

Episode Grouper Evaluation Criteria

Expert Panel Meeting

February 5 – 6, 2014

Evan Williamson
Ashlie Wilbon
Ann Phillips
Taroon Amin



NATIONAL
QUALITY FORUM

Welcome, Project Purpose, Scope and Timeline



NATIONAL
QUALITY FORUM

Welcome

- Restrooms
 - Exit main conference area, pass elevators, on right.
- Breaks
 - 10:15 – 15 Minutes
 - 12:45 – Lunch Provided by NQF
 - 3:15 – 15 Minutes
- Laptops and Cellphones
 - WiFi Network: “NQF-Guests” Password: “NQFguest”
 - Please mute your cellphone during the meeting
- Orientation to Discussion Guide

NQF Project Staff

- Evan Williamson
 - Project Manager, Performance Measurement
- Ashlie Wilbon
 - Managing Director, Performance Measurement
- Ann Phillips
 - Project Analyst, Performance Measurement
- Taroon Amin
 - Senior Director, Performance Measurement

Project Purpose

1. Define the characteristics and purpose of an episode grouper versus other measurement systems (e.g., risk adjustment systems)
2. Identify the key steps in episode grouper construction
 - Principles/considerations
 - Necessary elements and associated criteria
3. Identify key considerations and implications for endorsing episode groupers
 - Challenges and path forward

Timeline

<i>Process Step</i>	<i>Timeline</i>
<i>Expert Panel member orientation</i>	1/22/14
<i>Background Information Review</i>	1/29/14 – 2/5/14
<i>Expert Panel In-Person Meeting</i>	2/5/14 – 2/6/14
Post-Meeting Conference Calls	3/12/14 12-2pm ET AND 3/19/14 12-2pm ET
Draft report posted for NQF Member and Public Review and Comment	3/24/14 – 4/22/14
SC call to review and respond to comments	5/14/14 from 12-2pm ET
CSAC review and approval	6/3/14 – 6/17/14
Endorsement by the Board	6/18/14 – 6/30/14
Final Report Complete	7/1/14

Project Scope

- This project will focus on principles, considerations and criteria for Episode Grouper construction and evaluation
- This criteria is NOT intended to be used to evaluate a specific grouper/tool/product, but for ANY grouper that would be submitted
- We will NOT be evaluating/selecting “best” approaches or selecting specific tools

Disclosure of Interest

- **Ann Hammersmith, JD - General Counsel**
- Specific disclosures should include:
 - Involvement in the development of an episode grouper system(s);
 - Personal financial arrangement or affiliation with a specific product(s) or service based on a product;
 - Investment in a specific product(s) by your organization; and
 - Employment by or other affiliations with organizations, companies or other entities that own, develop or use episode groupers.

Expert Panel Roster

- Kristine Martin Anderson, MBA (Co-Chair)
- Joseph Cacchione, MD (Co-Chair)
- Stephen Bandeian, MD, JD
- David Bodycombe, MSc, ScD
- Francois de Brantes, MS, MBA
- Dan Dunn, PhD
- Nancy Garrett, PhD
- Jennifer Hobart, MBA, MSc
- David Hopkins, PhD
- Jim Jones, MBA
- Marjorie L King, MD, FACC, MAACVPR
- Mark Levine, MD, FACP
- Jim Loiselle
- Thomas MaCurdy, PhD
- Jelani McLean, PhD, MPA
- David Mirkin, MD
- James Naessens, ScD, MPH
- David Redfearn, PhD
- Andrew Ryan, PhD
- Tamara Simon, MD, MSPH, FAAP
- Christopher Tompkins, PhD

Agenda for the Meeting

Day 1

- Breakfast Buffet
- Welcome, Project Purpose, Scope and Timeline
- Committee Introductions and Disclosure of Interest
- Setting the Stage
- Review of Key Definitions
- Review of Existing NQF Resource Use Measure Evaluation Criteria
- Defining Key Modules for Episode Groupers
- Public and Member Comment

