NATIONAL QUALITY FORUM

Cost and Efficiency, Fall 2019 Cycle

Standing Committee Orientation
December 16, 2019

Ashlie Wilbon, Senior Director
Taroon Amin, Consultant
Janaki Panchal, Project Manager
Hannah Ingber, Project Analyst

Welcome

Project Team

- Ashlie Wilbon, Senior Director
- Janaki Panchal, Project Manager
- Hannah Ingber, Project Analyst
- Taroon Amin, Consultant

Agenda for the Call

- Standing Committee Introductions
- Overview of NQF, the Consensus Development Process(CDP), and Roles
- Overview of NQF's Portfolio of Cost and Efficiency Measures
- Overview of Measure Evaluation Process
- Cost and Resource Use Measure Evaluation Criteria Overview
- Social Risk Trial Overview
- SharePoint Tutorial
- 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
Srinivas Sridhara, PhD, MHS
Mahil Senathirajah*
Danny van Leeuwen, RN, MPH*

*newly appointed Cost and Efficiency Standing Committee member

Overview of NQF, the CDP, and Roles

The National Quality Forum: A Unique Role

Established in 1999, NQF is a nonprofit, nonpartisan, membership-based organization that brings together public and private sector stakeholders to reach consensus on healthcare performance measurement. The goal is to make healthcare in the U.S. better, safer, and more affordable.

Mission: To lead national collaboration to improve health and healthcare quality through measurement

- An Essential Forum
- Gold Standard for Quality Measurement
- Leadership in Quality

NQF Activities in Multiple Measurement Areas

- **Performance Measure Endorsement**
  - 600+ NQF-endorsed measures across multiple clinical areas
  - 15 empaneled standing expert committees
- **Measure Applications Partnership (MAP)**
  - Advises HHS on selecting measures for 20+ federal programs
- **National Quality Partners**
  - Convenes stakeholders around critical health and healthcare topics
  - Spurs action: recent examples include antibiotic stewardship, advanced illness care, shared decision making, and opioid stewardship
- **Measurement Science**
  - Convenes private and public sector leaders to reach consensus on complex issues in healthcare performance measurement
    - » Examples include HCBS, rural issues, telehealth, interoperability, attribution, risk-adjustment for social risk factors, diagnostic accuracy, disparities
- **Measure Incubator**
  - Facilitates efficient measure development and testing through collaboration and partnership

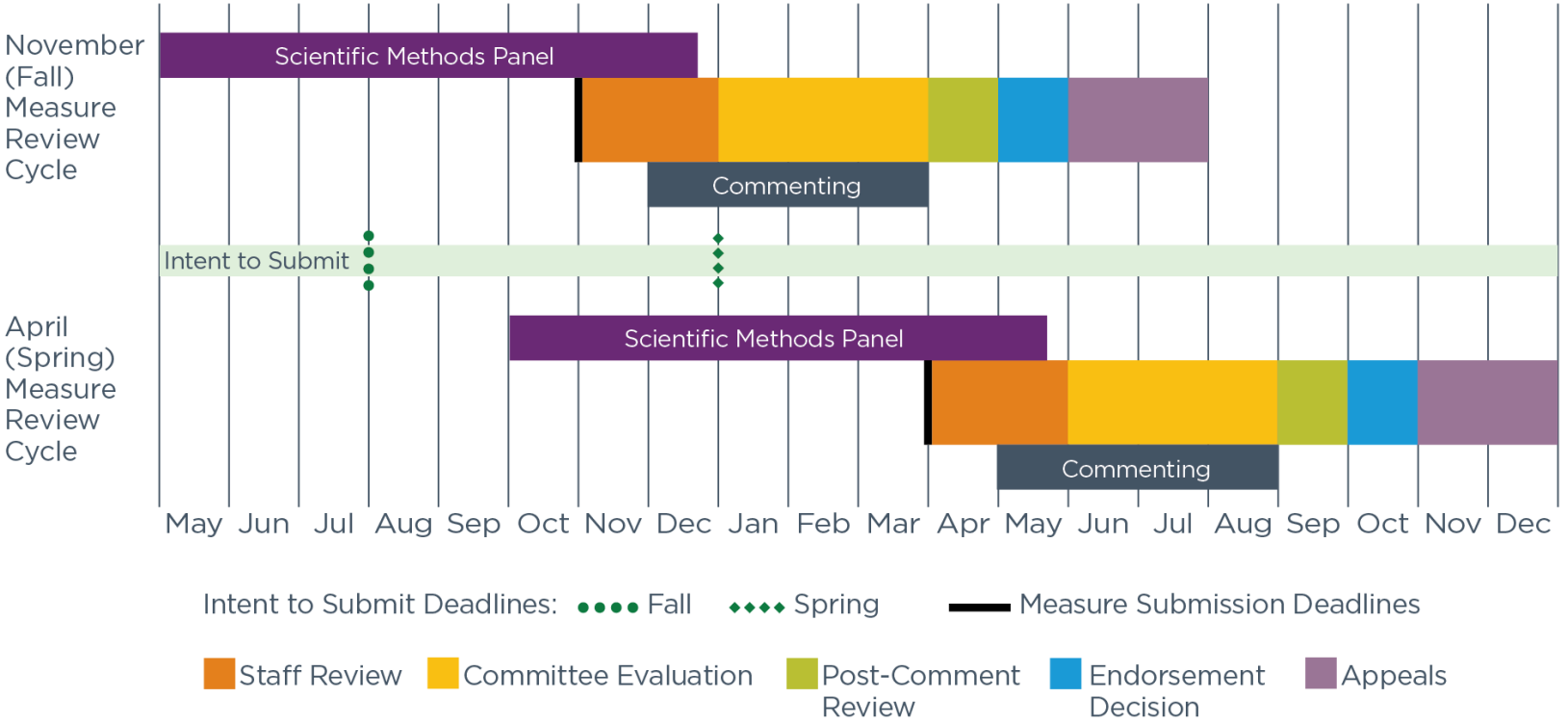
NQF Consensus Development Process (CDP)
6 Steps for Measure Endorsement

