# National Consensus Standards for Surgery

*Standing Committee Orientation/Q&A Webinar*

*Melinda Murphy*
*Andrew Lyzenga*
*Juliet Feldman*
*Alexandra Ogungbemi*

NATIONAL QUALITY FORUM

## Welcome & Introductions

## NQF Project Staff

- Melinda Murphy, RN, MS
  - Senior Director, Quality Measurement
- Andrew Lyzenga, MPP
  - Senior Project Manager, Quality Measurement
- Juliet Feldman
  - Project Manager, Quality Measurement
- Alexandra Ogungbemi
  - Project Analyst, Quality Measurement



## Surgery Standing Committee

- Anthony Asher, MD, FAANS, FACS
- Robert Cima, MD, MA
- Richard Dutton, MD, MBA
- Elisabeth Erekson, MD, MPH
- Lee Fleisher, MD
- Frederick Grover, MD
- William Gunnar, MD, MPH
- John Handy, MD
- Mark Jarrett, MD, MBA
- Clifford Ko, MD, MS, MSHS, FACS
- Barbara Levy, MD, FACOG, FACS
- Barry Markman, MD
- Kelsey McCarty, MS, MBA
- Lawrence Moss, MD
- Amber Slichta, MHSA
- Amy Moyer, MS, PMP
- Keith Olsen, PharmD, FCCP, FCCM
- Collette Pitzen, RN, BSN, CPHQ
- Lynn Reede, DNP, MBA, CRNA
- Gary Roth, DO, FACOS, FCCM, FACS
- Christopher Saigal, MD, MPH
- Robert Sawin, MD, MS
- Allan Siperstein, MD
- Larissa Temple, MD
- Melissa Thomason
- A.J. Yates, MD

## Agenda for the Call

- Role of the Standing Committee
- Project Scope
- Surgery Portfolio of Measures
- Process Updates
- Preliminary Analysis Walk-through
- SharePoint Tutorial



## Role of the Standing Committee
*General Duties*

- Act as a proxy for the NQF multi-stakeholder membership
- Serve 2-year or 3-year terms
- 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

## Role of the Standing Committee
### *Measure Evaluation Duties*

- All Members review ALL measures
- Lead discussants will be assigned to each measure
  - Responsible for thorough review of measures before March in-person meeting and presenting during meeting
- 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 Surgery portfolio of measures



## Surgery Portfolio of Measures

- This project will address measures in the areas of general and specialty surgery, including:
  - pre and post-surgical care
  - adverse surgical outcomes
  - other related topics
- The Surgery Standing Committee is responsible for overseeing NQF's portfolio of surgery-related measures
  - Currently, there are over 100 NQF-endorsed measures related to surgery, **69** of which are assigned to the Surgery Standing Committee for maintenance purposes

## Surgery Portfolio (cont.)

- Topic areas:
  - Abdominal and Colorectal Surgery (**5**)
  - Adverse Outcomes (**3**)
  - Antibiotic Prophylaxis (**8**)
  - Cardiac Surgery (**24**)
  - GU and GYN (**3**)
  - Orthopedic Surgery (**2**)
  - Pediatric Surgery (**7**)
  - Perioperative Care (**3**)
  - Thoracic Surgery (**4**)
  - Vascular Surgery (**9**)
  - VTE Prophylaxis (**1**)
- **New measures endorsed in last cycle (5):**
  - **2052**: Reduction of Complications through the use of Cystoscopy during Surgery for Stress Urinary Incontinence
  - **2063**: Performing cystoscopy at the time of hysterectomy for pelvic organ prolapse to detect lower urinary tract injury
  - **2558**: Hospital 30-Day All-Cause Risk-Standardized Mortality Rate Following CABG
  - **2561**: STS Aortic Valve Replacement (AVR) Composite Score
  - **2563**: STS Aortic Valve Replacement (AVR) + Coronary Artery Bypass Graft (CABG) Composite Score
- **Measures with endorsement removed in last cycle (3):**
  - **0264**: Prophylactic Antibiotics (IV) – Antibiotic Timing
  - **0453**: Urinary catheter removed on Postoperative Day 1 (POD 1) or Postoperative Day 2 (POD 2) with day of surgery being day zero
  - **0458**: Pulmonary Function Tests Before Major Anatomic Lung Resection (Pneumonectomy, Lobectomy, or Formal Segmentectomy)