Agenda for the Meeting

Day 1 (*continued*)

- Lunch
- Break-Out Sessions: Constructing and Evaluating an Episode Grouper by Individual Modules
- Review of the Principles for Constructing and Evaluating an Episode Grouper
- Public and Member Comment
- Adjourn
- Expert Panel Dinner (Optional)

Setting the Stage



NATIONAL
QUALITY FORUM

NQF Mission

Board of Directors

Steering Committees

8 Membership Councils

Measures Application
Partnership (MAP)

National Priorities
Partnership (NPP)

CSAC, HITACH

Neutral Convener

Standards Setting
Organization

- 1 Build
Consensus
- 2 Endorse National
Consensus Standards
- 3 Education and
Outreach

Setting the Stage

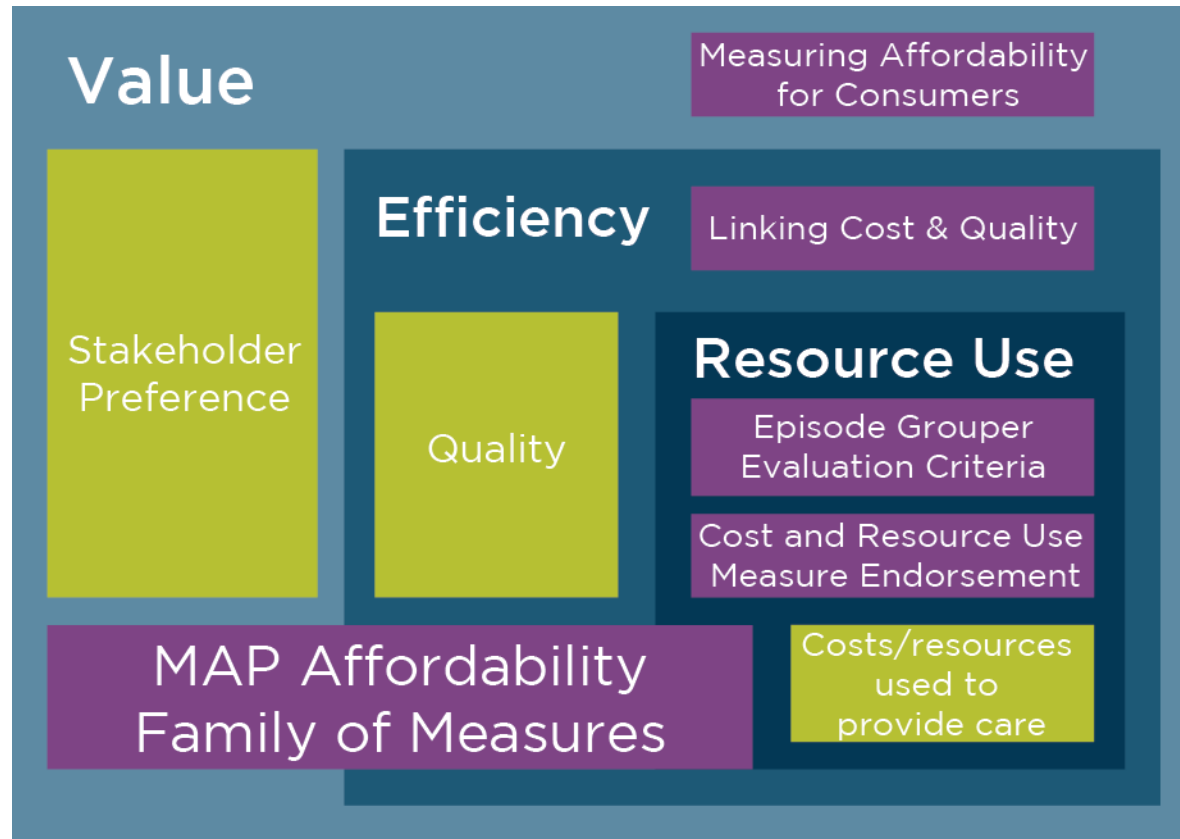
- Why now?
- The Purpose of NQF Endorsement
- Need for National Consensus Standards
- Balancing the Flexibility in Grouping Methods