- Intent to Submit
- Call for Nominations
- Measure Evaluation
- Public Commenting Period with Member Support
- Measure Endorsement
  - *Consensus Standards Approval Committee (CSAC)*
- Measure Appeals

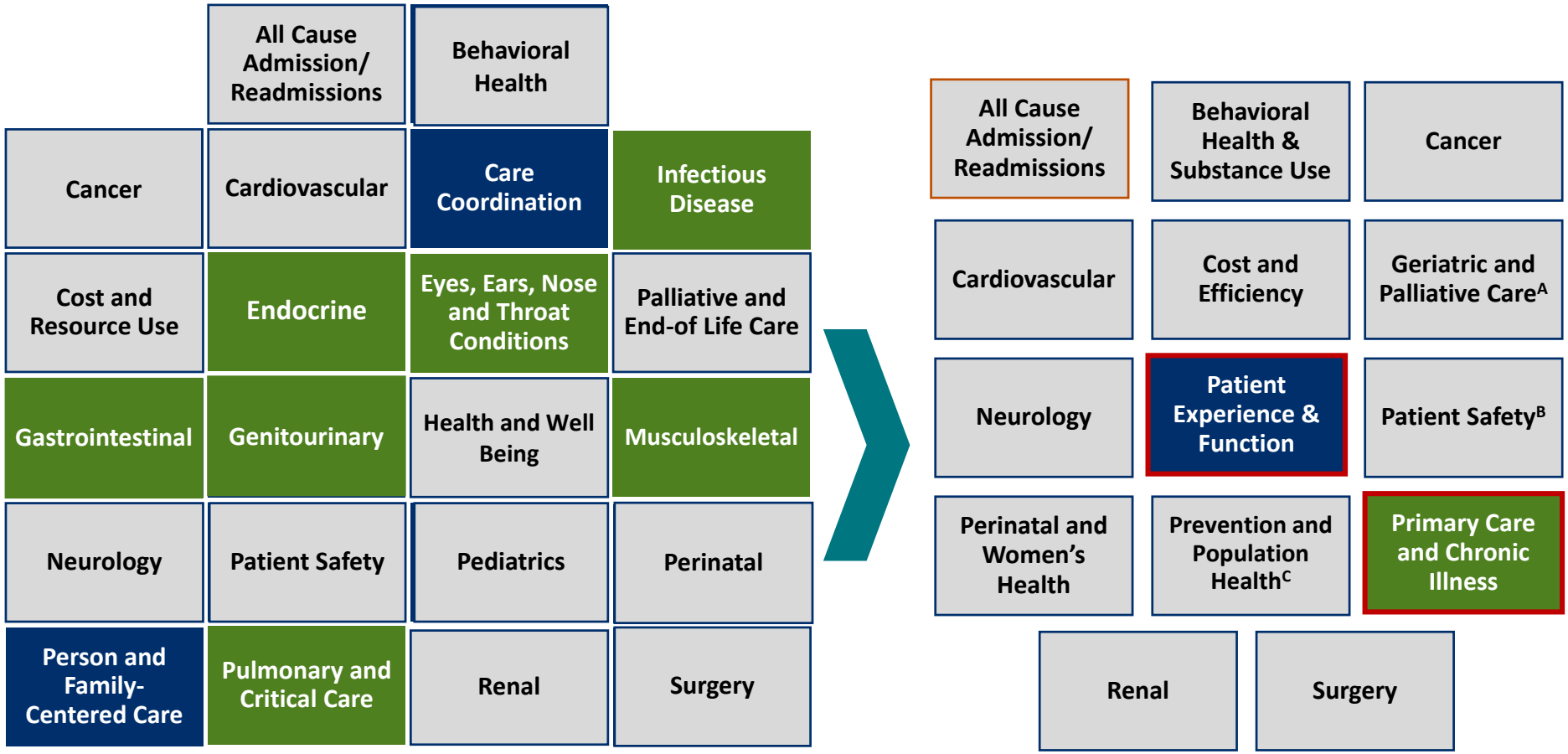
Measure Review: Two Cycles Per Year

Consensus Development Process:

Two Cycles Every Contract Year



14 Measure Review Topical Areas



[A]Geriatric & Palliative Care includes pain-focused measures from other domains
[B] Patient Safety will include acute infectious disease and critical measures
[C] Prevention and Population Health is formerly Health and Well Being

Role of the Standing Committee
General Duties

- Act as a proxy for the NQF multistakeholder membership
- Serve initial 2-year or 3-year terms
  - *Opportunity to renew for 2 additional years (4 cycles)*
- Work with NQF staff to achieve the goals of the project
- Evaluate candidate measures against the measure evaluation criteria
- Respond to comments submitted during the review period
- Respond to any directions from the CSAC
- Refer to the Standing Committee Guidebook for more information

