NATIONAL QUALITY FORUM

Scientific Methods Panel 2019 In-Person Meeting

Karen Johnson
Ashlie Wilbon
Andrew Lyzenga
Michael Abrams
Sam Stolpe

June 11, 2019

Welcome, Introductions, and Review of Meeting Objectives

Scientific Methods Panel Members

- David Cella, PhD, (Co-chair)
- David Nerenz, PhD (Co-chair)
- J. Matt Austin, PhD
- Bijan Borah, MSc, PhD
- John Bott, MBA, MSSW
- Lacy Fabian, PhD
- Marybeth Farquhar, PhD, MSN, RN
- Jeffrey Geppert, EdM, JD
- Laurent Glance, MD
- Sherrie Kaplan, PhD, MPH
- Joseph Kunisch, PhD, RN-BC, CPHQ

Scientific Methods Panel Members (continued)

- Paul Kurlansky, MD
- Zhenqiu Lin, PhD
- Jack Needleman, PhD
- Eugene Nuccio, PhD
- Jennifer Perloff, PhD
- Sam Simon, PhD
- Michael Stoto, PhD
- Christie Teigland, PhD
- Ronald Walters, MD, MBA, MHA, MS
- Susan White, PhD, RHIA, CHDA

Meeting Objectives

- Reach consensus on methodologic issues related to reliability
- Reach consensus on methodologic issues related to validity
- Discuss potential recommendations for changes to the NQF measure evaluation criteria and guidance
- Discuss ideas for disseminating panel recommendations
- Discuss updates on panel processes
- If time allows, provide a brief informational update on other methodologic efforts at NQF

Measures Update

Performance Metrics – Spring 2019

- 47 measures evaluated
- 25 measures discussed on calls (53% of total)
  - *17 where consensus wasn't initially reached*
  - *3 pulled by panelists for discussion*
  - *5 pulled by staff for discussion*
- Final results
  - *Passed, will go to SCs: n=30 (64%)*
  - *Consensus not reached, will go to SCs: n=6 (13%)*
  - *Did not pass, will not go to SCs: n=11 (23%)*

Rationale for Spring 2019 Measures that Did Not Pass

- Required testing not conducted
- Testing methodology unclear
- Testing methodology inappropriate
- Inadequate data in testing sample (e.g., too few states for population-based measures)
- Inadequate/low testing results (reliability)
- Lack of risk adjustment

Performance Metrics

Metrics	Fall 2017	Spring 2018	Fall 2018	Spring 2019
Total number of complex measures submitted for evaluation by the Scientific Methods Panel (SMP)	8 (7 new)	21 (9 new)	39 (21 new)	47 (19 new)
Unanimous "pass"	*2*	*4*	*17*	*19*
Unanimous "did not pass"	*1*	*4*	*2*	*2*
Split decision: co-chairs arbitrated	*5*	*13*	*n/a*	*n/a*
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%)	10 (26%)	11 (23%)
Percent of measures where the standing committee ratings aligned with SMP recommendations	75%	100%	23/29 (79%)	TBD
Percent of measures where the standing committee ratings did NOT align with SMP recommendations	25%	0%	6/29 (21%)	TBD
Average turnover rate of SMP membership	0%	0%	4%	4%

SMP Internal Disagreement

- Initial evaluations
  - *Fall 2017:* *63%*
  - *Spring 2018:* *62%*
  - *Fall 2018:* *51%*
  - *Spring 2019:* *55%*
- During calls (CNRs that go to SCs)
  - *Fall 2018:* *10%*
  - *Spring 2019:* *13%*

SC Nonalignment with SMP Ratings Results from Fall 2018 To Date*

* Note that post-comment discussions have not yet occurred

Measure	SMP decision	SC decision (to-date) and concerns
3456	Passed	Did not pass validity Concerns due to lack of clinical risk adjustment (in part due to small sample size) and concern that differences in the measure reflect differences in underlying populations being compared, rather than differences in quality of care. NOTE that these concerns were raised by the SMP.
3366	Passed	CNR on validity Concerns on the results of the risk-adjustment model (c-statistic=0.61) and possibly, lack of adjustment for social risk (dual status)