Current Landscape for Cost Measurement

- The cost imperative
- Policy Implications
 - Legislation for Cost Measurement and Episode Groupers
 - Physician Feedback Program[§]
 - Value-based Payment Modifier[§]
 - Commercial Pay for Performance Programs
 - Potential Bundled Payment and ACO Applications

§ Section 1848(p) of the Social Security Act as established by Section 3003 and 3007 of the Affordable Care Act of 2010 (ACA)

NQF's Current Portfolio of Cost Measurement Work



Current Landscape for Episode Groupers

- Current use of groupers
 - Episode Groupers for Commercial Populations
 - Episode Grouper for Medicare
 - Use of tools varies by region
- Challenges
 - Complexity
 - Transparency
 - Multiple signals for the same providers
 - Market Implications
 - Proprietary Components
- NQF's Role in Evaluation Episode Groupers

Definitions



NATIONAL
QUALITY FORUM

Group Discussion

Grouper Terminology

- Episode
- Episode Grouper
- Risk Adjustment /Case Mix Adjustment
- Measurement System
- Patient/Clinical Data Registries

Group Discussion

Discussion Questions

- How would you describe the purpose and function of an episode grouper?
- What are the various uses of groupers?
- How does an episode grouper differ from a case mix/risk adjuster, or other measurement systems?
- How does an episode grouper differ from clinical data registries or other databases?