Role of the Standing Committee
Meeting Participation

- Meeting attendance
  - *Must notify NQF staff if unable to attend in advance of the meeting*
- Quorum requirements
  - *NQF Quorum = 66% of active members*
  - *Committee recommendations can only be made with a quorum of Committee votes*
    - » Not based on Robert's Rules of Order
  - *Votes may be requested via email if quorum is not reached during the meeting*
    - » Materials (i.e., recording, transcripts) will be sent to inform votes
  - *Meetings may be cancelled (and rescheduled) if quorum not reached and vote is required*
- Measure-specific disclosure of interest
  - *Must be completed to participate in the measure evaluation discussion (each cycle)*

Role of the Standing Committee
Measure Evaluation Duties

- All members evaluate ALL measures being considered for endorsement
- Evaluate measures against each criterion
  - *Indicate the extent to which each criterion is met and rationale for the rating*
- Make recommendations to the NQF membership for endorsement
- Oversee Cost and Efficiency portfolio of measures
  - *Promote alignment and harmonization*
  - *Identify gaps*

Role of the Standing Committee Co-chairs

- Co-facilitate Standing Committee (SC) meetings with NQF staff
- Work with NQF staff to achieve the goals of the project
- Assist NQF in anticipating questions and identifying additional information that may be useful to the SC
- Keep SC on track to meet goals of the project without hindering critical discussion/input
- Represent the SC at CSAC meetings
- Participate as a SC member

Role of NQF Staff

NQF project staff works with SC to achieve the goals of the project and ensure adherence to the consensus development process:

- Organize and staff SC meetings and conference calls
- Guide SC through the CDP and advise on NQF policy and procedures; ensure NQF evaluation criteria is appropriately applied and process is followed
- Review measure submissions and prepare materials for Committee review
- Draft and edit reports for SC review
- Ensure and facilitate communication among all project participants (including SC and measure developers)
- Facilitate collaboration between different NQF projects

Role of NQF Staff
Communication

- Respond to NQF member or public queries about the project
- Maintain documentation of project activities
- Post project information to NQF's website
- Work with measure developers to provide necessary information and communication for the SC to fairly and adequately evaluate measures for endorsement
- Publish final project report

Role of Methods Panel

- Scientific Methods Panel was created to ensure higher-level and more consistent reviews of the scientific acceptability of measures
- The Methods Panel is charged with:
  - *Conducting evaluation of complex measures for the Scientific Acceptability criterion, with a focus on reliability and validity analyses and results*
  - *Serve in advisory capacity to NQF on methodologic issues, including those related to measure testing, risk adjustment, and measurement approaches.*
- The Methods Panel review will help inform the Standing Committee's endorsement decision. The Panel will not render endorsement recommendations.

Role of the Technical Expert Panel (TEP)

For the Cost and Efficiency Project, TEPs are convened to provide clinical input on clinically focused cost measures.

- Charged with review and providing feedback on clinical logic, episode trigger and end definitions, clinical exclusions, clinical risk factors for risk adjustment
  - *Focus on validity of measure specifications*
- Qualitative feedback only, no votes collected
- Time-limited convening
- Experts are pulled from existing standing committees, targeted outreach, and public nominations to fill other relevant gaps in expertise

Questions?

Overview of NQF's Cost and Efficiency Portfolio

Cost and Efficiency Portfolio of Measures

- This project will evaluate measures related to Cost and resource use that can be used for accountability and public reporting for all populations and in all settings of care.
- NQF currently has 10 endorsed measures within this topic area.

Cost and Efficiency Portfolio of NQF-endorsed Measures

- 2579 Hospital-level, risk-standardized payment associated with a 30-day episode of care for pneumonia (PN)
- 3474 Hospital-level, risk-standardized payment associated with a 90-day episode of care for elective primary total hip and/or total knee arthroplasty (THA/TKA)
- 1598 Total Resource Use Population-Based PMPM Index
- 1604 Total Cost of Care Population-Based PMPM Index
- 2158 Medicare Spending Per Beneficiary (MSPB) - Hospital
- 2431 Hospital-level, risk-standardized payment associated with a 30-day episode-of-care for Acute Myocardial Infarction (AMI)
- 2436 Hospital-level, risk-standardized payment associated with a 30-day episode-of-care for heart failure (HF)

Most Recently Endorsed

- 3509 Routine Cataract Removal with Intraocular Lens (IOL) Implantation
- 3510 Screening/Surveillance Colonoscopy
- 3512 Knee Arthroplasty

Summary of Portfolio

- All rely on administrative claims data
- Typically count costs based on what is paid by the health plan to the provider/hospital
- Only two endorsed measures consider out-of-pocket costs from patients
- All are risk-adjusted (e.g., HCCs, ACGs)
- All focus on adult population; majority for Medicare population
- Varying levels of analysis (e.g., hospital, clinician, clinician groups)
- Do not capture or represent activity-based costs, production costs (fixed or variable), administrative costs, government funding to support healthcare delivery, or societal costs (e.g., lost wages, sick days)

Cost, Resource Use, Efficiency, and Value

Measures Included in the Scope of Cost and Efficiency Committee Evaluations

- Cost Measures
  - the actual price paid by health plans for health plan member
  - may also include a member (consumer) cost based on member co-pays, coinsurance, and deductibles
- Resource use measures
  - broadly applicable and comparable measures of health services counts (in terms of units or dollars) that are applied to a population or event (broadly defined to include diagnoses, procedures, or encounters).
  - counts the frequency of defined health system resources; some may further apply a dollar amount (e.g., allowable charges, paid amounts, or standardized prices) to each unit of resource use.
- Does not include measures of appropriateness, ED throughput

