Infectious Disease Steering Committee

Conference Call: Orientation

NATIONAL QUALITY FORUM

July 10, 2012

Welcome and Introductions

NQF Project Staff

- Reva Winkler, MD, MPH
  - Senior Director, Performance Measures
- Alexis Morgan, MPH
  - Senior Project Manager, Performance Measures
- Adeela Khan, MPH
  - Project Analyst, Performance Measures

Agenda for the Call

- Overview of NQF
- Consensus Development Process
- Overview of the Infectious Disease project
- Introduction to Measure Evaluation Process
- Logistics
- Questions at any time

Steering Committee Handbook

- New document
- Provides important background and contextual information for Steering Committees
  - What is NQF?
  - How does NQF endorse consensus standards?
  - Glossary
- On SharePoint

Overview of NQF

What is NQF ?

A private, non profit voluntary consensus standards-setting organization.

- Public-private partnership
- Multi-stakeholder Board of Directors
- Membership: 400+ Member organizations – organized into 8 Stakeholder Councils

Consumers	**Health Plans**
Purchasers	**Community/public health**
Professionals	**Quality measurement, research, improvement**
Providers	
Supplier/industry	

NQF Mission

The National Quality Forum (NQF) operates under a three-part mission to improve the quality of American healthcare by:

- Building consensus on national priorities and goals for performance improvement and working in partnership to achieve them.
- Endorsing national consensus standards for measuring and publicly reporting on performance.
- Promoting the attainment of national goals through education and outreach programs.

NQF's Roles

- **Standard setting organization**
  - Voluntary consensus standards:
    - *Performance measures*
    - *Serious reportable events*
    - *Preferred practices*
    - *Frameworks*
- **Neutral convener**
  - National Priorities Partnership (NPP)
  - Measure Applications Partnership (MAP)

Questions?

NATIONAL QUALITY FORUM

Consensus Development Process

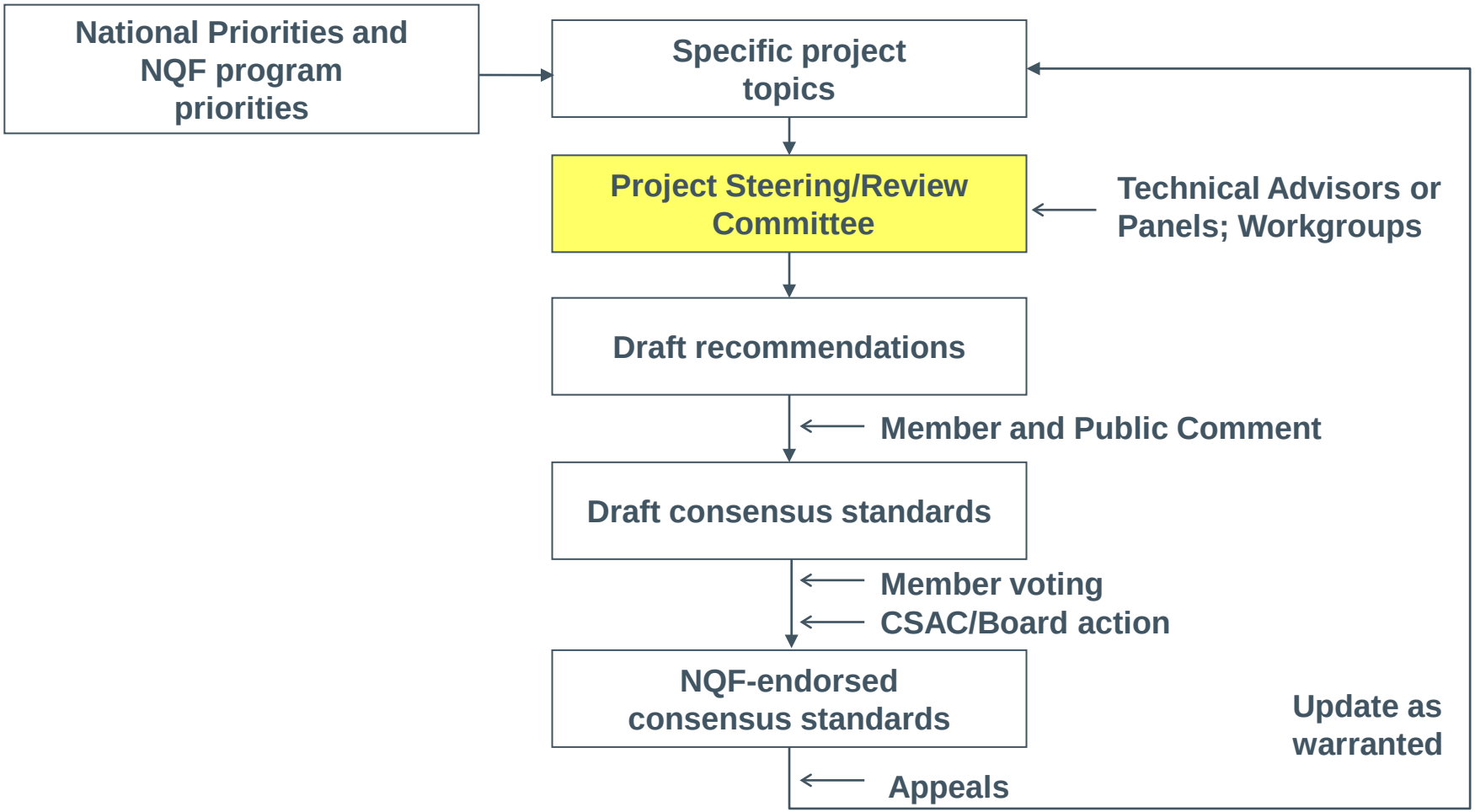
Why NQF Endorsement?