Break



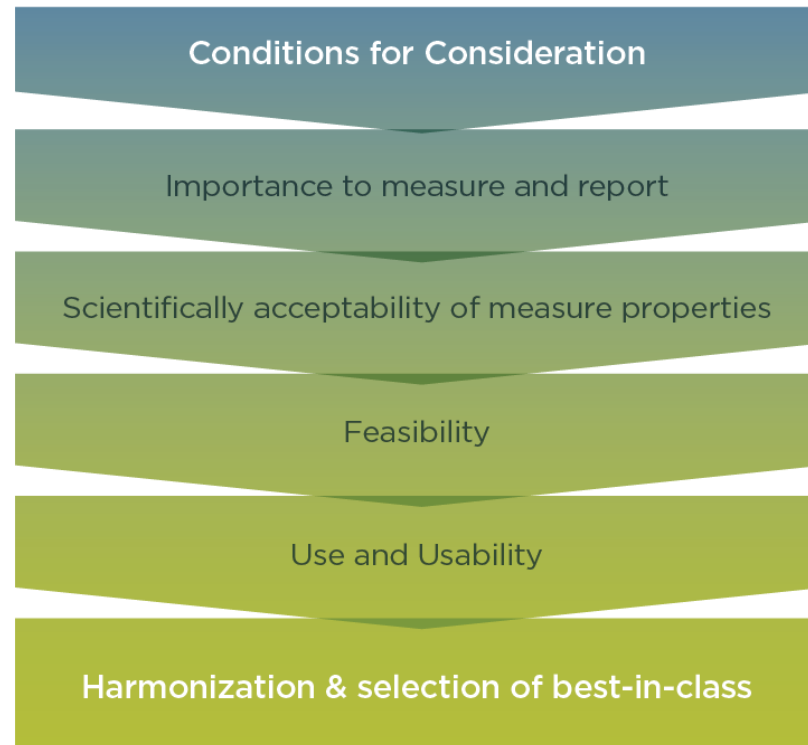
NATIONAL
QUALITY FORUM

Review of Existing NQF Resource Use Measure Evaluation Criteria



NATIONAL
QUALITY FORUM

Review of Existing NQF Resource Use Measure Evaluation Criteria



NQF Measure Criteria

Principles of the Current Measure Criteria

- Some of the key principles that guide the application of these criteria include:
 - There are two must-pass criteria. If a measure does not meet the minimum requirements for *Importance to Measure and Report* or *Scientific Acceptability of Measure Properties*, it cannot be recommended for endorsement and will not be evaluated against the remaining criteria.
 - Subcriteria delineate how to demonstrate that the major criteria are met.
 - The criteria were developed to parallel best practices for measure development.
 - The application of these criteria requires both evidence and expert judgment.
 - The assessment of each criterion is a matter of degree (rather than all or nothing), generally rated on a scale of high, moderate, low, and insufficient.

NQF Measure Criteria

Discussion: Developing Episode Grouper Criteria

- Which of these criteria are relevant to the evaluation of episode groupers? :
 - Can these five major criteria be applied to episode groupers?
 - For the major criteria that apply, how might the subcriteria discussed below apply to episode groupers?
 - Are there other major or subcriteria that should be considered when evaluating groupers?

Importance to Measure and Report

Used to determine:

- If the measure focus (topic) is important to making significant contributions toward understanding healthcare costs for a specific high-impact aspect of healthcare
- There is variation or a demonstrated high-impact aspect of healthcare or overall poor performance.

Subcriteria:

- The measure addresses a specific national health goal/priority or a demonstrated high-impact aspect of healthcare
- There is a demonstrated resource use or cost problem and opportunity for improvement.
- The intent of the measure is clearly described and the types of costs captured (resource service categories) are consistent with the intent.

Scientific Acceptability: Reliability

Used to determine the extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the cost or resources used to deliver care.

- Reliability Subcriteria:
 - Preciseness of the specifications facilitates consistent implementation.
 - Reliability testing results demonstrates the results are repeatable.

Scientific Acceptability: Validity

- Validity:
 - The measure specifications are consistent with the measure intent;
 - Validity testing;
 - Exclusions;
 - Evidence-based risk-adjustment strategy;
 - Adequate discrimination and calibration of the risk model;
 - Statistically significant and practically/clinically meaningful differences in performance;
 - Multiple data sources/methods should produce comparable results;

Feasibility

Used to assess the extent to which the required data are:

- Readily available
- Can be captured without undue burden, and
- can be implemented

Subcriteria:

- The required data elements are routinely generated during care delivery.
- The required data elements are available in electronic sources.
- The data collection strategy can be implemented without undue burden (includes an assessment of financial burden due to the cost of the measures).

Usability and Use

Used to assess the extent to which potential audiences (e.g., consumers, purchasers, providers, policymakers) are using or could use performance results for both accountability and performance improvement.

Subcriteria:

- Current and planned use of the measure in public reporting and accountability application
- The measure has demonstrated improvement of cost/resource performance
- The benefits of the measure outweigh unintended consequences of the measure
- The measure can be deconstructed to facilitate transparency and understanding

Related or Competing Measures

If a measure meets the 4 previous criteria and there are related or competing measures, the measures are compared to determine areas of potential harmonization and/or selection of the best measure.

- Related measures may:
 - Share the same measure type (e.g., per episode, per capita), or
 - measure the same costs/resources (e.g., actual prices paid vs. standard prices, resource service categories), or
 - address the same population (e.g., people with diabetes).
- Competing measures would share all of the characteristics previously listed.

Criteria Discussion

- Can these five major criteria be applied to episode groupers?
- For the major criteria that apply, how might the subcriteria apply to episode groupers?
- Are there other major or subcriteria that should be considered when evaluating groupers?

Defining Key Modules for Episode Groupers



NATIONAL
QUALITY FORUM

General Principles of Episode Grouper Construction

Discussion: Preliminary Modules

- As the Panel reviews each of these modules, please consider the following:
 - Do these modules reflect the major elements of episode grouper construction?
 - Are elements missing?
 - Are there specific items within each module that should be associated with other modules?