## Surgery Portfolio (cont.)

**By data source*:**

| | |
|---|---|
| Administrative Claims | 27 |
| Electronic Administrative Data/Claims | 5 |
| Electronic Clinical Data: Electronic Health Record | 12 |
| Electronic Clinical Data: Registry | 37 |
| Electronic Clinical Data: Laboratory/Pharmacy/Imaging/Diagnostic | 3 |
| Paper Medical Records | 19 |

**By level of analysis*:**

| | |
|---|---|
| Clinician: Individual | 19 |
| Clinician: Group/Practice | 39 |
| Facility | 62 |
| Integrated Delivery System | 1 |
| Population: National/Regional/State/County/City | 26 |

**By care setting*:**

| | |
|---|---|
| Ambulatory Care : Ambulatory Surgery Center (ASC) | 6 |
| Ambulatory Care : Clinician Office/Clinic | 4 |
| Home Health | 1 |
| Hospital/Acute Care Facility | 58 |



*Note: because many measures are applicable to multiple categories, there is significant overlap in these numbers.

## Activities and Timeline

| Process Step | Timeline |
|---|---|
| **Measure submission deadline** | January 14, 2015 6:00 pm EST |
| **SC member orientation/Q&A webinar** | January 22, 2015 3:00-5:00 pm EST |
| **SC in-person meeting** | March 19-20, 2015 |
| **Post-meeting webinar** | March 27, 2015; 2:00-4:00 pm EST |
| **SC call to review and respond to comments** | June 8, 2015; 1:00-3:00 pm EST |



## Measures Under Review

1) 0115: Risk-Adjusted Surgical Re-exploration
2) 0116: Anti-Platelet Medication at Discharge
3) 0118: Anti-Lipid Treatment Discharge
4) 0120: Risk-Adjusted Operative Mortality for Aortic Valve Replacement (AVR)
5) 0121: Risk-Adjusted Operative Mortality for Mitral Valve (MV) Replacement
6) 0122: Risk-Adjusted Operative Mortality MV Replacement + CABG Surgery
7) 0123: Risk-Adjusted Operative Mortality for Aortic Valve Replacement (AVR) + CABG Surgery
8) 0130: Risk-Adjusted Deep Sternal Wound Infection Rate
9) 0236: Coronary Artery Bypass Graft (CABG): Preoperative Beta-Blocker in Patients with Isolated CABG Surgery
10) 0354: Hip Fracture Mortality Rate (IQI 19)
11) 0360: Esophageal Resection Mortality Rate (IQI 8)
12) 0361: Esophageal Resection Volume (IQI 1)
13) 0465: Perioperative Anti-platelet Therapy for Patients undergoing Carotid Endarterectomy

## Measures Under Review

14) 0533: Postoperative Respiratory Failure Rate (PSI 11)

15) 0696: The STS CABG Composite Score

16) 0732: Surgical Volume for Pediatric and Congenital Heart Surgery: Total Programmatic Volume and Programmatic Volume Stratified by the Five STS-EACTS Mortality Categories

17) 0733: Operative Mortality Stratified by the Five STS-EACTS Mortality Categories

18) 1501: Risk-Adjusted Operative Mortality for Mitral Valve (MV) Repair

19) 1502: Risk-Adjusted Operative Mortality for MV Repair + CABG Surgery

20) 2038: Performing vaginal apical suspension at the time of hysterectomy to address pelvic organ prolapse

21) 2677: Preoperative evaluation for stress urinary incontinence prior to hysterectomy for pelvic organ prolapse

22) 2681: Perioperative Temperature Management

23) 2683: Risk-Adjusted Operative Mortality for Pediatric and Congenital Heart Surgery