- Standardized performance measures are tools to assess quality that can be used to compare.
- An NQF endorsement reflects rigorous scientific and evidence-based review, input from patients and their families, and the perspectives of people throughout the healthcare industry.

Consensus Development Process

1. Call for Nominations
2. Call for Candidate Standards
3. Candidate Consensus Standard Review – Recommendations and Draft Report
4. Public and Member Comment
5. Member Voting
6. Consensus Standards Approval Committee (CSAC) Decision
7. Board Ratification
8. Appeals

Endorsement Process



Endorsement Maintenance Process

- Purpose: To ensure the currency and relevance of NQF-endorsed consensus standards through a regularized schedule of reviewing measures for continued endorsement
- Timeline: Review of endorsed measures every 3 years
- Process:
  - Implementation comments are sought and considered
  - Measures are reviewed against all the evaluation criteria
  - New and endorsed measures are reviewed within same project
    - Harmonize measure specifications
    - Endorse “best in class” measures

Overview of the Infectious Disease Project

Infectious Disease Project Overview

This project will evaluate and endorse measures for accountability/public reporting and quality improvement for:

- HIV/AIDS - 13 measures
- Hepatitis – 10 measures
- Upper respiratory infections - 2 measures
- Sepsis – 1 measure
- Central line infections – 1 measure
- Ventilator associated pneumonia -1 measure

Infectious Disease Endorsement Maintenance

- 29 measures for maintenance review
  - 6 have been retired by developer
- 5 new measures submitted
- 28 measures for review in this project
- 22 additional Infectious Disease measure not due for maintenance
  - Table of Infectious Disease measures

Current Use of Measures

Measures being reviewed in this project are used for accountability/public reporting by:

- Medicare Physician Quality Reporting System (PQRS)
- NCQA HEDIS measures
- States (Maine, California, Minnesota)

…… and many others for quality improvement.

Roles of Steering Committee

- Evaluate submitted measures against the NQF measure evaluation criteria and make recommendations for endorsement
- Act as a proxy for the NQF multi-stakeholder membership for a specific project
- Work with NQF staff to achieve the goals of the project
- Respond to comments submitted during the review period
- Co-chairs represent the Steering Committee at CSAC meeting

Roles of Steering Committee (continued)

Individual members:

- Assigned to workgroups for in-depth review of selected measures
- Evaluate measures against each criterion
  - Indicate the extent to which each criterion is met and rationale for the rating
- Vote on a recommendation for all measures

Expectations of Steering Committee Members

- Attend meetings and conference calls
- Identify and acknowledge potential biases (real or perceived)
- Individually evaluate all measures using NQF evaluation criteria and guidance and submit evaluations in tools provided
  - If large number of measures, will be assigned a subset of measures for in-depth review and evaluation
- Participate in discussion and vote on ratings and recommendations for all measures
  - Lead discussion of some measure reviews as requested
- Review meeting summaries and draft reports
- Review public comments and suggest responses

Roles of the Steering Committee Co-Chairs

- Facilitate Steering Committee (SC) meetings
- Represent the SC at the CSAC meetings
- Keep SC on track to meet goals of the project without hindering critical discussion/input
- Assist NQF in anticipating questions and identifying additional information that may be useful to the SC
- Work with NQF staff to achieve the goals of the project
- 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
- Ensure communication among all project participants (including SC and measure developers)
- Facilitate necessary communication and collaboration between different NQF projects
- Respond to NQF member or public queries about the project
- Maintain documentation of project activities
- Post project information to NQF website

Project Activities and Timeline

(* Dates are tentative)

Meeting	Date
Steering Committee Orientation	July 10
Steering Committee Tutorial Call	July 24
Workgroup Calls	August 15, 16, 22, 23, 2012
Steering Committee In-person Meeting	August 28-29, 2012
Public & Member Comment (30 days)	October 2012*
Member Voting (15 days)	December 2012*
CSAC Review	January 2013*
Board Endorsement	January 2013*
Appeals (30 days)	February 2013*

Questions?