Getting to Efficiency and Value

Current State of NQF evaluation and endorsement of cost and efficiency measures

- Evaluation of cost and resource use measures by Cost and Efficiency Standing Committee
- Quality measures are submitted and reviewed separately, by the relevant (clinical) standing committees
- Developers are able to submit "paired" measures, but to date, no measure pairs for cost and quality have been submitted (i.e., no efficiency measures)
- No evaluation process or criteria to assess how cost/resource use and quality measures should be used together; left to users to determine as part of implementation

Challenges with Cost and Resource Use Measure Evaluation

- Discerning whether there is an opportunity to improve costs/resource use by implementing the measure
- Determining whether social factors should be included in the risk-adjustment model; what impact does it have on performance and rankings?
- Validity testing methodology
- Threshold for reliability estimates
- Attribution approach
- Impact of the inclusion/exclusion of pharmacy costs

Cost and Resource Use Measure Evaluation Pipeline

- CMS had a statutory requirement to develop cost measures that cover 50% of Medicare costs (primarily through MIPS)
- Mix of measurement approaches
  - *Episode-based, clinician level*
  - *Cost per capita/spending per beneficiary*
- Maintenance reviews
- Ongoing assessment of NQF capacity and process
  - *Clinical TEP reviews*
  - *Timelines*
  - *Staff and Committee capacity*

Cost Measure Pipeline—Tentative

- Spring 2020 (6 measures)
  - *Total Per Capita Cost*
  - *Medicare Spending per Beneficiary (MSPB) (clinician level)*
  - *(4) Medicare Spending per Beneficiary (PAC setting)*
- Fall 2020 (Maintenance Review + new measures)
  - ***1598** Total Resource Use Population-Based PMPM Index*
  - ***1604** Total Cost of Care Population-Based PMPM Index*
  - ***2158** Medicare Spending Per Beneficiary (hospital level)*
  - ***2431** Hospital-Level, Risk-Standardized Payment Associated with a 30-Day Episode of Care for Acute Myocardial Infarction (AMI)*
  - ***2436** Hospital-Level, Risk-Standardized Payment Associated with a 30-Day Episode of Care for Heart Failure*
  - ***2579** Hospital-Level, Risk-Standardized Payment Associated with a 30-Day Episode of Care for Pneumonia*

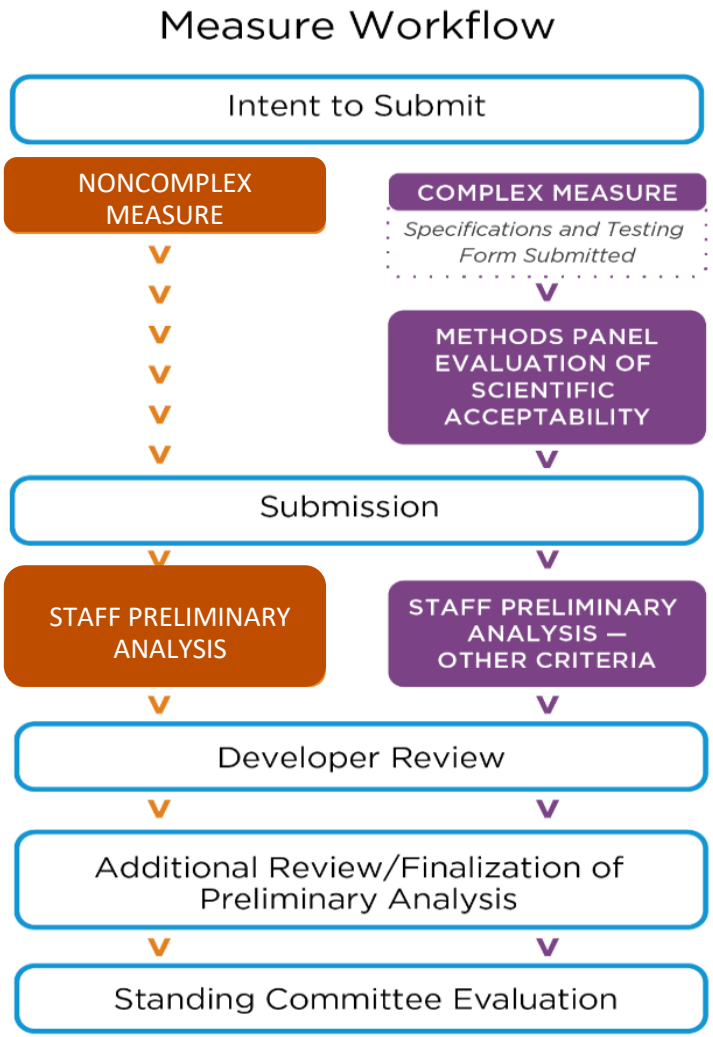
Activities and Timeline

- Fall 2019 cycle
  - There were no measures submitted for evaluation during this cycle
  - Strategic Web Meeting: Thursday, March 19, 1:30 pm - 3:30 pm ET
- Spring 2020 cycle
  - Scientific Methods Panel review: January-April 2020
  - Standing Committee review: April-June 2020
    - Dates TBD
    - Anticipate 1-day in-person meeting for measure review (~June 2020)

Questions?