24) 2687: Hospital Visits after Hospital Outpatient Surgery

13

## Process Updates

- Staff to perform preliminary analyses of measures
  - Will inform committee review of measures before and at the in-person meeting
  - Analyses to be shared with Committee and developers before meeting
- No Committee workgroup calls
  - Committee members will be assigned specific measures for thorough review and will complete preliminary evaluations surveys before March meeting

## Measure Evaluation – #0119: Risk-Adjusted Operative Mortality for CABG

- **Measure Steward:** The Society of Thoracic Surgeons
- **De.3. Brief Description of Measure:** Percent of patients aged 18 years and older undergoing isolated CABG who die, including both 1) all deaths occurring during the hospitalization in which the CABG was performed, even if after 30 days, and 2) those deaths occurring after discharge from the hospital, but within 30 days of the procedure
- **S.4. Numerator Statement:** Number of patients undergoing isolated CABG who die, including both 1) all deaths occurring during the hospitalization in which the operation was performed, even if after 30 days, and 2) those deaths occurring after discharge from the hospital, but within 30 days of the procedure
- **S.7. Denominator Statement:** All patients undergoing isolated CABG
- **De.1. Measure Type:** Outcome
- **S.23. Data Source:** Electronic Clinical Data : Registry
- **S.26. Level of Analysis:** Clinician : Group/Practice, Facility



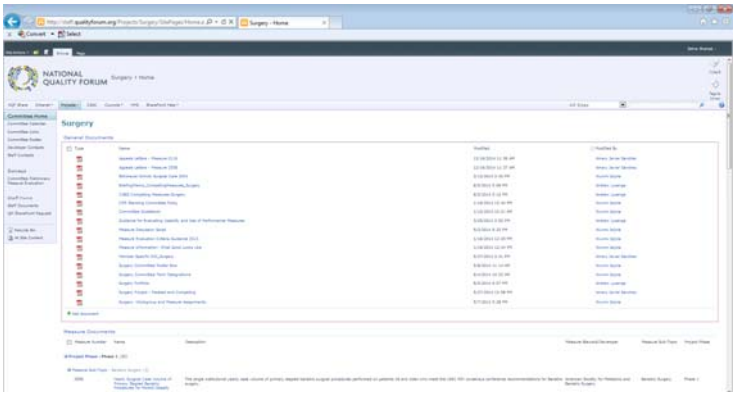
## SharePoint Overview

http://share.qualityforum.org/Projects/Surgery/SitePages/Home.aspx
**[link to Project SharePoint Site]**

- Accessing SharePoint
- Standing Committee Guidebook
- Measure Document Sets
- Meeting and Call Documents
- References
- Survey Tool

## SharePoint Overview

- **Screen shot of homepage:**






## Questions?

## Next Steps for Committee

- Complete your preliminary evaluation surveys: Will be distributed by February 19, 2015 and due on March 11, 2015
- Travel logistics information sent by **February 19, 2015**
- Full Committee meeting: March 19-20, 2015 in Washington, DC



## Project Contact Info

- *Juliet Feldman, jfeldman@qualityforum.org*
- *Alexandra Ogungbemi, aogungbemi@qualityforum.org*
- *Andrew Lyzenga, alyzenga@qualityforum.org*
- *Melinda Murphy, mlmurphy@qualityforum.org*
- NQF Phone: 202-783-1300
- SharePoint site: http://share.qualityforum.org/Projects/Surgery/SitePages/Home.aspx

APPENDIX – Measure Evaluation Overview

NATIONAL QUALITY FORUM

*Please note page numbers denoted correspond to the **Steering Committee Guidebook.**

21

## NQF Measure Evaluation Criteria

- Conditions for Consideration
- Importance to measure and report
- Scientific acceptability of measure properties
- Feasibility
- Use and Usability
- Harmonization & selection of best-in-class

NATIONAL QUALITY FORUM

22

## Major Endorsement Criteria Hierarchy and Rationale (page 32)

- **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 resource use; 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**: 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

**Please note page numbers denoted correspond to the Steering Committee Guidebook.*



## Criterion #1: Importance to Measure & Report (page 36-38)

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

   **1a. Evidence –** the measure focus is evidence-based.

   **1b. Opportunity for Improvement -** demonstration of quality problems and opportunity for improvement, i.e., data demonstrating considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or disparities in care across population groups (pages 41-42)

