- Matthew Pickering, Senior Director
- Janaki Panchal, Project Manager
- Funmilayo Idaomi, Project Analyst
- Taroon Amin, Consultant
- Ashlie Wilbon, Senior Technical Expert

Agenda for the Call

- Welcome and Roll Call
- Overview of NQF's Validity Criterion
- Validity Testing in Cost Measurement
- Next Steps

Cost and Efficiency Standing Committee

Cheryl Damberg, PhD (Co-chair)
Sunny Jhamnani, MD (Co-chair)
Kristine Martin Anderson, MBA
Robert Bailey, MD*
Bijan Borah, MSc, PhD*
John Brooks, PhD*
Cory Byrd*
Michael Chernew, PhD*
Amy Chin, MS*
Lindsay Erickson, MPH*
Troy Fiesinger, MD, FAAFP
Emma Hoo*
Sean Hopkins, BS*
Rachael Howe, MS, BSN, RN
Donald Klitgaard, MD, FAAFP*
Lisa Latts, MD, MSPH, MBA, FACP
Jason Lott, MD, MHS, MSHP, FAAP
Alefiyah Mesiwala*
Jack Needleman, PhD
Janis Orlowski, MD, MACP
John Ratliff, MD, FACS, FAANS
Mahil Senathirajah*
Srinivas Sridhara, PhD, MHS
Danny van Leeuwen, RN, MPH*

