



TO: Consensus Standards Approval Committee (CSAC)

FR: Ashlie Wilbon, Elizabeth Carey and Taroon Amin

RE: Episode Grouper Evaluation Criteria Draft Recommendations

DA: July 3, 2014

The CSAC will review the draft recommendations from the Episode Grouper Expert Panel on July 9th during the in-person meeting.

The draft report includes the Expert Panel's draft recommendations for submission elements and evaluation criteria, as well as discussion on the characteristics and challenges of constructing, implementing and endorsing episode groupers. This memo includes a summary of the project, an overview of the Expert Panel's recommendations and themes identified from and responses to the public and member comments.

Accompanying this memo are the following documents:

1. [Episode Grouper Evaluation Criteria Draft Report](#). The draft report has been updated to reflect the changes made following Standing Committee discussion of public and member comments. The complete draft report and supplemental materials are available on the project page.
2. [Comment table](#). Staff has identified themes within the comments received. This table lists 40 comments received and the NQF/Standing Committee responses.

CSAC ACTION REQUIRED

- Approve report and recommendations.
- Discuss operational guidance and implications for NQF endorsement of episode groupers.

BACKGROUND

NQF has undertaken several projects relating to episodes of care and cost measurement including the development of episode of care evaluation framework, endorsement of stand-alone cost and resource use measures, as well as the consideration of measures developed from episode grouper methodology. Despite these efforts, there remained significant gaps in stakeholders' understanding of episode groupers, their function, purpose and use.

In basic terms, episode groupers are software tools used to create condition-specific episodes of care (EOC) as a means of measuring costs, and potentially evaluating quality, most often using administrative claims data. Thus, the term "cost" used in this report generally reflects the payment rendered for services as opposed to other notions of cost, such as the costs of rendering care or opportunity costs borne by the patient. Grouping claims at the EOC level affords the opportunity to evaluate the scope of services delivered to a group of patients for identified clinical conditions. Grouping claims in this way allows users to make meaningful comparisons among providers about the use of resources, and other potential innovative uses, such as new payment models and delivery innovations. Experts note that there is no "gold standard" for how claims are grouped and attributed to a particular condition or

episode, and thus involve subjective judgments that may vary significantly between developers. Further exploration is needed to better understand how to achieve consensus on these approaches, their application and implications for evaluation. As such, the goals of this project are to:

- Define the characteristics and challenges of constructing episode groupers;
- Determine the key elements of episode groupers that should be submitted to NQF for evaluation;
- Establish an initial set of criteria by which episode groupers should be evaluated for NQF endorsement; and
- Identify implications and considerations for NQF-endorsement of episode groupers.

To guide this effort, NQF convened a 21-member Expert Panel (Appendix A in the draft report), comprised of stakeholders representing purchasers, health plans, providers and clinicians with expertise in performance measurement, measurement methodologies, clinical quality improvement, and the development of episode groupers. The Expert Panel gathered for a two-day in-person meeting in Washington, DC on February 5th and 6th, 2014 to discuss the key issues identified above and provide recommendations on the evaluation of episode groupers. The focus of the panel was not specific to a particular grouper or product, but broad in nature such that the recommended criteria could be applied to any episode grouper that may be submitted for evaluation.

DRAFT REPORT

In addition to describing and illustrating the basic functions and outputs of an episode grouper, the report lays out key challenges, advantages and disadvantages of measuring costs using episode groupers. The report also discusses the implications for the evaluation of groupers on the endorsement process and the broader field. The Panel's recommendations focused on submission elements for groupers seeking endorsement as well as proposed evaluation criteria.

The key categories of submission items include: descriptive information on the intent and planned use of the grouper; the clinical logic and data required for grouping claims; and reliability and validity testing, methods and results. The Panel emphasized the importance of understanding the intent and planned use for evaluating potential threats to validity and possible unintended consequences of using the grouper.

The recommended evaluation criteria for episode groupers are based on the standard NQF Measure Evaluation Criteria, and include: scientific acceptability (reliability and validity); feasibility; and usability and use. The Panel did not recommend the application of the importance to measure and report or related and competing criteria.

COMMENTS AND THEIR DISPOSITION

NQF received about 40 comments on the draft report from 12 NQF member organizations and members of the public. To focus the Panel's discussion, NQF staff identified the major themes expressed in the comments:

1. Patient-centered episode of care approach
2. Demonstrating validity of the grouper and addressing threats to validity
3. Distinguishing between grouper, episode, and measure levels of evaluation and endorsement
4. Linking quality to grouper episodes/measures

5. Use of clinical guidelines in developing/defining episodes of care

Panel discussion of these themes is summarized below.

1. Patient-centered episode of care approach

Several commenters supported the patient-centered episode of care approach as described in the draft report, while acknowledging the challenges of using this approach for provider profiling. The draft report also describes an alternative approach, provider-centric episodes.

- *Panel response: The evaluation process and criteria should accommodate the submission of provider, patient, or procedure-focused groupers. The panel agreed that the appropriateness of each of these model designs would depend on the intended use of the grouper, and thus the developer should state its intended use in the evaluation form. However, each of these applications should recognize the patient's full care experience and outcomes.*

2. Demonstrating validity of the grouper

Several commenters acknowledged the importance and challenges of ensuring the validity of the grouper methodology. In particular, identifying the threats to validity was of great importance.