SC Nonalignment with SMP Ratings Results from Fall 2018 To Date*

* Note that post-comment discussions have not yet occurred

Measure	SMP decision	SC decision (to date) and concerns
3443	CNR on validity	Did not pass validity SC not convinced that risk-adjustment strategy accounted for differences in underlying populations being compared
3445	CNR on validity	Did not pass validity SC not convinced that risk-adjustment strategy accounted for differences in underlying populations being compared
0964	CNR on validity	Passed validity SMP divided on relatively low correlation results for validity; the SC did not have these concerns
0753	CNR on reliability	Passed reliability SMP divided on relatively low reliability estimates; the SC agreed that results were low but decided to pass the measure, given lack of reliability thresholds

Fall 2019 Measure Evaluation Updates

- Complex measures scheduled for maintenance: ~25
  - *New measure submissions: TBD*
- Expiring of terms for half of SMP
  - *Members can elect to continue participation for 2 years (4 cycles)*
  - *Complete brief survey by the end of the day to indicate your decision*
- Expansion of the SMP to ~40 members
- Anticipate a (in-person) measure evaluation meeting during week of October 21 *(pending approval)*
  - *Specific dates to be determined*

Update on White Papers

- NQF Guidelines for Evaluating the Scientific Acceptability of Risk-Adjusted Clinical Outcome Measures
- The NQF Scientific Methods Panel: Enhancing the Review and Endorsement Process for Performance Measures
- Reliability paper (Title TBD)
  - *Stage of writing/Current status*
  - *Plans for submission*

Reliability

Criterion #2: Reliability–Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) results about the quality of healthcare delivery

2a. Reliability (must-pass)

2a1. Precise specifications including exclusions

2a2. Reliability testing—data elements or measure score

Definitions of Reliability

- Repeatability (consistency, reproducibility, stability)
- Precision

	Repeatability	Precision
Data element	X	
Performance measure score		X

- Data Element Reliability: Repeatability and reproducibility of the data elements for the same population in the same time period
- Measure Score Reliability: Precision: Proportion of variation in the performance scores due to systematic differences across the measured entities (signal) in relation to random error (noise)

Current Assumptions about Reliability

- There will always be some error in performance measurement
  - *Random error affects reliability; systematic error affects validity*
- Reliability is **not a static property of a measure** (it can vary under conditions of implementation)
- Reliability is not an all-or-none property and is instead a matter of degree
  - *Considerations are scope of testing, method used, and results obtained*
- Reliability does not guarantee validity

Some Basic NQF Terminology

- Data elements – building blocks of a measure; "variables" used to calculate a measure
  - Examples include diagnosis codes, medications, admission date, birth date, questions/items from surveys

- Measure score – the computed results of the measure
  - Examples include rates, averages, proportions

Reliability Testing – Data Element

- Reliability of the **data elements** refers to the repeatability/ reproducibility of the data for the same population in the same time period
  - *Common Approaches*
    - » inter-rater/abstractor or intra-rater/abstractor studies
    - » **internal consistency for multi-item scales**
    - » **test-retest for survey items**
- Current NQF Guidance
  - *All critical data elements must be tested (not just agreement of one final overall computation for all patients).*
    - » At a minimum, the numerator, denominator, and exclusions (or exceptions) must be assessed and reported separately.

Reliability Testing — Data Element

Panel Considerations:

- Data element:
  - *Cronbach's alpha*
    - » Exception for single item factors?
  - *Test-retest*
  - *Kappa (inter- or intra-rater)*

Reliability Testing – Measure Score

- 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).
- Current focus of NQF criteria is ***precision*** of the measure score
- Other definitions: e.g., stability with time?

Reliability Testing Measure Score

Considering some of the common approaches for demonstrating of reliability:

- Distinguishing differences and demonstrating accurate classification (signal-to-noise), typically between providers.
- Split half with ICC vs Split half with Pearson's or rank ordered correlations
- Split half with assessment of providers movement across quintiles
- Bootstrapping

Panel Consideration:

- Is it necessary to perform more than one test to adequately demonstrate reliability of the measure score? Are any of these approaches sufficient to demonstrate reliability on their own?
- Are the scores from these various tests comparable (i.e., ICC 0.7 to IUR)? How should they be interpreted in relationship to each other?
- What guidance can we provide about when/how to use such methods including selection of an approach and interpretation of results?