*Newly appointed Cost and Efficiency Standing Committee member

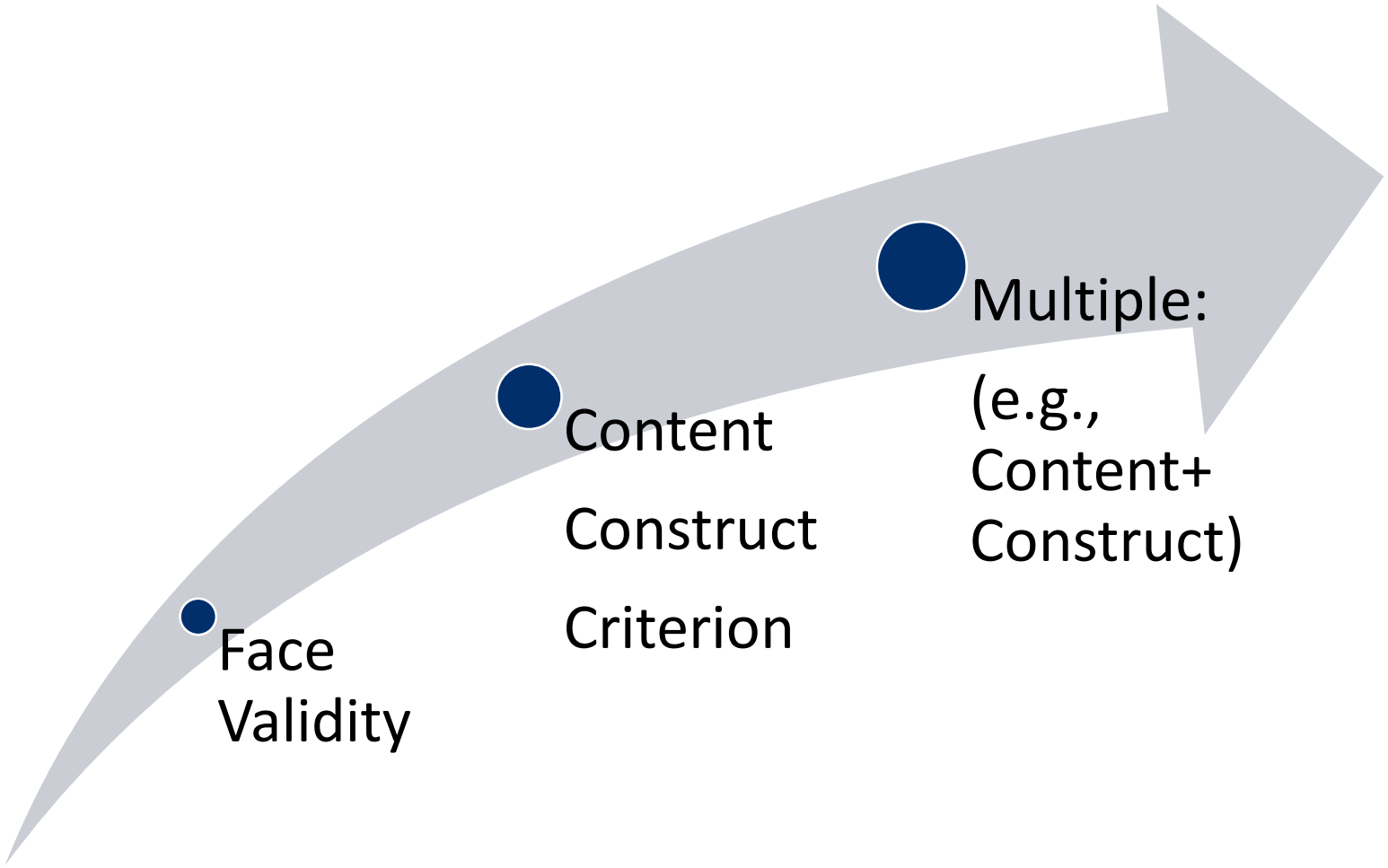
NQF Measure Evaluation Criteria

- **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. (**must-pass**)
- **Scientific Acceptability of Measure Properties:**
  - *Reliability and* ***Validity****: **Extent to which the measure, as specified, produces** consistent (reliable) and **credible (valid) results about the costs or resources to deliver healthcare (must-pass)***
- **Feasibility**: Extent to which the required data are readily available, or could be captured without undue burden, and can be implemented for performance measurement.
- **Usability and Use** (**must-pass** for maintenance measures): 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.
- **Comparison to related or competing measures**

What Is Validity?

- Correctness of measurement (data/data element testing)
- 4 Types of Validity
  - *Construct*
    - » Assesses how the measure performs based on the theory or intent of the construct.
  - *Content*
    - » Assesses whether a measure is representative and captures what it aims to measure
  - *Criterion*
    - » Assesses the correlation of the computed measure score against some criterion determined to be valid (relationship to other measures)
      - *Includes concurrent and predictive*
  - *Face*
    - » Assesses whether the measure appears to measure what it is intended to measure

Demonstrating Validity



Validity Testing

- 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.
- 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 with a "gold standard"

Threats to Validity

- Conceptual
  - *Measure focus is not a relevant outcome of healthcare or not strongly linked to a relevant outcome*
- Unreliable
  - *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 different data sources/methods
- Systematic missing or “incorrect” data (unintentional or intentional)

NQF Validity Criterion

- Demonstrating validity: The measure data elements are correct and/or the measure score correctly reflects the cost of care, adequately identifying differences in cost.
- NQF requirements for validity testing
  - *Data element OR measure score validity testing is acceptable; some measure types require more (e.g., PRO-PM)*
    - » Face validity of measure score is acceptable for initial submission
    - » Empirical validity testing is required for maintenance review
    - » If data element validity is performed and acceptable, reliability testing is not required
    - » NQF does not specify requirements for the type or how many different types of validity testing

NQF Validity Criterion Elements

- Validity testing – data elements or measure score
- Justification of exclusions
- Risk adjustment – calibration, discrimination, risk factors
- Identification of differences in performance
- Comparability of data sources/methods
- Missing data

Committee Discussion

- What is the goal of validity testing for cost/resource use measure construct?
  - *Is it different from other types of measures?*

Demonstrating Validity: Data Element

- Example: Validating administrative claims data elements, where codes are used to represent the primary clinical data (ICD, CPT, CPT-II/G)
  - *Validity of coded data from claims as compared to some criterion authoritative source of the same data*
  - *Analysis of agreement using appropriate statistical analyses (e.g., sensitivity, specificity, positive predictive value, negative predictive value with:*
    - » some other source of the same information considered to be valid (e.g., original data collection such as survey or observation, vital statistics)
    - » manual abstraction from the full medical record as the authoritative source