NATIONAL QUALITY FORUM

Measure Evaluation Overview

Why Measure?

- Measures drive improvement
- Measures inform consumers and other stakeholders
- Measures influence payment

Characteristics of Measures

- Measures are different from concepts or ideas
  - Quality of care is an abstract construct
  - A quality measure is a numeric quantification of healthcare quality
- Measures have detailed specifications
  - What to count (including codes, definitions)
  - Who is included and/or excluded
  - When to count
  - Where to find data
  - How to compute score

Types of Performance Measures

- Quality
  - Structure
  - Process
  - Intermediate clinical outcome
  - Outcome
    - Use of services (used as proxy for outcome, cost)
  - Patient experience
- Resource use/cost
- Efficiency (combination of quality and resource use)
- Composite (combination of two or more individual measures in a single measure that results in a single score)

Conditions for Consideration

Reviewed by NQF Staff

- Measure Steward Agreement
  - All non-government organizations
- Entity and process to maintain and update the measure as needed/at least every 3 years
- Intended use of the measure includes accountability/public reporting as well as quality improvement
- Measure is fully specified and tested for reliability and validity
- Attests that harmonization and competing measures are considered and addressed
- Measure submission is complete

Evaluation Criteria

- Subcriteria delineate how to demonstrate that the major criteria are met
  - How do you know a measure is important, scientifically acceptable, etc.?
- Criteria parallel best practices for measure development
  - For example, begin with identifying what is important to measure, and later what is feasible
- Most criteria/subcriteria involve a matter of degree rather than all-or-nothing determination
  - Requires both evidence and expert judgment

Rating Scale

Rating	Definition
High	Based on the information submitted, there is high confidence (or certainty) that the criterion is met
Moderate	Based on the information submitted, there is moderate confidence (or certainty) that the criterion is met
Low	Based on the information submitted, there is low confidence (or certainty) that the criterion is met
Insufficient	There is insufficient information submitted to evaluate whether the criterion is met (e.g., blank, incomplete, or not relevant, responsive, or specific to the particular question)

New versus Endorsed Measures

- All measures—both new and endorsed—are expected to meet current criteria and guidance
- Endorsed measures
  - Data from implementation of the measure as specified for 1b (Opportunity for Improvement)
    - Potential for reserve status
  - Reliability and validity testing expanded unless meet high rating
  - Usability: Actual use in public reporting/other accountability and improvement OR specific plans and timeline
  - Feasibility: Problems with implementation or unintended consequences

Criterion # 1: Impact, Opportunity, Evidence– Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in health care quality and improving health outcomes for a specific high-impact aspect of healthcare where there is variation in or overall less-than-optimal performance.

- Must pass criterion
- Must pass all three subcriteria
  - 1a. High impact
  - 1b. Performance gap/opportunity for improvement**
    - Including disparities
  - 1c. Evidence supports measure focus

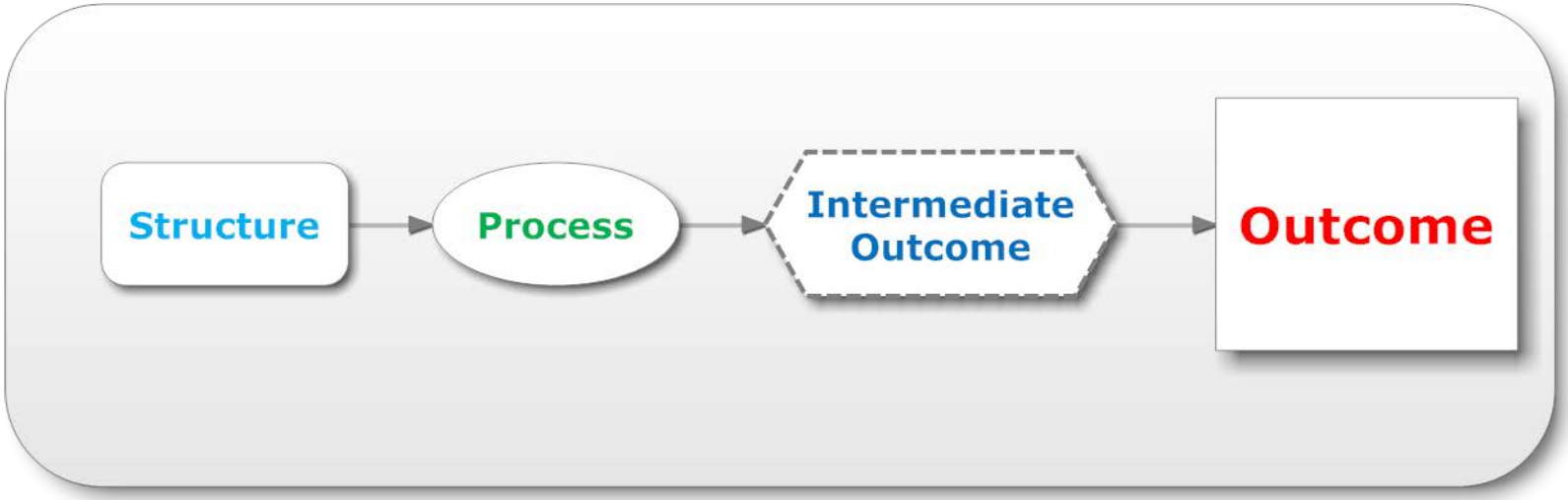
** Measures being reviewed for endorsement maintenance may qualify for reserve status if they address an important aspect of quality but fail to demonstrate a gap in performance and certain other criteria are met. Such measures should be rated on all evaluation criteria