   **1c. High Priority** – the measure addresses a specific national health goal or priority and/or a high-impact aspect of healthcare. (page 42)

   **1d. Quality construct and rationale (composite measures)**

## 1a Evidence **(page 36-37)**

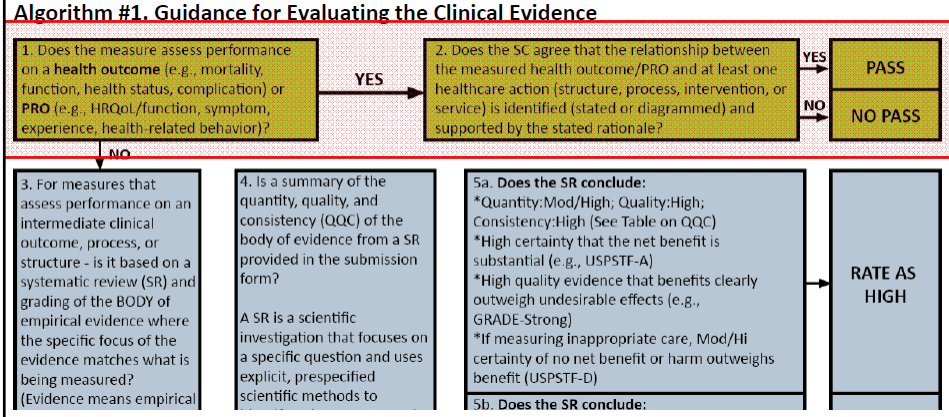
**Requirements for 1a.**

- Outcome measures –a rationale (which often includes evidence) for how the outcome is influenced by healthcare processes or structures.
- Process, intermediate outcome measures - the quantity, quality, and consistency of the body of evidence underlying the measure should demonstrate that the measure focuses on those aspects of care known to influence desired patient outcomes
  - Empiric studies (expert opinion is not evidence)
  - Systematic review and grading of evidence
    - » Clinical Practice Guidelines – variable in approach to evidence review



## Algorithm #1 – page 37

**Algorithm #1. Guidance for Evaluating the Clinical Evidence**



## Criterion # 2: Reliability and Validity – Scientific Acceptability of Measure Properties (page 43 -46)

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

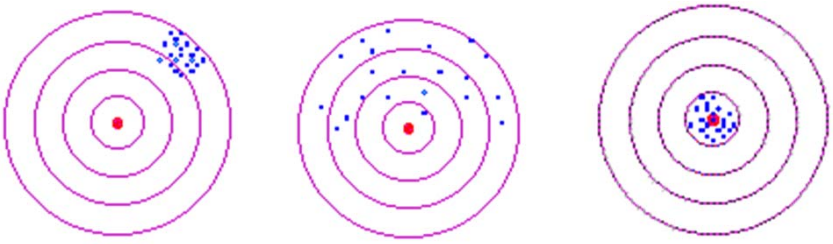
2a. **Reliability (must-pass)**
- 2a1. Precise specifications including exclusions
- 2a2. Reliability testing—data elements or measure score

2b. **Validity (must-pass)**
- 2b1. 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



## Reliability and Validity

**Assume the center of the target is the true score…**



| Reliable Not Valid | Neither Reliable Nor Valid | Both Reliable And Valid |
|---|---|---|
| Consistent, but wrong | Inconsistent & wrong | Consistent & correct |