Overview of Measure Evaluation Process

Measure Evaluation Workflow



NQF Consensus Development Process (CDP) Measure Evaluation

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**

Noncomplex Measures

- **Process measures**
- **Structural measures**
- **Previously endorsed complex measures with no changes/updates to the specifications or testing**

Measure Evaluation Inputs to the Standing Committee

- Multistakeholder Committee
- Evaluates all evaluation criteria
- Makes recommendation for endorsement

- Statistical & methodological expertise
- Clinical expertise
- Evaluates clinical elements of measure
- Multistakeholder comments

When Measures are Submitted to NQF

- NQF team reviews measures for the following:
  - *All required submission form items have a response*
  - *Submission meets the minimum requirements to be reviewed (e.g., testing is performed at requisite levels (data element and/or measure score)*
- Committee completes measure-specific disclosures of interest
- NQF staff creates a measure worksheet for each measure
  - *Includes: all submission materials (i.e., measure specifications, testing information, evidence information), staff analysis, and summary of Methods Panel review*

Complex Measure Evaluation

- Complex measures are reviewed by the SMP when:
  - *Newly submitted*
  - *Maintenance measures with updated testing*
  - *NQF staff requests (e.g., expert opinion needed to support review of testing, review of unfamiliar methodology)*
- All measures reviewed by the SMP can be discussed by the Standing Committee
  - *Standing Committee will evaluate and make recommendations for endorsement for:*
    - » Measures that pass SMP review
    - » Measures on which the SMP did not reach consensus
  - *Measures that did not pass the SMP can be pulled by a Standing Committee member for further discussion*

Committee Measure Evaluation Process

- Committee members are notified of Methods Panel evaluation results (if complex measures reviewed by SMP)
- Members have the opportunity to pull failed measures for discussion (and re-vote for eligible measures)

Committee Consideration of Measures that Do Not Pass the SMP

- Any measure pulled by a Standing Committee member will be discussed
  - *Request should be submitted with a brief rationale*
- Some measures may be eligible for vote by the Standing Committee
  - *Eligibility will be determined by NQF Staff and SMP co-chairs*
  - *Measures that failed the SMP due to the following will not be eligible for re-vote:*
    - » Inappropriate methodology or testing approach applied to demonstrate reliability or validity
    - » Incorrect calculations or formulas used for testing
    - » Description of testing approach, results, or data is insufficient for SMP to apply the criteria
    - » Appropriate levels of testing not provided or otherwise did not meet NQF's minimum evaluation requirements

Committee Consideration of Measures that Do Not Pass the SMP

- For measures eligible for vote by the Committee:
  - *The full Committee must vote on whether to uphold the SMP's vote on reliability and validity*
    - » Vote to Uphold→ No further discussion of the measure
    - » CNR or vote to overturn SMP vote→ SC discusses and votes on reliability and/or validity
- Maintenance Measures
  - *Endorsement will be removed for maintenance measures not pulled for discussion.*

Committee Measure Evaluation Process

~3 week review period for Measure Worksheets:

- Maximum of ~8 measures per cycle
- Measure Information Form (MIF): describes measure and specifications (e.g., title, description, numerator, denominator)
- Preliminary analysis by NQF staff
- Committee preliminary ratings
- Member and public comments
- Information submitted by the developer
  - *Evidence and testing attachments*
  - *Spreadsheets*
  - *Additional documents*

Committee Measure Evaluation Process

- **Staff Preliminary analysis (PA)**: To assist the Committee evaluation of each measure against the criteria, NQF staff and the Methods Panel (if applicable) will prepare a PA of the measure submission and offer preliminary ratings for each criteria.
  - *The PA will be used as a starting point for the Committee discussion and evaluation*
  - *Methods Panel will complete review of the Scientific Acceptability criterion for complex measures*
- **Member individual evaluation**: Each Committee member will conduct an in-depth evaluation on all measures under review
  - *Committee members will be assigned a measure (or parts of criteria for a measure) for which they will serve as lead discussant in the evaluation meeting*

Committee Measure Evaluation Process

- NQF staff compiles votes and redistributes measure worksheet with summary of all members preliminary analyses
- **Lead discussants** are assigned to each measure for committee evaluation meetings
- **Measure evaluation and recommendations at the in-person/web meeting**: The entire Committee will discuss and rate each measure against the evaluation criteria and make recommendations for endorsement.

Evaluation Process Continues

- **Staff will prepare a draft report** detailing the Committee's discussion and recommendations
  - *This report will be released for a 30-day public and member comment period*
- **Post-comment call**: The Committee will re-convene for a post-comment call to discuss comments submitted
- **Final endorsement decision by the CSAC**
- **Opportunity for public to appeal endorsement decision** (for endorsed measures only)

Cost and Resource Use Measure Evaluation Criteria Overview

NQF Cost and Resource Use Measure Evaluation Criteria for Endorsement

NQF endorses measures for accountability applications (public reporting, payment programs, accreditation, etc.) as well as quality improvement.