1. Importance to Measure and Report

- 1a. High impact
  - National health goal or priority, i.e., related to the National Quality Strategy
  - Data on numbers of persons affected, high resource use, severity of illness, consequences of poor quality

- 1b. Opportunity for improvement
  - Data demonstrating considerable variation in performance **OR** overall less than optimal performance
  - Data on disparities in care

Subcriterion 1c: Evidence for Measure Focus



- Hierarchical preference for
  - Outcomes linked to evidence-based processes/structures
  - Outcomes of substantial importance with plausible process/structure relationships
  - Intermediate outcomes
  - Processes/structures

Most closely linked to outcomes

Subcriterion 1c: Evidence Guidance for Evaluation

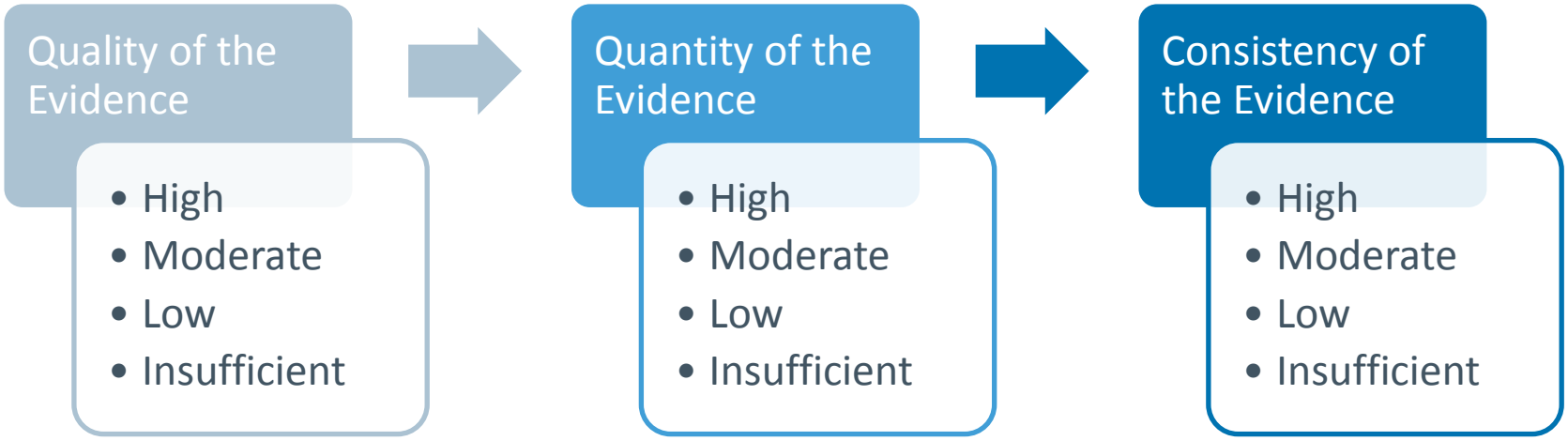
If measure focus IS a health outcome

- Rating of quantity, quality, consistency of the body of evidence is not required
- A rationale supports the relationship of the health outcome to processes or structures of care

If measure focus IS NOT a health outcome

- Explicit, transparent information on the **quantity**, **quality**, **consistency** of the body of evidence
  - Not selected individual studies

Specific Rating Scale
1c-Evidence



Quantity of Body of Evidence

Rating	**Quantity of Body of Evidence:** Total number of studies (not articles or papers)
High	5+ studies
Moderate	2-4 studies
Low	1 study
Insufficient to evaluate	• No empirical evidence OR • Only selected studies from a larger body of evidence