Demonstrating Validity: Construct

- How the measure performs based on the theory or intent of the construct
- Example (measure score level):
  1. *Tested correlation of cost with other known indicators of resource or service utilization (e.g., complications)*
  2. *Compared the ratio of observed over expected spending for cost measure episodes with and without complications related to care episodes occurring in the post-trigger period.*
  3. *This analysis sought to confirm the expectation that variation in service utilization is captured by the cost measure (i.e., cost goes up when utilization goes up)*

Demonstrating Validity: Content

- Assesses whether a measure is representative and captures what it aims to measure (e.g., does the measure accurately capture the targeted clinical population?)
- Example: (not a cost measure example)
  - *Record eligibility assessment: Six hospitals participating in the registry reported all TAVT and Mitral cases performed at their facility during a specified time frame. These were compared with those records included in the registry to verify that cases were not missed. N=366 records*

Demonstrating Validity: Criterion

- Assesses the correlation of the computed measure score against some criterion determined to be valid (relationship to other measures)
  - *Includes concurrent and predictive*
- Example (measure score level)
  - *Examined correlation of one clinician-based cost measure with another cost measure aimed to estimate resource use by clinicians.*
  - *Compared clinician scores on the measure under consideration with scores for the comparator measure in the same time frame for the same clinician.*
  - *Hypothesis: Positive correlation among the cost measures if they succeed at measuring resource use.*
  - *Acknowledged overlap between these measures as both the measures assess clinician performance in providing care in acute inpatient hospital settings, but stated that, while overlap exists, these measures intend to capture distinct populations of patients or services for each clinician; as such, did not expect to see extremely high correlations in analysis results.*

Demonstrating Validity: Face

- Assesses whether the measure appears to measure what it is intended to measure
- Example (measure score level)
  - *To systematically assess face validity, we surveyed the Technical Expert Panel (TEP) and asked each member to rate the following statement using a six-point scale (1=Strongly Disagree, 2=Moderately Disagree, 3=Somewhat Disagree, 4=Somewhat Agree, 5= Moderately Agree, and 6=Strongly Agree): "The Hip/Knee Payment measure as specified will provide a valid assessment of the relative costs of a 90-day hip/knee arthroplasty episode of care for Medicare patients admitted to a given hospital."*
  - *Among the 13 of 15 TEP members who provided a response, two responded "Somewhat Agree," six responded "Moderately Agree," and five reported "Strongly Agree" that this measure provides a valid assessment of payments for Medicare patients for a 90-day THA/TKA episode of care, removing policy adjustments unrelated to care decisions, risk-adjusting based upon case mix, and providing CMS with a tool it can use to compare payments across hospitals and identify hospitals with notably higher and lower payments.*

Committee Discussion

- Are these validity testing approaches commonly used by developers answering the questions we need to assess validity?
  - *How would you interpret them and what do they tell us?*
- Are there other examples that would demonstrate validity (of any type)?

Validity Testing in Cost Measures

Validity Challenges with Cost and Resource Use Measure Evaluation

- Validity concerns raised by SMP in Spring 2020 cycle
  - *Claims-based measures should not be validated with other claims-based measures, as that can lead to validating systematic inaccuracies in the claims data (threat to content validity)*
  - *Testing completed validates the measure construct, but does not validate clinical accuracy or representation of the appropriate clinical populations in the measure*

Committee Discussion

- How should developers be approaching validity testing? What methods or approaches should be considered?
  - *Are these approaches feasible given data availability and resources typically available for measure testing/ development?*
  - *Is it different for clinically focused episode measures vs. total cost (non-condition-specific) measures?*
    - » Procedure-based vs. chronic conditions?

Next Steps

Next Steps

- Availability survey for Spring 2020 measure evaluation meetings
- Six cost measures were submitted by CMS for Spring 2020 cycle; all will be evaluated by SMP prior to standing committee review; results will be shared in April

Project Contact Info

- Email: efficiency@qualityforum.org
- NQF phone: 202-783-1300
- Project page: http://www.qualityforum.org/Project_Pages/Cost_and_Efficiency.aspx
- SharePoint site: http://share.qualityforum.org/Projects/costEff/SitePages/Home.aspx

Questions?

Thank you.