Reliability Testing — Measure Score

Panel Consideration

- Is there agreement that data element reliability cannot supplant the need for score level reliability?
- Measures that address rare events face specific challenges with demonstrating reliability (low R. estimates with due to small sample size, significant variation)
  - *Can rates of rare events be measured reliably?*
  - *What are some methodological approaches that should be considered by developers when testing these measures?*

Reliability

Additional issues for consideration to address in the reliability white paper

- Anchoring concept: misclassification
  - *References: Adams and Paddock, c. 2016; Zaslavsky, 2001*

Questions to Consider

- Repeatability, consistency, reproducibility, stability: Are these interchangeable? Should we pick one or two?
- The idea of stability of the measure score as an important facet of reliability is new to NQF.
  - *Compared to the ability to distinguish differences, is stability as important? Less? More?*
  - *Would you expect to see <u>both</u> types of analysis for score-level testing?*
- Is it useful or helpful to use the term "signal to noise" when talking about score-level reliability?
  - *Why or why not? When?*

Questions to Consider

- Any recommendations regarding "signal-to-noise" reliability estimates in submissions?
  - *Mean and variance (or other stats such as median, percentile values, IQR, etc.)*
  - *Stratified by sample size*
- Any statement regarding signal-to-noise testing that is limited to providers with a minimum sample size?
- Other recommendations for submissions?
  - *Examples include: sample size calculations; when testing multiple samples, average shift in rank or proportion of units in the top or bottom quintile, along with distribution of differences; standard error of measurement*

Break

NQF Member and Public Comment

Lunch Break

Validity

Criterion #2: Validity – Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces credible (valid) results about the quality of health care delivery

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*

Conceptual Definition of Validity

- The correctness of measurement
  - *The extent to which one can draw correct conclusions about a particular attribute based on the results of a measure*

- The extent to which a measure assesses what it intends to measure

Cronbach (1971) quote => context matters

"The phrase 'validation of a test' is a source of much misunderstanding. On validates, not a test but an 'interpretation of data arising from a special procedure.' A single instrument is used in many different ways—Smith's reading test may be used to screen applicants for professional training, to plan remedial instruction…etc. Since each application is based on a different interpretation, the evidence that justifies one application may have little relevance to the next. Because every interpretation has it own degree of validity, one can never reach the simple conclusion that a particular test 'is valid'."

(Traub and Rowley, Applied Psychological Measurement, 1980)

Current Definitions: Data Element and Measure Score Validity

- Data Element Validity
  - *Correctness of the data elements as compared to an authoritative source*
- Measure Score Validity
  - *Correctness of conclusions about quality that can be made based on the measure score (i.e., a higher score on a quality measure reflects higher quality)*

	Accuracy (at patient level)	Correct conclusion about performance (at provider level)
Data element	**X**	
Performance measure score		**X**

Validity Testing — Data Element

- Empirical testing
  - Data element – assesses the correctness of the data elements compared to a "gold standard"
  - Validity testing of data elements typically analyzes agreement with another authoritative source of the same information.

- Panel Consideration
  - Is accuracy the right word? Truth? True to concept? Meaningful as intended?
  - What guidance can be given to developers about how to contextualize and overall describe their validity testing approach, results, interpretation?

Validity Testing — Measure Score

- 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; or, 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
  - Assesses a hypothesized relationship of the measure results to some other concept; assesses the correctness of conclusions about quality

Validity Testing — Measure Score

Empirical Validity Testing

Common Approaches

- Known group validity: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method
- Construct validity: correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures).

Validity Testing — Measure Score

Construct Validity: Challenging Examples

- Comparing CAHPS measures to themselves
- Cost measures comparing to other claims-based measures with the same data elements
- Behavioral health (substance use disorder (SUD) screening versus depression and infectious disease screening) versus actual better SUD outcomes

Validity Testing — Measure Score

Panel Considerations:

- Face Validity
  - How should face validity be considered when empirical validity is also provided for new measures?
    - » Should the face validity still be considered as meeting the minimum requirement or should the focus of evaluation be on empirical testing?
    - » How should measure score validity be evaluated when data element validity testing is submitted in conjunction with face validity?
- Construct Validity
  - What are the best practices for selecting a comparator for demonstrating construct validity?
  - Is there additional guidance to define how the measure and its comparator should be conceptually related?