Quality of Body of Evidence

Rating	**Quality of Body of Evidence:** Certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence
High	RCTs; direct evidence for specific measure focus; adequate size to obtain precise estimates of effect; without serious flaws that introduce bias
Moderate	Non-RCTs w/control for confounders; large, precise estimates of effect **OR** RCTs without serious flaws, but either indirect evidence or imprecise estimate of effect
Low	RCTs w/flaws introduce bias **OR** Non-RCTs w/small or imprecise estimate of effect or without control of confounders
Insufficient to evaluate	• No empirical evidence **OR** • Only selected studies from a larger body of evidence

Consistency of Results of Body of Evidence

Rating	**Consistency of Results of Body of Evidence**: Stability in both the direction and magnitude of clinically/practically meaningful benefits and harms to patients (benefit over harms) across studies in the body of evidence
High	Estimates of clinically/practically meaningful benefits & harms to patients consistent in direction & similar in magnitude across preponderance of studies
Moderate	Estimates of benefits & harms consistent in direction but may differ in magnitude (If 1 study then estimate of benefits greatly outweigh harms)
Low	Estimates of benefits & harms differ in both direction and magnitude **OR** wide confidence intervals prevent estimating net benefit (If 1 study then estimate of benefits do not greatly outweigh harms)
Insufficient to evaluate	No assessment of magnitude and direction of benefits and harms to patients

Subcriterion 1c: Evidence Decision Logic

Quantity	Quality	Consistency	Does the measure pass subcriterion 1c?
Moderate or High	Moderate or High	Moderate or High	YES
Low	Moderate or High	Moderate	YES, IF additional research unlikely to change conclusion that benefits to patients outweigh harms. Otherwise NO.
Moderate or High	Low	Moderate or High	YES, IF potential benefits to patients clearly outweigh potential harms. Otherwise NO.
Low, Moderate, or High	Low, Moderate, or High	Low	NO

Note: Insufficient evidence – does not pass 1c

Distinguishing Between a Low Rating Versus a Rating of Insufficient Evidence

- A low rating generally means the evidence/information demonstrates that a criterion is not met
  - For evidence – depends on combination of quantity, quality, consistency
- Insufficient evidence means either:
  - The evidence does exist and was presented but is not adequate for a definitive answer OR
  - The submission was incomplete or deficient in presenting evidence/information that does exist
- Ratings of Low or Insufficient Evidence for a subcriterion result in not passing a criterion but signify different reasons
  - For evidence – depends on combination of quantity, quality, consistency

Submitted vs. Existing Evidence (1c)

- Individual committee member preliminary evaluation
  - Rate the measures based on evidence submitted
  - Note if aware of additional evidence

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 care when implemented

2a. **Reliability (must-pass)**
- 2a1. Precise specifications including exclusions (previously 2d)
- 2a2. Reliability testing—data elements or measure score

2b. **Validity (must-pass)**
- 2b1. Specifications consistent with evidence
- 2b2. Validity testing—data elements or measure score
- 2b3. Justification of exclusions—relates to evidence
- 2b4. Risk adjustment
- 2b5. Identification of differences in performance
- 2b6. Comparability of data sources/methods

2c. Stratification for disparities – *disparities now just addressed in 1b*

Reliability and Validity

- Reliability and validity are not all-or-none properties: they are a matter of degree
- Reliability and validity are not static: they can vary with different conditions of using the measure
- In order to be valid, a measure must be reliable
  - BUT, reliability does not guarantee validity

Characteristics of Measures

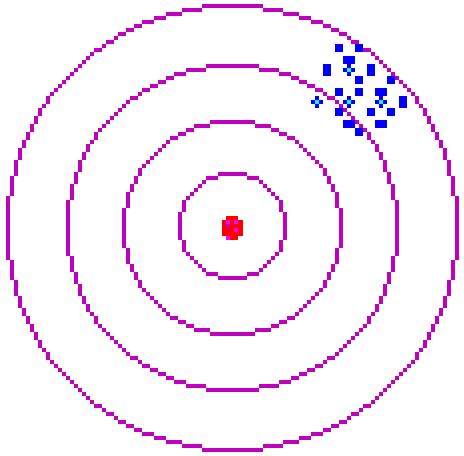
- Measures are not perfect
  - All measures result in a score:
    Obtained score = true score + error
  - Two basic types of error occur in measurement
    - Random error (chance disturbances) → Affects reliability
    - Nonrandom (systematic) error → Affects validity
- Measures can be evaluated in terms of **reliability** and **validity**

Reliability and Validity

- **Reliability**
  - Refers to the ***repeatability*** or ***precision*** of measurement
- **Validity**
  - Refers to the ***correctness*** of measurement