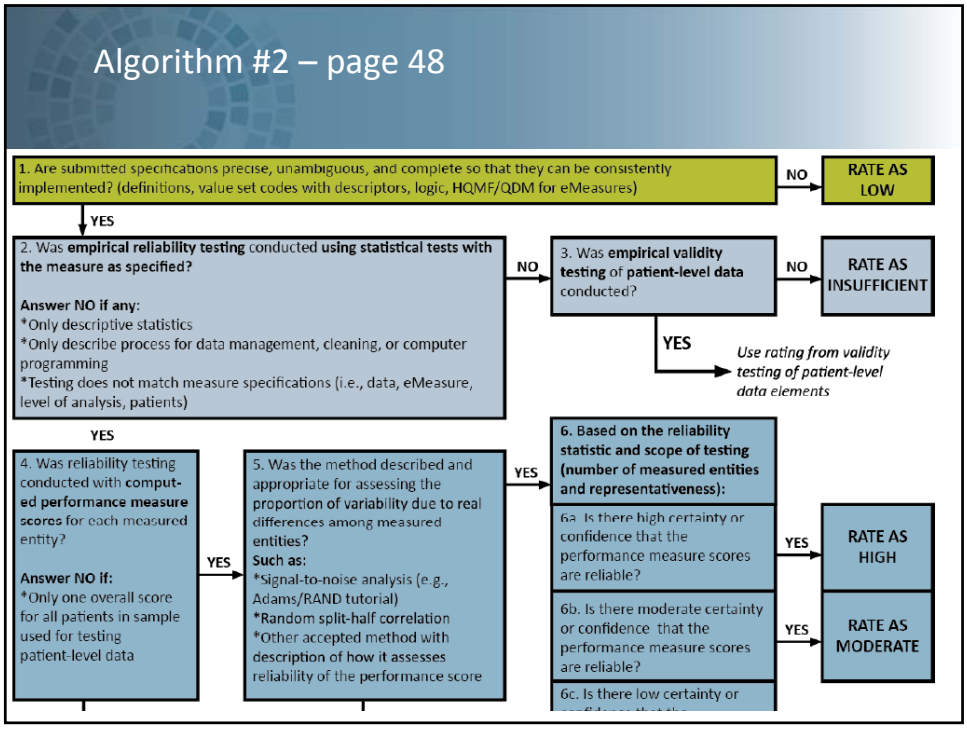
## Measure Testing – (Key Points page 46)

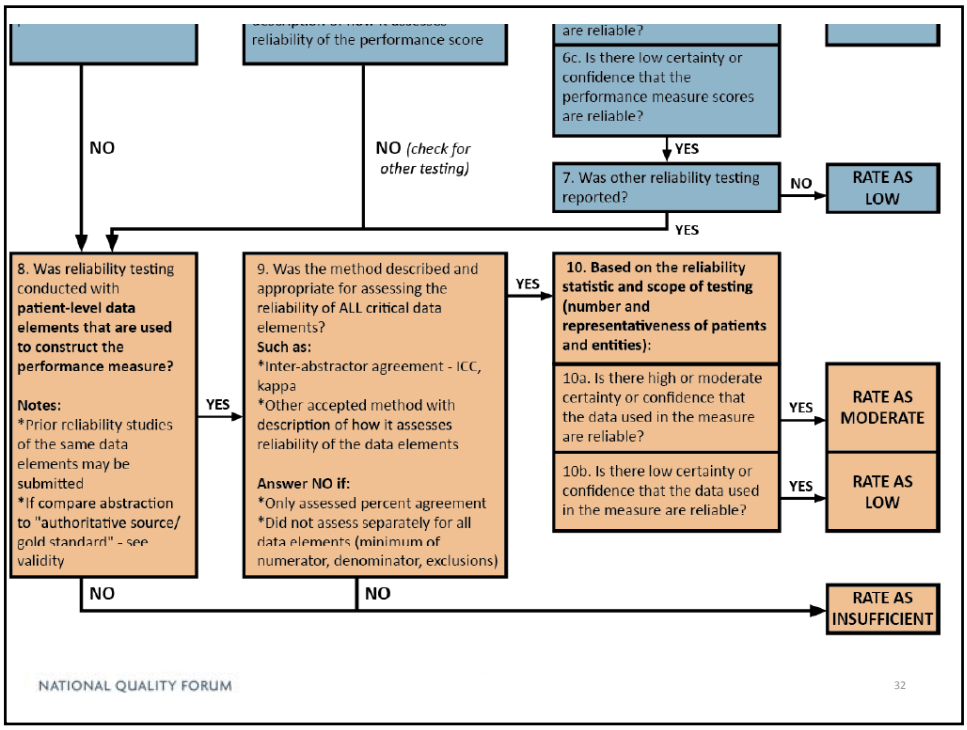
**Empirical analysis** to demonstrate the reliability and validity of the *measure as specified,* including analysis of issues that pose threats to the validity of conclusions about quality of care such as exclusions, risk adjustment/stratification for outcome and resource use measures, methods to identify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