- Standardized evaluation criteria
- Criteria have evolved over time in response to stakeholder feedback
- The performance measurement enterprise is constantly growing and evolving—greater experience, lessons learned, expanding demands for measures—the criteria evolve to reflect the ongoing needs of stakeholders

Major Endorsement Criteria *(page 32 in the SC Guidebook)*

- **Importance to measure and report**: Goal is to measure those aspects with greatest potential of driving improvements; if not important, the other criteria are less meaningful (**must-pass**)
- **Reliability and Validity-scientific acceptability of measure properties**: Goal is to make valid conclusions about quality; if not reliable and valid, there is risk of improper interpretation (**must-pass**)
- **Feasibility**: Goal is to, ideally, cause as little burden as possible; if not feasible, consider alternative approaches
- **Usability and Use** (**must-pass** for maintenance measures): Goal is to use for decisions related to accountability and improvement; if not useful, probably do not care if feasible
- **Comparison to related or competing measures**

Criterion 1: Importance to Measure and Report *(page 34-42)* – (Must Pass)

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.

1a. ***High Impact:*** *The measure focus addresses a demonstrated high-impact aspect of healthcare (e.g., affects large numbers, leading cause of morbidity/mortality, high resource use [current and/or future], severity of illness, and patient/societal consequences of poor quality).*

AND

1b. ***Opportunity for Improvement:*** *Demonstration of resource use or cost problems and opportunity for improvement, i.e., data demonstrating considerable variation in cost or resource use across providers.*

Criterion 2: Scientific Acceptability of the Measure Properties (pages 42-54) – (Must Pass)

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

- 2a. Reliability
  - *2a1. Precise specifications including exclusions*
  - *2a2. Reliability testing—data elements or measure score*
- 2b. Validity
  - *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*

Criterion 3: Feasibility
(pages 54-55)

Extent to which the required data are readily available, or could be captured without undue burden, and can be implemented for performance measurement.

3a: Clinical data routinely generated and used during care delivery

3b: Electronic sources

3c: Data collection strategy can be implemented

Criterion 4: Usability and Use *(pages 55-56)*

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.

Use (4a): Must-pass for maintenance measures

4a1: Accountability and Transparency*: Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement.*

4a2: Feedback by those being measured or others*: Those being measured have been given results and assistance in interpreting results; those being measured, and others have been given opportunity for feedback; the feedback has been considered and incorporated into the measure.*

Usability (4b)

4b1: Improvement*: Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.*

4b2: Benefits outweigh the harms*: The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).*

4b3. *Data and result detail are maintained such that the resource use measure, including the clinical and construction logic for a defined unit of measurement, can be deconstructed to facilitate transparency and understanding*

Criterion 5: Related or Competing Measures *(pages 57-58)*

If a measure meets the four criteria and there are endorsed/new **related** measures (same measure focus or same target population) or **competing** measures (both the same measure focus and same target population), the measures are compared to address harmonization and/or selection of the best measure.

- 5a. The measure specifications are harmonized with related measures **OR** the differences in specifications are justified.
- 5b. The measure is superior to competing measures (e.g., is a more valid or efficient way to measure) **OR** multiple measures are justified.

Questions?

Social Risk Trial Overview

Background

- NQF conducted a two-year trial period from 2015-2017. During this time, adjustment of measures for social risk factors was no longer prohibited
- The NQF Board of Directors reviewed the results of the trial period and determined there was a need to launch a new social risk initiative
- As part of the Equity Program, NQF will continue to explore the need to adjust for social risk
- Each measure must be assessed individually to determine if SDS adjustment is appropriate (included as part of **validity** subcriterion)
- The Standing Committee will continue to evaluate the measure as a whole, including the appropriateness of the risk-adjustment approach used by the measure developer
- Efforts to implement SDS adjustment may be constrained by data limitations and data collection burden

Standing Committee Evaluation

The Standing Committee will be asked to consider the following questions:

- Is there a conceptual relationship between the SDS factor and the measure focus?
- What are the patient-level sociodemographic variables that were available and analyzed during measure development?
- Does empirical analysis (as provided by the measure developer) show that the SDS factor has a significant and unique effect on the outcome in question?
- Does the reliability and validity testing match the final measure specifications?

Questions?