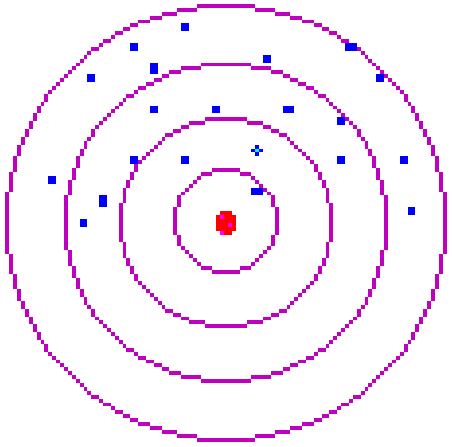
Reliability and Validity

Assume the center of the target is the true score…



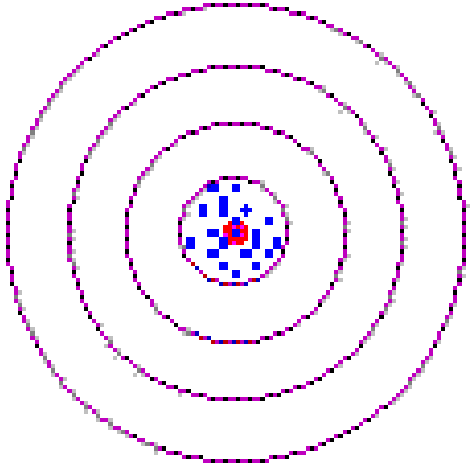
Reliable Not Valid

Consistent, but wrong



Neither Reliable Nor Valid

Inconsistent & wrong

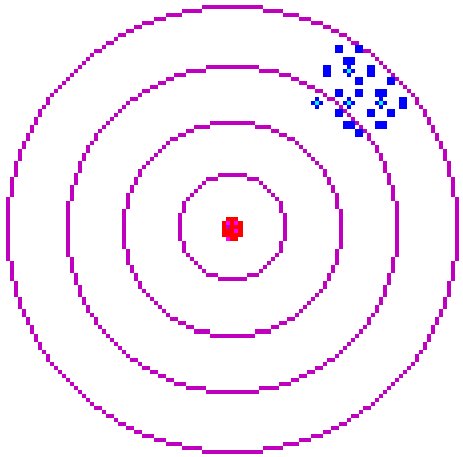


Both Reliable And Valid

Consistent & correct

Measure Is Reliable But Not Valid

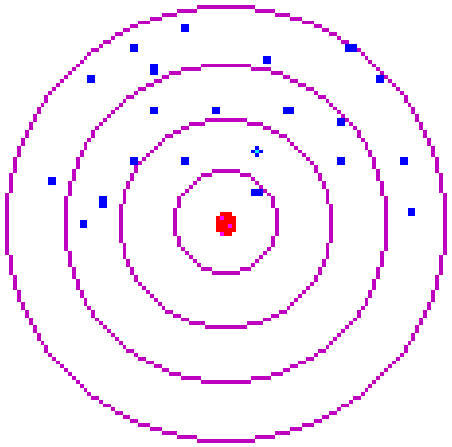
Example: Scale set at 5 lbs instead of 0 lbs



Consistent,
but
wrong

- Repeated weights are consistent, but when compared to measures taken from a calibrated scale, the results are consistently ~5 lbs heavier
- Systematic error (consistent direction/size) affects validity: these measures will not give an accurate weight

Measure Is Neither Reliable Nor Valid

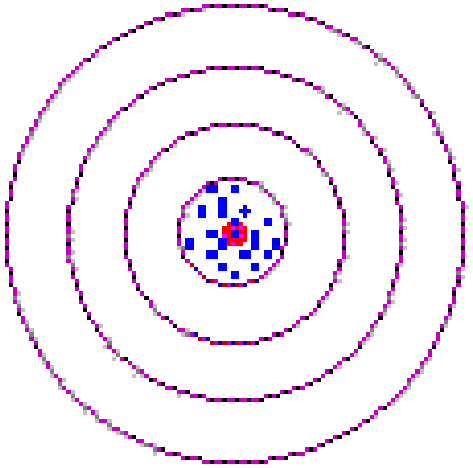


Inconsistent
and
wrong

- Repeated weights fluctuate widely; when compared to a calibrated scale, the results are always heavier

Measure Is Both Reliable And Valid

Example: Scale set at 0 lbs



Consistent and correct

- Repeated weights are consistent; when compared to measures taken from a calibrated scale, the results are close to the correct weight

Threats to Reliability

- Ambiguous measure specifications (e.g., instructions for data collection or scoring)
  - Can result in random error
- Small case volume or sample size, or rare events
  - Can affect the precision (reliability) of the measure score
- Other random errors
  - Random errors in coding or transcription (e.g., data coded for claims)
  - Random missing data

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)

Measure Testing

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.