## Reliability Testing (page 46) Key points - page 47

- 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 results are within acceptable norms
- Algorithm #2 – page 48


Algorithm #2 – page 48
1. Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented? (definitions, value set codes with descriptors, logic, HQMF/QDM for eMeasures)
NO
RATE AS LOW
YES
2. Was empirical reliability testing conducted using statistical tests with the measure as specified?
Answer NO if any:
*Only descriptive statistics
*Only describe process for data management, cleaning, or computer programming
*Testing does not match measure specifications (i.e., data, eMeasure, level of analysis, patients)
NO
3. Was empirical validity testing of patient-level data conducted?
NO
RATE AS INSUFFICIENT
YES
Use rating from validity testing of patient-level data elements
YES
4. Was reliability testing conducted with computed performance measure scores for each measured entity?
Answer NO if:
*Only one overall score for all patients in sample used for testing patient-level data
YES
5. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities?
Such as:
*Signal-to-noise analysis (e.g., Adams/RAND tutorial)
*Random split-half correlation
*Other accepted method with description of how it assesses reliability of the performance score
YES
6. Based on the reliability statistic and scope of testing (number of measured entities and representativeness):
6a. Is there high certainty or confidence that the performance measure scores are reliable?
YES
RATE AS HIGH
6b. Is there moderate certainty or confidence that the performance measure scores are reliable?
YES
RATE AS MODERATE
6c. Is there low certainty or


reliability of the performance score
are reliable?
6c. Is there low certainty or confidence that the performance measure scores are reliable?
YES
NO
NO (check for other testing)
7. Was other reliability testing reported?
NO
RATE AS LOW
YES
8. Was reliability testing conducted with patient-level data elements that are used to construct the performance measure?
Notes:
*Prior reliability studies of the same data elements may be submitted
*If compare abstraction to "authoritative source/ gold standard" - see validity
YES
9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?
Such as:
*Inter-abstractor agreement - ICC, kappa
*Other accepted method with description of how it assesses reliability of the data elements
Answer NO if:
*Only assessed percent agreement
*Did not assess separately for all data elements (minimum of numerator, denominator, exclusions)
YES
10. Based on the reliability statistic and scope of testing (number and representativeness of patients and entities):
10a. Is there high or moderate certainty or confidence that the data used in the measure are reliable?
YES
RATE AS MODERATE
10b. Is there low certainty or confidence that the data used in the measure are reliable?
YES
RATE AS LOW
NO
NO
RATE AS INSUFFICIENT
NATIONAL QUALITY FORUM
32

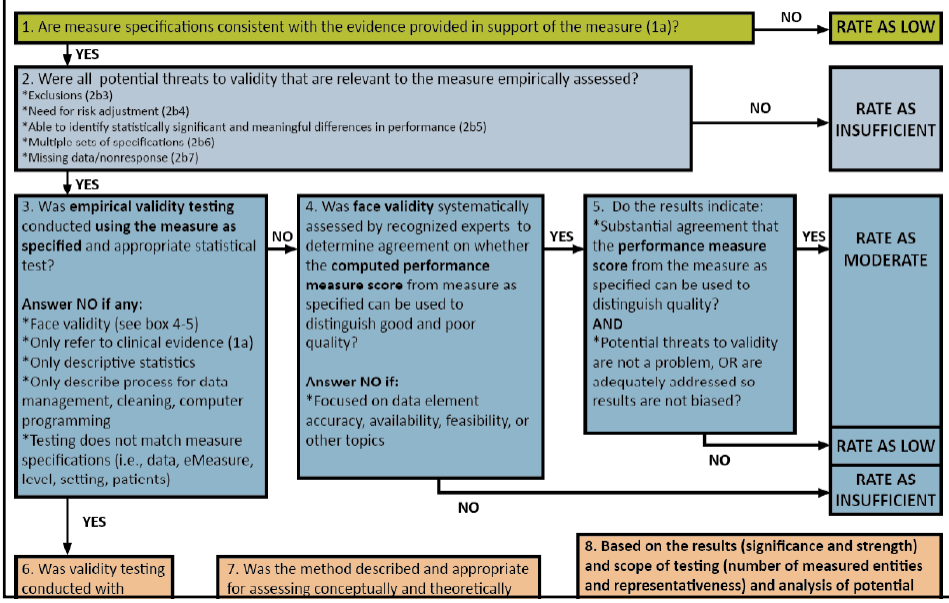
## Validity testing (pages 49- 51)
## Key points – page 51