SharePoint Overview

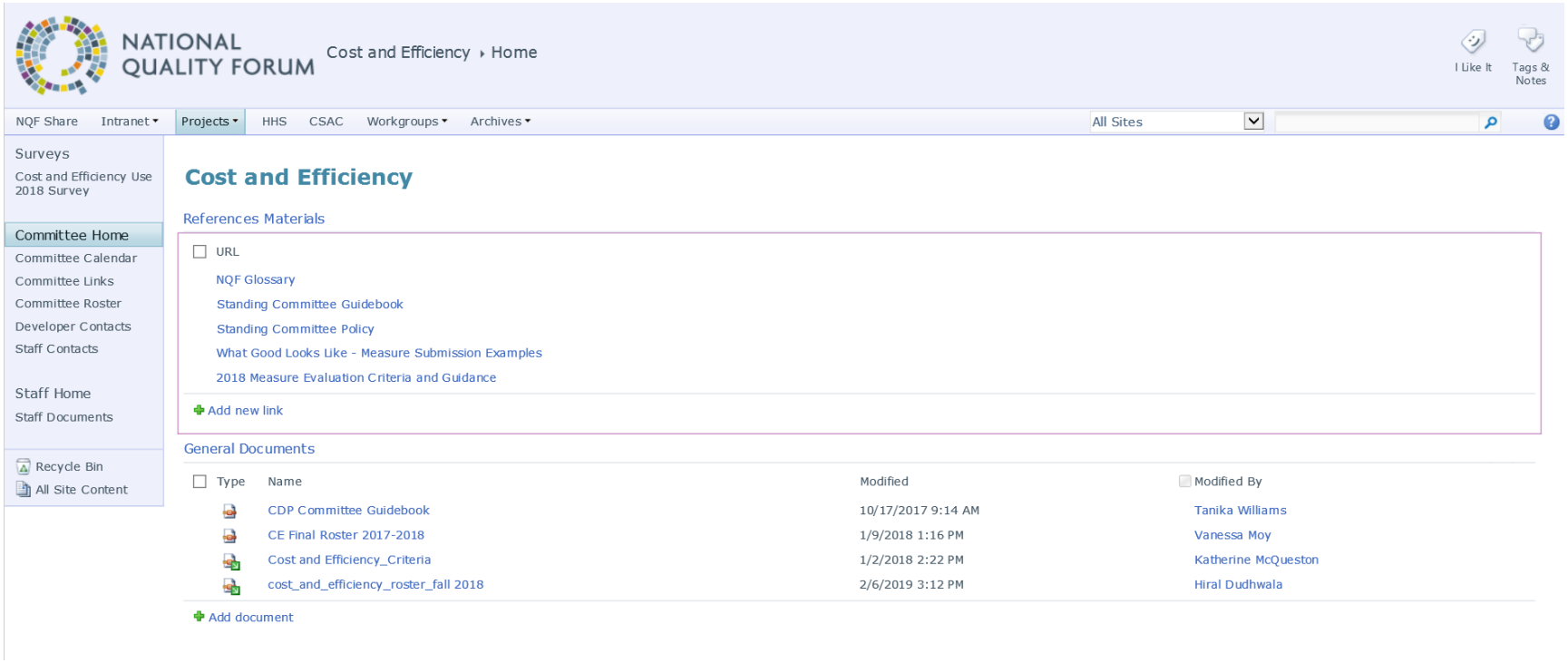
SharePoint Overview

http://share.qualityforum.org/Projects/costEff/SitePages/Home.aspx

- Accessing SharePoint
- Standing Committee Policy
- Standing Committee Guidebook
- Measure Document Sets
- Meeting and Call Documents
- Committee Roster and Biographies
- Calendar of Meetings

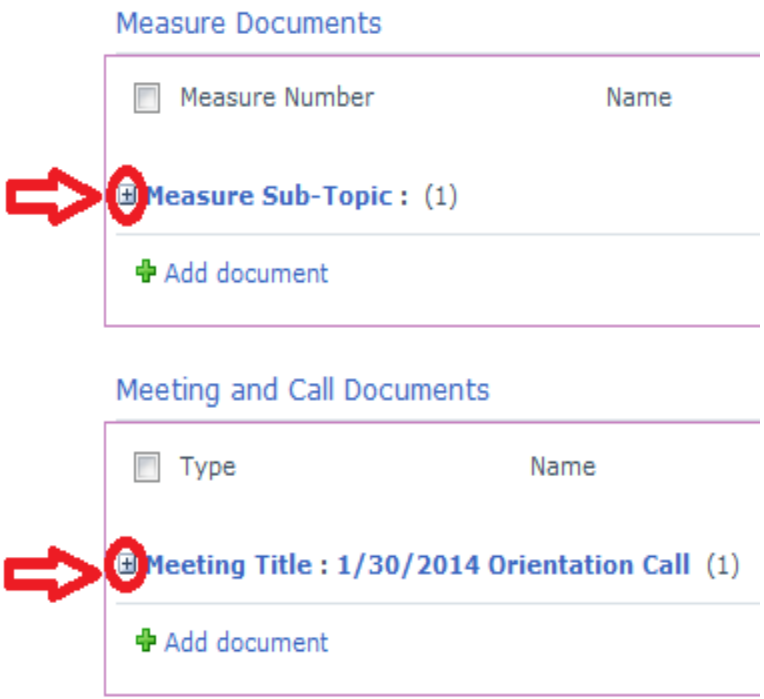
SharePoint Overview

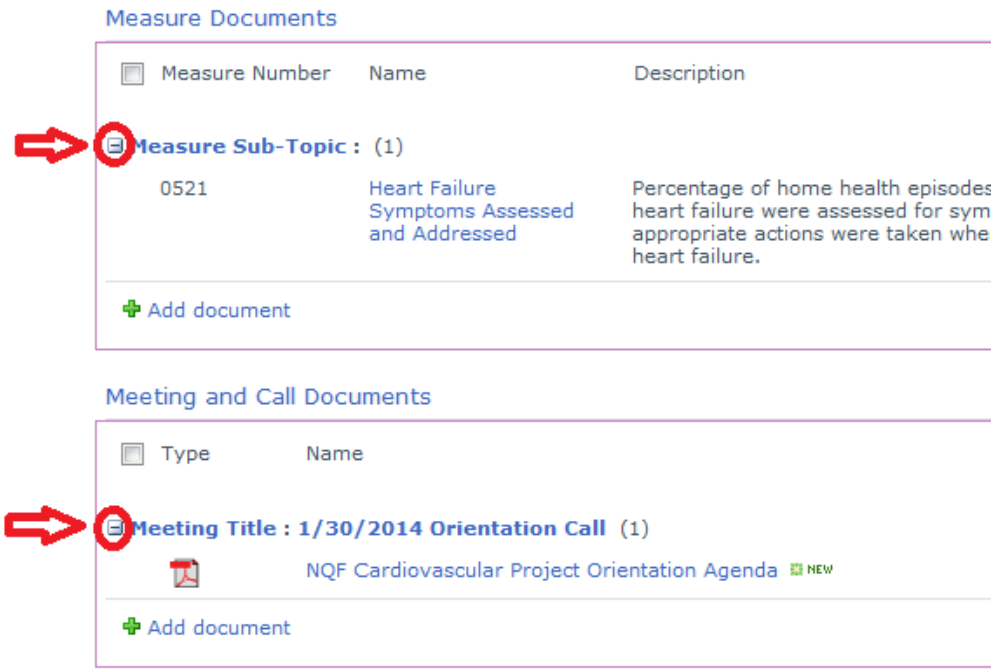
- Screen shot of homepage:



SharePoint Overview

- Please keep in mind:
- + and – signs :





Next Steps

Next Steps

- Strategic Web Meeting
  - *Cost measure evaluation during the social risk trial*
  - *Thursday, March 19 at 1:30-3:30 pm*
- NQF staff will distribute prep materials prior to the webinar for your review

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

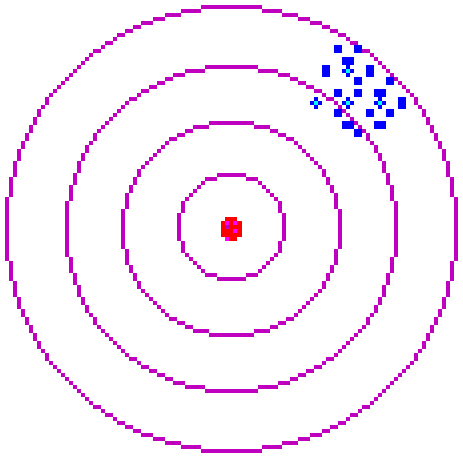
Questions?

Thank you.

Appendix

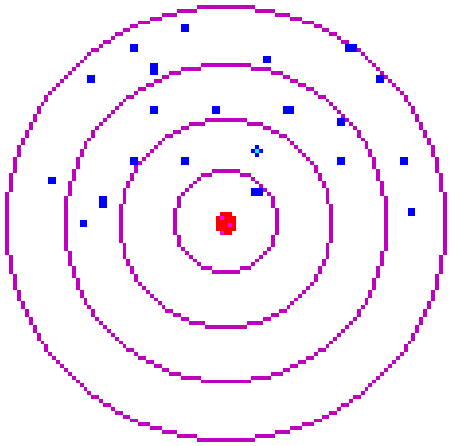
Reliability and Validity *(page 46)*

Assume the center of the target is the true score.



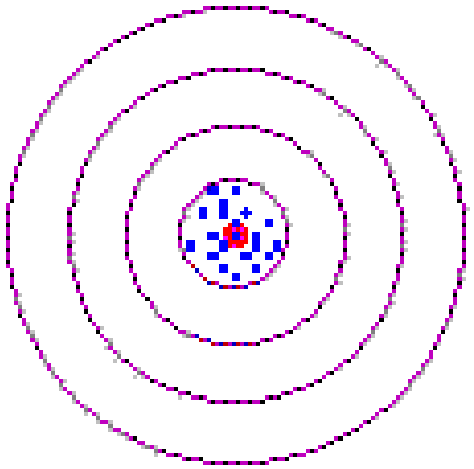
Reliable Not Valid

Consistent, but wrong



Neither Reliable Nor Valid

Inconsistent & wrong



Both Reliable And Valid

Consistent & correct

Evaluating Scientific Acceptability – Key Points *(page 45)*

Empirical analysis to demonstrate the reliability and validity of the measure as specified, including analysis of issues that pose threats to the validity of conclusions about quality of care such as exclusions, risk adjustment/stratification for outcome and resource use measures, methods to identify differences in performance, and comparability of data sources/methods.

Reliability Testing – Key points *(page 48)*

- 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).
  - *Example – Statistical analysis of sources of variation in performance measure scores (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*
- Consider whether testing used an appropriate method and included adequate representation of providers and patients and whether results are within acceptable norms
- Algorithm 2

Rating Reliability: Algorithm 2 *(page 47)*

[Screen share Reliability algorithm]

Validity Testing *(pages 48-54)*

- 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 to a "gold standard"
- 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.

Rating Validity: Algorithm 3 *(page 53)*

[Screen share Validity algorithm]

Threats to Validity

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

Criterion 2: Scientific Acceptability

New measures	Maintenance measures
• Measure specifications are precise with all information needed to implement the measure	NO DIFFERENCE: Require updated specifications
• Reliability • Validity (including risk adjustment)	**DECREASED EMPHASIS**: If prior testing adequate, no need for additional testing at maintenance with certain exceptions (e.g., change in data source, level of analysis, or setting) Must address the questions regarding use of social risk factors in risk-adjustment approach

Criteria 3 and 4: Feasibility and Usability and Use

Feasibility

New measures	Maintenance measures
• Measure feasible, including eMeasure feasibility assessment	NO DIFFERENCE: Implementation issues may be more prominent

Usability and Use

New measures	Maintenance measures
• Use: used in accountability applications and public reporting	**INCREASED EMPHASIS**: Much greater focus on measure use and usefulness, including both impact and unintended consequences
• Usability: impact and unintended consequences	

Updated Guidance for Measures that Use ICD-10 Coding

- For CY2019 and beyond, reliability testing should be based on ICD-10 coded data.
- Validity testing should be based on ICD-10 coded data
- If providing face validity (FV), both FV of the ICD-10 coding scheme and FV of the measure score as an indicator of quality is required update