Evaluation of Validity

The evaluation of validity encompasses of the assessment of multiple elements, some of which are methodological and others are clinical

- Are there specific elements or subcriteria that should not solely be under the purview of the SMP?
- Are there elements that should be considered by the Standing Committee before a final vote on validity is rendered? Face validity? Justification of face validity for maintenance measures?
  - 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

SC Nonalignment with SMP Ratings Results from Fall 2018 To Date*

* Note that post-comment discussions have not yet occurred

Measure	SMP decision	SC decision (to-date) and concerns
3456	Passed	Did not pass validity Concerns due to lack of clinical risk adjustment (in part due to small sample size) and concern that differences in the measure reflect differences in underlying populations being compared, rather than differences in quality of care. NOTE that these concerns were raised by the SMP.
3366	Passed	CNR on validity Concerns on the results of the risk-adjustment model (c-statistic=0.61) and possibly, lack of adjustment for social risk (dual status)

SC Nonalignment with SMP Ratings Results from Fall 2018 To Date*

* Note that post-comment discussions have not yet occurred

Measure	SMP decision	SC decision (to date) and concerns
3443	CNR on validity	Did not pass validity SC not convinced that risk-adjustment strategy accounted for differences in underlying populations being compared
3445	CNR on validity	Did not pass validity SC not convinced that risk-adjustment strategy accounted for differences in underlying populations being compared
0964	CNR on validity	Passed validity SMP divided on relatively low correlation results for validity; the SC did not have these concerns
0753	CNR on reliability	Passed reliability SMP divided on relatively low reliability estimates; the SC agreed that results were low but decided to pass the measure, given lack of reliability thresholds

Validity — Risk Adjustment

Issues from risk-adjustment paper for Panel discussion

- General analytics/methods regarding risk-adjustment/missing data/exclusions
  - *Deciding if it is necessary*
  - *Correct methods*
    - » Simple stratification
    - » Multiple regression
    - » Hierarchical modeling
- How can sensitivity analyses be used to support validity?
- Risk adjustment's impact on reliability

Other Validity Issues

- Are there any assumptions about validity that should be questions, or facets of validity that we are missing?
  - *Assumption: to be valid, a measure must be reliable*
    - » Alternative way of thinking about this is that reliability and validity are two separate and distinct characteristics of performance measures

Evaluation Criteria Discussion

Criterion #2: Reliability and Validity — Scientific Acceptability of Measure Properties

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*

Current Evaluation Guidance — Reliability

- Beginning Summer 2019: Reliability of unstructured (i.e., not standardized) data must be demonstrated at the data element level.
  - *If data element testing is not possible, justification is required and must be accepted by the Standing Committee/Methods Panel*
  - *If sufficient data are available for testing, testing of reliability and validity at the score level is encouraged in addition to required data element testing.*
- Prior evidence of reliability of data elements for the data type specified in the measure (e.g., hospital claims) can be used as evidence for those data elements.

Current Evaluation Guidance — Reliability

Measure Type	Data Element Reliability	Measure Score Reliability
Instrument-based	**X**	**X**
Composite		**X**
eCQMs	**X**	
All other measures	**Either/Or**	

Current Evaluation Guidance — Reliability

- For some measure types, data element reliability is not required if data element validity testing is completed.

Current Evaluation Guidance — Validity

Measure Type	Data Element Validity	Measure Score Validity
Instrument-based	**X**	**X**
Composite		**X**
eCQMs	**X**	
All other measures	**Either/Or**	

Evaluation Criteria Considerations

- Should score level reliability testing be required for all measures? Should both data element AND score level testing be required for all measures?
- Should data element validity testing be required for all measures?