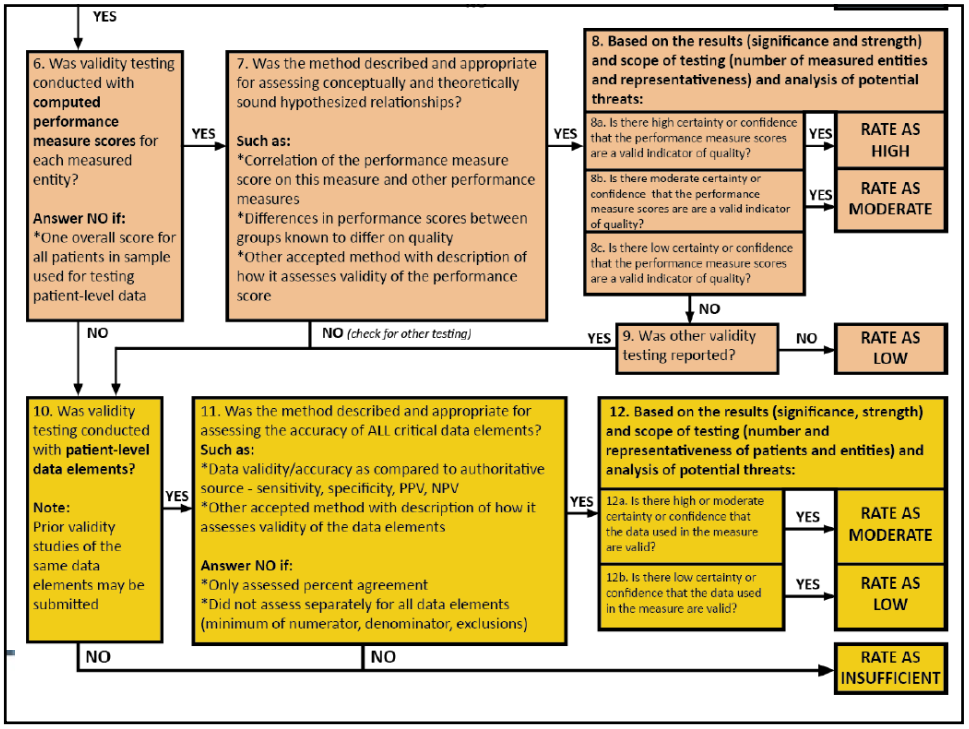
- **Empiric testing**
  - *Measure score* – assesses a hypothesized relationship of the measure results to some other concept; assesses the correctness of conclusions about quality
  - *Data element* – assesses the correctness of the data elements compared to a "gold standard"
- **Face validity**
  - Subjective determination by experts that the measure appears to reflect quality of care



## Algorithm #3 – page 52

**Algorithm #3. Guidance for Evaluating Validity**






## 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 #3: Feasibility (page 53-54)
## Key Points – page 55

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

**3a: Clinical data generated during care process**

**3b: Electronic sources**

**3c: Data collection strategy can be implemented**



## Criterion #4: Usability and Use (page 54)

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.

**4a: Accountability:** 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

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

**4c: 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).

**4d. Transparency:** 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.

## 5. Related or Competing Measures (page 55-56)

**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.