Defining Key Modules for Episode Groupers

- Construction Logic
- Clinical Logic
- Adjustments for Comparability

Construction Logic Module

Includes steps used to cluster, group, or assign claims beyond those associated with the measure's clinical logic and an assessment of how the various components of the measure work together. This module includes logic to address the following components:

- Concurrency of clinical events, measure redundancy or overlap, disease interactions
- Complementary Services
- Clinical Hierarchies
- Missing Data
- Identification of Resource Use service categories
- Attribution Approach

Clinical Logic Module

The clinical logic includes steps to identify the condition or event of interest and any clustering of diagnoses or procedures.

- Clinical Logic
- Evidence to Support Clinical Logic
- Trigger and end mechanisms
- Clinical Severity Levels
- Comorbids and Interactions

Adjustments for Comparability Module

These steps are used for adjusting measures to increase comparability, and may include:

- Inclusion and Exclusion Criteria
- Risk Adjustment
- Stratification Details/Variables
- Costing Method
- Scoring

General Principles of Episode Grouper Construction

Discussion: Preliminary Modules

- Do these modules reflect the major elements of episode grouper construction?
 - Are elements missing?
- Are there specific items within each module that should be associated with other modules?

Public and Member Comment



NATIONAL
QUALITY FORUM

Lunch



NATIONAL
QUALITY FORUM

Break-out Discussions:
Principles and Proposed
Criteria for Episode Group
Evaluation



NATIONAL
QUALITY FORUM

Breakout Group Charge

Discuss the principles for constructing a episode grouper and criteria for evaluation:

1. Examine key questions for each module (using the discussion guide)
2. Further define the module and key elements
3. Identify principles and considerations for evaluation
4. Identify the criteria for evaluating the module and its components

Breakout Group Charge

Further, the breakout groups will explore the following key questions for each module:

- What are relevant criteria that may be used to evaluate the module?
- How might the reliability and validity subcriteria be applied to the module?
- Is there specific guidance on how testing should be completed for the module?
- How might the use of the grouper impact the specifications/logic for the module?
- In addition to the description and rationale for each component of the module, what other information would be needed in order to appropriately evaluate this portion of the grouper logic?

Break-out Discussion: Groups

Starting at 1:15pm

Group A – Construction Logic -9A Main Meeting Room

1. Steve Bandien
2. Francios DeBrantes
3. Nancy Garrett
4. Tom MaCurdy
5. Jim Loiselle
6. David Hopkins

Group B – Clinical Logic 9B Conference Room

1. Joe Cacchione
2. Tamara Simon
3. Chris Tompkins
4. Marjorie King
5. David Mirkin
6. James Naessens
7. Jennifer Hobart

Group C – Adjustments for Comparability 8th Fl Conference Room

1. David Bodycombe
2. David Redfearn
3. Jim Jones
4. Mark Levine
5. Jelani McLean
6. Kristine Martin Anderson

*Group A Facilitator:
Taroon Amin*

*Group B Facilitator:
Ashlie Wilbon*

*Group C Facilitator:
Evan Williamson*

Break



NATIONAL
QUALITY FORUM

Review of the Principles for Constructing and Evaluating an Episode Grouper (Part 1)



NATIONAL
QUALITY FORUM

- 
- Insert Slides From Breakout Discussions

Public and Member Comment



NATIONAL
QUALITY FORUM

Adjourn

Dinner: PJ Clarke's – 6:00 PM

1600 K Street NW
Washington, DC 20006



NATIONAL
QUALITY FORUM

Welcome, Goals, Agenda
Review, Recap of Day 1



NATIONAL
QUALITY FORUM

Agenda for the Meeting

Day 2

- Breakfast Buffet
- Welcome, Goals, Agenda Review, Recap of Day 1
- Review of the Principles for Constructing and Evaluating an Episode Grouper
- Finalize Proposed Criteria for Episode Grouper Evaluation
- Public and Member Comment
- Lunch
- Implications for Episode Grouper Endorsement – Challenges and the Path Forward
- Next Steps / Wrap up
- Public and Member Comment
- Adjourn