Break

Next Steps:
Methods “Toolkit” and Beyond

NQF "Toolkit"

- A resource at the NQF website for those seeking guidance regarding scientific acceptability methods
- Potential Items
  - *Definitions of important terms*
  - *Descriptions of methods, including when to use, how to deploy (software), and how to report and interpret.*
  - *Look-up charts for measure use, thresholds of acceptability (e.g., Landis), and other quick fact, with references.*
  - *Published literature that educates end users about methods.*
  - *"What good looks like" methods examples, per NQF submission forms.*
  - *General descriptions of NQF standards*
- **What is the starter set of materials that should be in the "Toolkit"? Top 3 items?**
- **Recommendations for organizing resources (e.g., based on criteria, measure type)?**

White Papers

In Progress

- NQF Guidelines for Evaluating the Scientific Acceptability of Risk-Adjusted Clinical Outcome Measures (Larry G. et al.)
- The NQF Scientific Methods Panel (David N. et al.)
- Reliability paper?

Next

- Patient-reported outcomes paper
- Others?

Discussion of Methods Panel Processes for Measure Evaluation

Process Improvement Efforts

- NQF staff has been iterating on process improvements since the SMP began
  - *Incremental improvements each cycle*
    - » Independent Reviews→ Subgroup calls
    - » Increasing developer involvement in SMP discussions
    - » Development of discussion guide
    - » CNR measures no longer go to Co-chairs for adjudication
  - *Goals*
    - » Reduce SMP burden, workload
    - » Reduce number of CNR measures that go to Committee
    - » Increase transparency of SMP process
    - » Improve efficiency of processes (i.e., reduce rework/re-evaluation of measures)
- NQF staff recently met for process improvement "workout" to assess the intent to submit period and identify areas for improvements

Intent to Submit Period SMP Evaluation

- Business Case/Problem Statements
  - There is a significant amount of process involved during the ITS process that has changed since the implementation of the process. It is unclear whether these steps are being implemented efficiently
  - There is lack of transparency around the SMP review process
  - It is unclear how the ITS period is used for managing non-SMP measures
  - Developers want more opportunities to respond to SMP comments
- Goal Statements
  - Improve the efficiency of staff, developer, and volunteer time to accomplish goals of ITS timeframe
  - Identify necessary improvements to the SMP process for fall 2019 implementation
  - Increase transparency of SMP review

Process Improvement Efforts

- Approach
  - *Stakeholder Surveys*
  - *Process Mapping*
  - *Eliminate Waste*
  - *Address problem statements*
- Survey of many stakeholder perspectives
  - *Measure developers*
  - *Standing committee members*
  - *Scientific Methods Panel (SMP) members*
  - *NQF members and public commenters*
  - *CSAC members*
  - *NQF CDP staff*

Number of Responses

Stakeholder group	Total # of respondents
Developers	10
SMP members	14
Standing committee members	52
CSAC members	7
Commenters	55*
NQF CDP staff	14

* Many did not complete the survey

Preliminary Results: Perception of SMP Value

Stakeholder group	Number of responses	Yes	No
Developers	9	22%	78%
SMP members	n/a	—	—
SC members	40	93%	8%
Commenters	43	79%	21%
CSAC members	5	*	*
NQF CDP staff	10	100%	0%

* Open-end: 3 of the 4 believe SMP is helpful

Preliminary Results: SMP's "Gatekeeper" Role

Stakeholder group	Number of responses	Keep as is	SC should evaluate	Other
SC members	36	86%	8%	6%

Preliminary Results: Comments on the SMP Gatekeeper Role

- NQF CDP staff (n=10)
  - *Protects rigor of the criteria*
  - *More work for SCs if changed*
  - *SCs don't want to evaluate measures that are not "ready"*
  - *Believe some SCs resent this role*
- Standing committees (n=36)
  - *No point in SC evaluation if fail SMP*
  - *Don't want SCs to recommend measures that don't have acceptable level of scientific rigor*
  - *I like having the SMP "weed out" measures*
  - *Want a clearly written justification for a "fail"*
  - *Can save SCs time and effort*