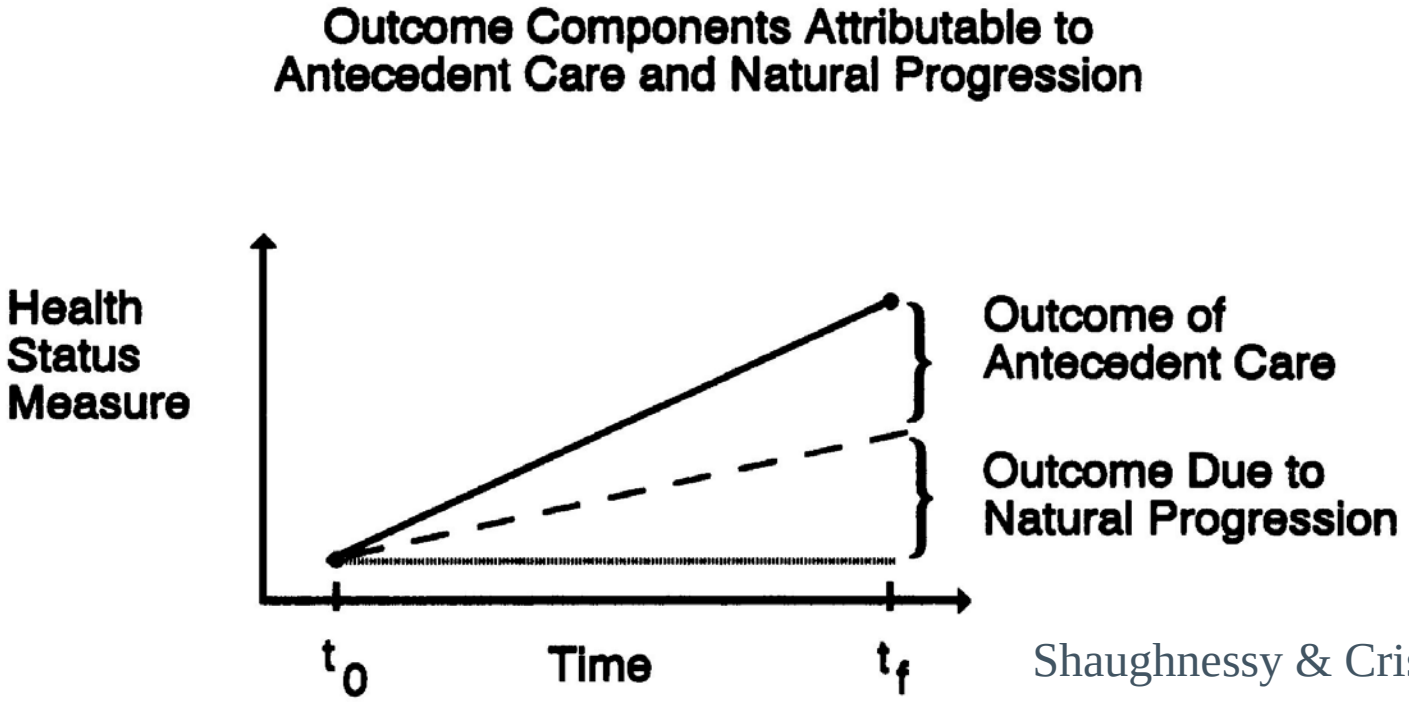
--Measure Testing Guidance Report

Evaluation of Testing

- Was the measure tested at the level of the data elements and/or the measure score?
  - High rating only if tested at both data element and measure score
  - Moderate – highest rating possible if only tested either data elements or measure score
  - Face validity acceptable only if systematically assessed
- Was an appropriate method used?
  - Consider level (data or score), data source, type of measure, topic, potential sources of error, conceptual relationships, feasibility
- Was the scope of testing adequate?
  - If sample, consider number of entities, number of patients, representativeness
- Were the results within acceptable norms?

Risk Adjustment

Case mix (or risk) adjustment: Process of controlling for patient factors that could influence patient outcomes or resource use; factors used in risk adjustment should be present before care begins.



Shaughnessy & Crisler, 1995

Criterion # 3: Usability*

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making.

- 3a. Meaningful, understandable, and useful for public reporting
  - Is it in use for public reporting or an accountability application and if not, what is plan/progress?
  - Is the rationale for use in accountability credible?

- 3b. Meaningful, understandable, and useful for quality improvement
  - Is it in use for improvement, and if not what is the plan/progress?
  - Is the rationale for use in QI credible?

** Currently being revised*

Criterion # 4: Feasibility

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

- 4a. Clinical data generated and used during care process
  - Blood pressure, lab value vs. survey or observation
- 4b. Electronic sources
  - EHR, claims vs. abstracted and entered into database/registry
  - Is there a credible, near-term path to electronic collection?
- 4c. Susceptibility to inaccuracies/unintended consequences identified
  - Ability to audit and detect?
- 4d. Data collection strategy can be implemented
  - Is it already in operational use or testing indicated ready for operational use?