Recap of Day 1

- Update after Day 1

Review of the Principles for Constructing and Evaluating an Episode Grouper (Part 2)



NATIONAL
QUALITY FORUM

- 
- Insert Slides From Breakout Discussions

Break



NATIONAL
QUALITY FORUM

Finalize Proposed Criteria for Episode Grouper Evaluation



NATIONAL
QUALITY FORUM

[insert slides listing compiled principles and criteria for the modules- agreed upon by the group]

Finalize Proposed Criteria for Episode Grouper Evaluation

- Does evaluation of a grouper require evaluation of BOTH the individual modules and the grouper as a whole?
- Are there criteria that should be applied to the grouper as whole?
- Does the evaluation of a grouper require evaluation of the logic within each of the individual clinical episodes and the grouper overall? Is there a distinction?

Finalize Proposed Criteria for Episode Grouper Evaluation

- What impact does the potential use and implementation of the grouper have on the evaluation of the overall grouper, and potentially its components?
- To what extent does complexity of the episode grouper logic impact its transparency?
- How might the testing of the grouper software differ from the testing of the grouper logic? Is there a distinction?
- What are some examples of how the reliability and validity of a grouper might be tested?
- How might testing for the reliability and validity of grouper logic differ from testing a quality outcome measure? Other types of measures?

Finalize Proposed Criteria for Episode Grouper Evaluation : Reporting

- Should the evaluation of episode groupers include the review of reporting guidelines (e.g., discrimination statistics, sample size, comparative estimates)?
 - Does this depend on the intended use of the grouper?
- Should a minimum sample size be required for the reporting of episode grouper results?

Public and Member Comment



NATIONAL
QUALITY FORUM

Lunch



NATIONAL
QUALITY FORUM

Implications for
Endorsement of Groupers –
Challenges and the Path
Forward



NATIONAL
QUALITY FORUM

Implications of Endorsement

- What are some of the benefits and unintended consequences that might result from the endorsement of episode grouper systems?
- Given that many of the episode groupers that could be considered for endorsement are commercially owned, how might the evaluation and endorsement of these groupers impact the market?
- Should future efforts seek to align the public and private sector in the use of a single endorsed episode grouper system?

Implications of Endorsement

- Given that many groupers allow for user options (that may impact the measure score), is it feasible to endorse a national standard that could be consistently applied across users?
- Given the complexity of episode groupers, is it feasible for a multi-stakeholder group to evaluate them and make endorsement recommendations?
- How can grouper measure results be used in combination with quality measures?
- What are some considerations for the path forward in evaluating and endorsing episode groupers?

Implications of Endorsement – Measure Applications Partnership (MAP)

- What are the implications of NQF endorsement on the use of an episode grouper for specific applications, e.g. value-based payment modifier?
- Is there specific implementation guidance that the Committee would offer future MAP workgroups in the evaluation and selection of episode grouper measures for the Physician Feedback Program and the VBPM?
- The MAP currently evaluates the episode grouper by selecting individual measure concepts within it (e.g. lung cancer condition episode). Can the individual measures within a grouper be selected for implementation in VBPM without considering how costs are assigned among the other co-occurring conditions for an individual patient?
- What are some considerations for the path forward in selecting episode grouper measures for specific applications, e.g. value-based payment modifier?

Next Steps/Wrap up



NATIONAL
QUALITY FORUM

Next Steps

- Questions and Feedback
- Post Meeting Calls (*As needed*)
 - March 12 – 12:00 to 2:00 PM ET
 - March 19 – 12:00 to 2:00 PM ET
- Post Comment Call
 - May 14 – 12:00 to 2:00 PM ET

Public and Member Comment



NATIONAL
QUALITY FORUM

Adjourn



NATIONAL
QUALITY FORUM