Preliminary Results:
Comments on the Value of the SMP to the CDP

- Developer:
  - *Extra level of review, more requirements, not transparent*
  - *Increased complexity, don't have consensus on what meets requirements, harder to pass R/V*
  - *SMP knowledge of survey-based measures inadequate*
  - *Another hurdle; dislike gatekeeper function*
- Commenters
  - *Not enough expertise on the SMP*
  - *Risk of biasing SC votes*
  - *High level of effort but incremental benefit*
  - *Complex measures face more scrutiny than noncomplex*
  - *SMP confused; lack of clarity in criteria and submission forms*

Preliminary Results: Comments on the Value of the SMP to the CDP

- Standing Committees
  - *Greater sense that measure has received appropriate scientific scrutiny*
  - *As a patient, I look entirely to the SMP for scientific acceptability*
  - *It helps focus the reviews*
  - *Helps SC discuss real issues, not basic entry requirements; greatly informative*
  - *Takes the heat off SC members who may not have needed expertise*
  - *SMP doing a very needed deep dive*
  - *SMP lacks context for measures*
  - *SMP doesn't always reach consensus*
  - *More detailed, but simpler, summary would be helpful*

Intent to Submit Period: Proposed Improvements – Measure Intake

Current Process	Improvements
• Staff identifies minor edits needed by developer prior to sending to SMP ➢ Developers have 48 hours to update submission	• Staff will no longer perform this review
• Staff identifies measures that do not meet minimum criteria for endorsement review and notifies developer	• No change

Intent to Submit Period: Proposed Improvements – SMP Structure and Transparency

Current Process	Improvements
• SMP currently comprised of 22 experts	• SMP to be expanded to 35-45 people*
• SMP is convened over a series of 8 conference calls divided amongst 4 subgroups	• SMP to meet in person 2x/year*
• Subgroup meeting agenda posted publicly	• All SMP meeting materials will be posted publicly
• No public commenting during conference calls	• Allow opportunity for public commenting at SMP meeting

*To be proposed in upcoming contract cycle

Intent to Submit Period: Proposed Improvements – Developer Engagement

Current process	Improvements
• Developers can only respond verbally to questions/concerns (additional documentation after submission is not permitted)	• Developers will have the opportunity to respond in writing to SMP preliminary analyses before final vote, and verbally during the meeting

Intent to Submit Period: Proposed Improvements – SMP Gatekeeper Role

Current process	Improvements
• Measures that are CNR and pass R/V are provided to committee for consideration and final recommendation	• Remains the same
• Measures that do not pass the SMP do not go to committee for discussion ➢ Only short summary of rationale for not passing is provided	• Committee members will have the opportunity to pull a measure that did not pass SMP for discussion; cannot revote on R/V ➢ Detailed SMP summary, measure information form, and testing attachment will be provided to committee

Improvements to SMP Process

Any recommendations to improve the process that have not already been addressed?

- What worked particularly well this last cycle?
- What should be a focus for improvement in fall 2019?