5. Comparison to Related or Competing Measures

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.

Measure Evaluation Guidance

- Reports on guidance for measure evaluation:
  - Evidence for the Focus of Measurement and Importance to Measure and Report
  - Measure Testing and Scientific Acceptability of Measure Properties
  - Measure Harmonization
- Updated Measure Evaluation Criteria
- Specific rating scales for evidence (1c), reliability (2a), and validity (2b)
- Decision tables for Importance to Measure and Report and Scientific Acceptability of Measure Properties
- Revised Measure Submission Form
  - Most changes related to guidance on evidence (1c)
  - Some changes related to taxonomy (primarily response options, e.g., setting)
  - Some clarification in wording/instructions

Questions?

NATIONAL QUALITY FORUM

Logistics

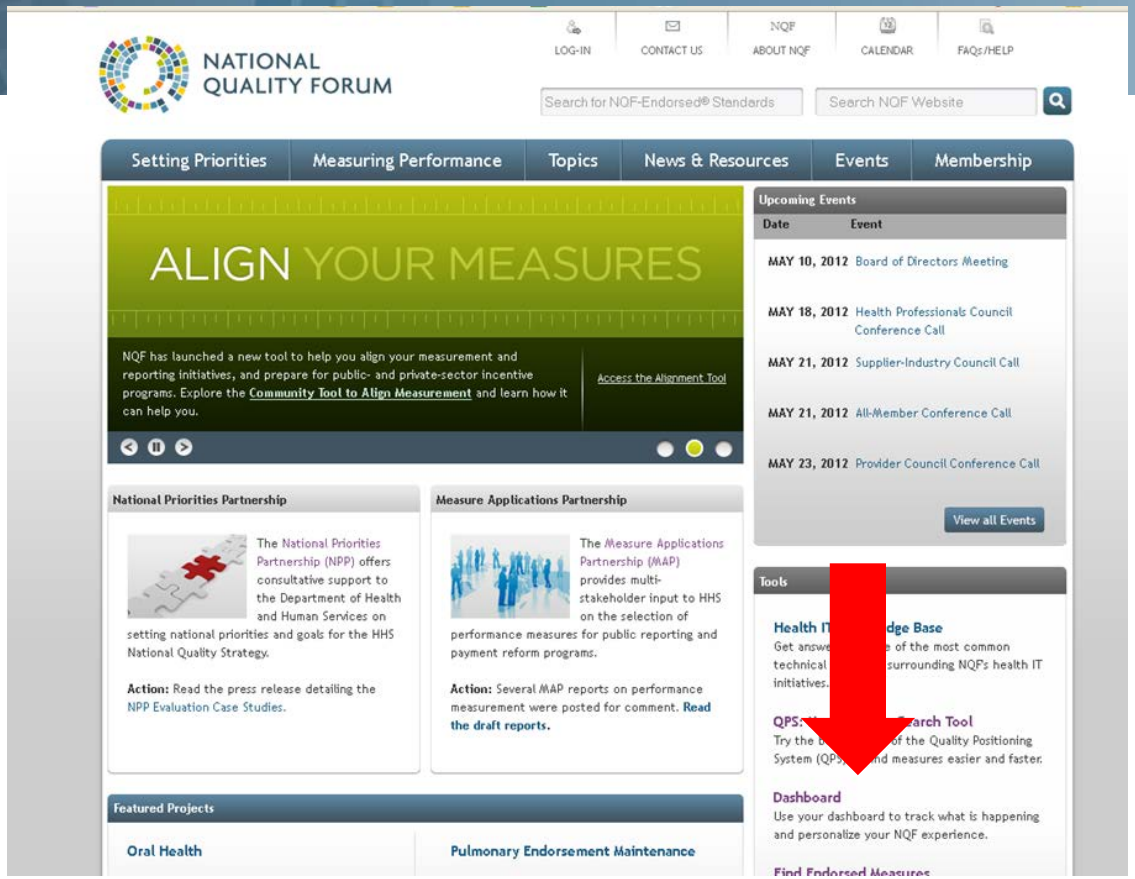
SharePoint

- NQF projects now use SharePoint to share documents with SC members
  - All materials will be posted to SP
  - Your NQF log-in provides access to this page

Next Steps

- Committee members will receive measures: July 18
- Tutorial call: July 24
- Workgroup calls: August 15, 16, 22, 23
- In person meeting: August 28-29

www.qualityforum.org



Enroll for an account and set up dashboard to follow this project at www.qualityforum.org – Click on DASHBOARD

Project Contact Info

- Reva: rwinkler@qualityforum.org
- Alexis: aforman@qualityforum.org
- Adeela: akhan@qualityforum.org
- NQF Phone: 202-783-1300

Questions?

NATIONAL QUALITY FORUM