Brief informational update on other methodologic efforts at NQF

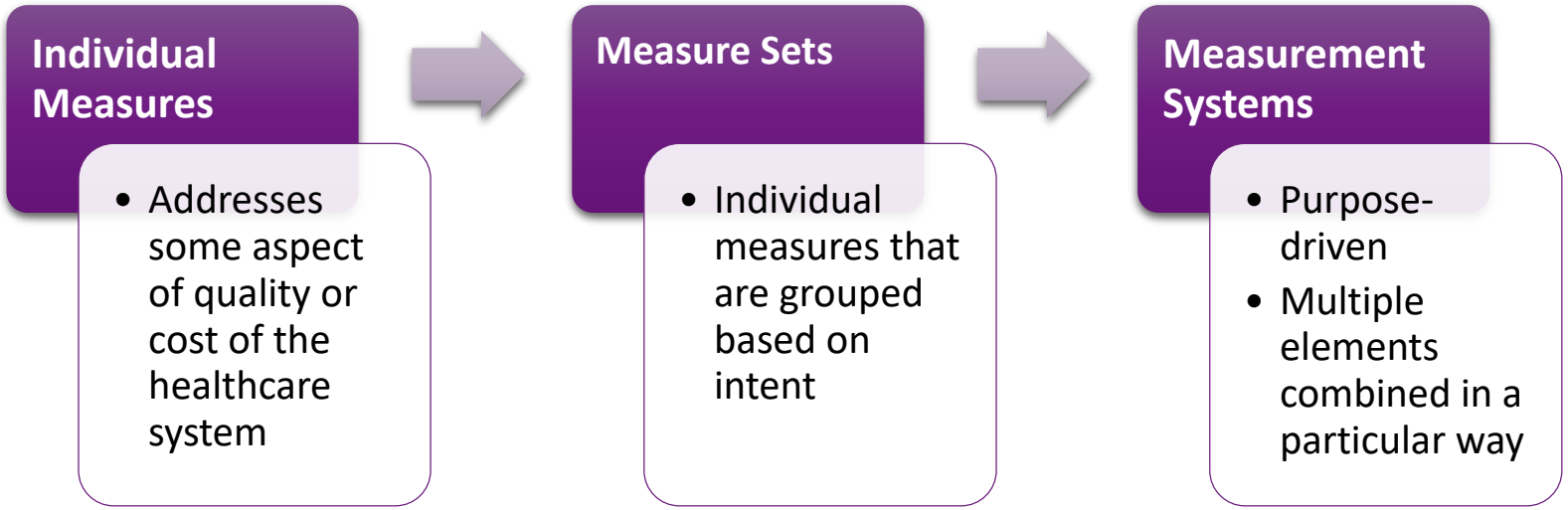
Other Methodological Efforts at NQF

- Disparities Project (http://www.qualityforum.org/ProjectDescription.aspx?projectID=80894)

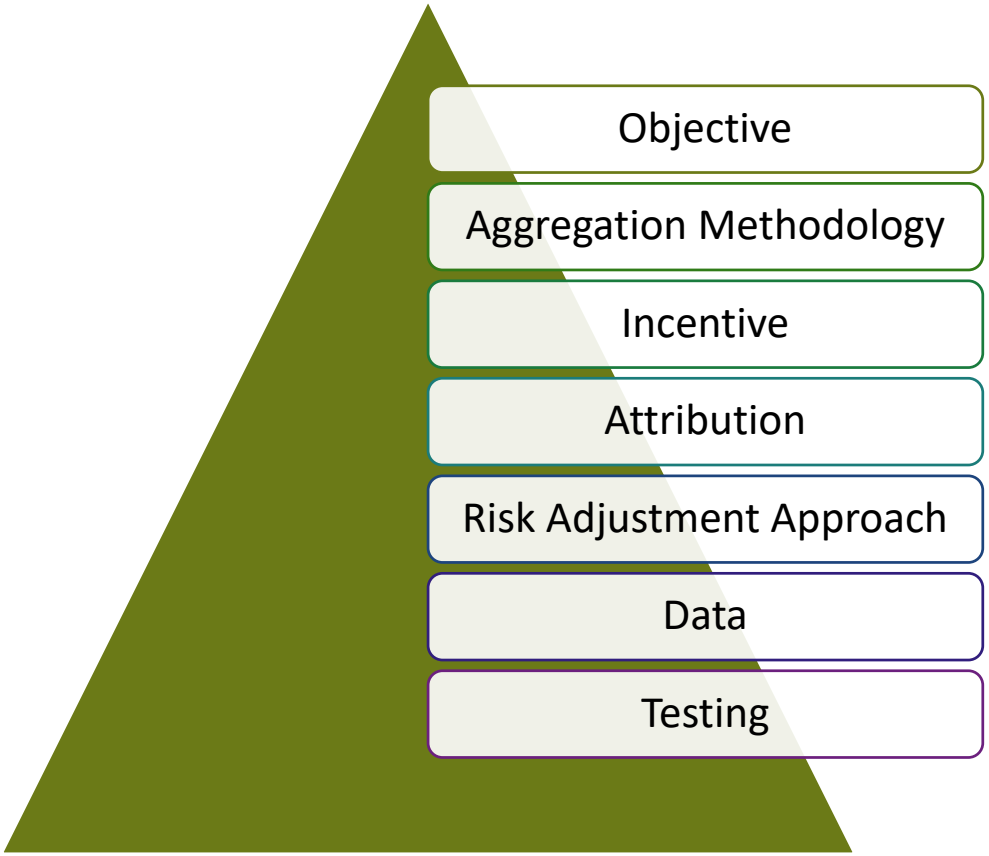
- Measure Sets and Measurement Systems (http://www.qualityforum.org/ProjectDescription.aspx?projectID=89799)

The Opportunity: A New Framework for a Quality Measurement Infrastructure

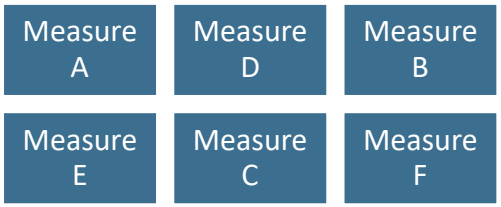
- Establish independent, transparent, and multistakeholder consideration of three interdependent levels of performance measurement:



What is a Measurement System?

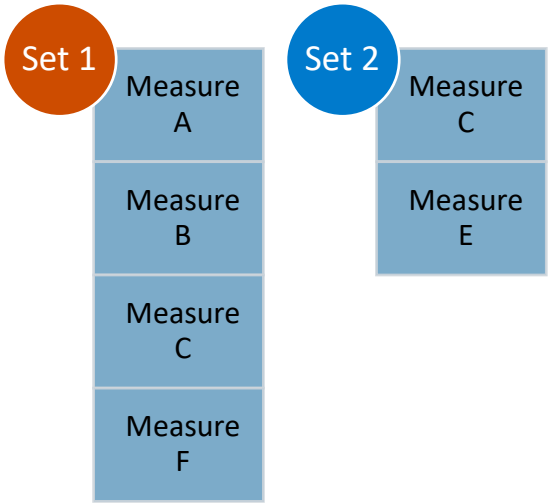


Conceptual Framework: Quality Measurement Infrastructure



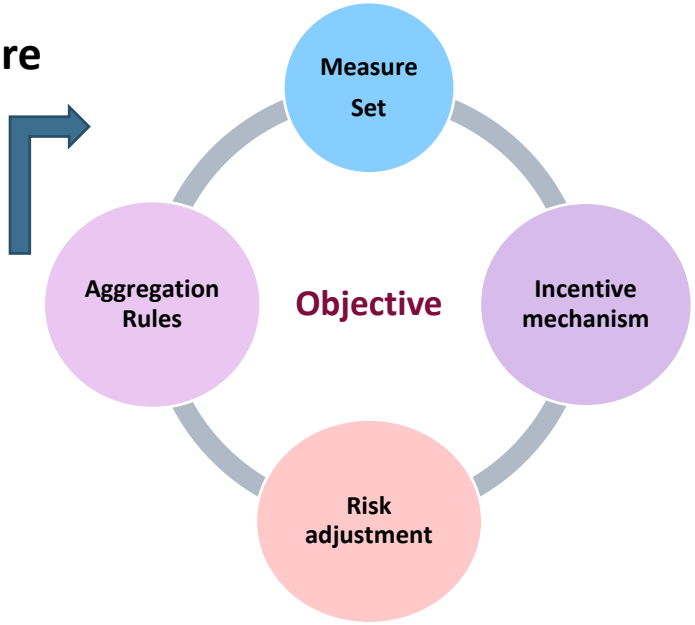
Level 1: Individual Measures

Example: NQF Measure 2158 Medicare Spending Per Beneficiary



Level 2: Measure Sets

Example: CMS Hospital Value-Based Purchasing Program (HVBP) uses a set of measures assessing four domains: safety, clinical care, efficiency and cost reduction, and patient experience.



Level 3: Measurement Systems

Example: Measures are aggregated as part of the HVBP program, and the score is used to determine hospital penalties or rewards.

NQF Member and Public Comment

Next Steps

Next Steps

- Monthly 1-hour calls
  - *Every 2nd Thursday of the month*
  - *Next call: July 11 at 3pm ET*
- Fall 2019 evaluation
  - *Measures will be distributed by September 3, 2019 (~4 weeks to evaluate assigned measures)*
  - *In-person meeting in October (approval pending)*
    - » Will determine availability over the next few weeks
- Contact information: methodspanel@qualityforum.org

Adjourn