Potential threats to validity may include:

- a. Defining chronic episodes in the presence of other co-occurring conditions: Acute episodes generally have defined time periods related to when the condition began and ended, while chronic conditions are generally defined by a pre-determined time window. Commenters agreed that a one year timeframe for chronic conditions seems appropriate, but cautioned about the challenges of identifying and attributing care for patients with multiple chronic conditions, complications, or sequela.
 - b. Data Limitations: Most episode grouper software currently relies on administrative claims data to identify episodes of care. Several commenters noted the limitations of claims-derived data, and urged the inclusion of clinical data from electronic health records or registries for grouping episodes. Some commenters highlighted more general data aggregation and data quality issues, while other focused more specifically on the impact the transition to ICD-10 will have on episode groupers and on administrative claims-based measures.
 - c. Risk and Severity Adjustment: A few commenters suggested the Panel provide additional guidance on how risk adjustment and patient severity should be addressed when evaluating an episode grouper. Given the various ways in which severity and risk adjustments can be applied within a grouper system, additional guidance and clarification on these approaches and the implications should be explored.
- *Panel Response: To the extent that these concepts are not already included in the recommendations for episode grouper submission elements, the Panel emphasized the need for the inclusion of detailed descriptions of the following concepts:*
 - *Statement of grouper intent and purpose, and description of the logic and functionality.*
 - *Description of the risk and severity model and methodology with justification for the choice of model and methodology.*
 - *Method by which the grouper manages multi-morbidity and outliers.*

- *Diagnoses or procedures that are excluded from the grouper and the reasoning behind these exclusions.*
- *Description of any user configurable features and their potential effects on grouper output.*

3. Distinguishing between grouper, episode, and measure levels of evaluation and endorsement

A few commenters noted that more clarity is needed regarding the levels of evaluation that may be required when evaluating a grouper. The Panel's current recommendations note that the grouper logic, the episodes, and measures that result from episode groupings should be subject to evaluation.

- *Panel Response: The Panel affirmed that these three components are important to evaluate; the evaluation of the grouper and the episodes should follow the recommendations outlined in the report. The approach for evaluating measures that result from groupers will need further exploration to determine whether the existing resource use measure evaluation criteria is sufficient and if the Resource Use Standing Committee is the appropriate expert body to facilitate the evaluation process.*

4. Linking quality to grouper episodes/measures

While quality measurement is not typically a key feature of an episode grouper, the grouper can enable the evaluation of both cost and quality signals. These signals are critical to understanding value and efficiency. Several commenters noted the importance of assessing utilization and costs in combination with quality measures. One commenter cautioned that assessing utilization and costs in isolation is likely to result in unintended consequences, especially for vulnerable populations and individuals with multiple co-occurring conditions.

- *Panel Response: The Panel agreed that the linkage of episode grouper-based measures to quality measures continues to be the goal. Given the various approaches to episode grouping, the linkages to quality measures may not always be done within the system.*

5. Use of clinical guidelines in developing/defining episodes of care

Commenters emphasized that episode groupers should be grounded in sound process and methods. This includes ensuring the methods and evidence base used in defining both the episode and grouper are transparent; time dimensions of the clinical episode are clear and objective; accounting for variations in the treatment population; and linking the episode to relevant, endorsed measures of clinical quality. Episode construction should include assumptions about current clinical practice guidelines, appropriate use criteria, and patient preference-sensitive interventions and technologies. The report does not currently address this concern in any detail.

- *Panel Response: The Panel agreed that the use of clinical guidelines in the development of episode groupers is important and should be integrated where possible. However, many groupers do not solely count claims or resources used based on "appropriate" care that would be outlined in guidelines, but rather sort all eligible claims and attribute them based on their relevance to the clinical condition.*

CSAC DISCUSSION: STRATEGIC AND OPERATIONAL GUIDANCE

This phase of work by the Expert Panel raises several issues for consideration related to NQF's strategic goals and operationalization of the evaluation and endorsement of episode groupers. Issues for consideration include:

- **Implications for endorsement:** While most of NQF's efforts to date have focused on the endorsement of measures, there is also a desire to explore the endorsement of measurement systems. Episode groupers are considered a type of measurement system by some and are unique in that they are built to inherently account for user inputs and enable flexibility. Given the variability of data inputs, user specifications and use cases, a single episode grouper can have significant variability in its outputs (i.e., episodes and related measures) and therefore may not produce consistent or comparable results. NQF endorsement has traditionally represented the consensus-based, multi-stakeholder approval of a standard that can be implemented consistently and that is comparable across measured entities. Based on this characterization, the endorsement of episode groupers would challenge this notion. Given the current and future state of NQF:

 - What would "endorsement" of a grouper mean? Is endorsement the appropriate mechanism for vetting and approving groupers?
- **Capacity and process:** The Expert Panel's recommendations for the evaluation of groupers involved evaluation at three different levels: 1) the grouper decision logic, 2) the episodes that result from the grouping, and 3) the performance measures that are developed from the episodes. Given the volumes of data and information that would be required to evaluate an episode grouper, the inherent complexity of the tools, potential time limitations and the limited pool of expertise for this work, operationalizing the evaluation process presents various challenges. In particular, the capacity of both NQF staff and volunteer experts would be a barrier to effective evaluation at this point in time. In order to mitigate these concerns, the process for evaluating episode groupers may require a significant departure from our standard process for evaluating performance measures. Addressing these concerns may have implications for the use of a multi-stakeholder body for assisting in the evaluation, which could require extensive education and training to undertake this type of evaluation. Further, a call for episode groupers could lead to increased evaluation burden if, in fact, multiple groupers were submitted in response.

 - What strategies or guidance should be considered when determining the process for evaluating episode groupers?
 - What strategies for accommodating the gaps in NQF and expert capacity should be considered in order to implement an episode grouper evaluation process?
- **Endorsement of commercial groupers:** While NQF has reviewed and endorsed measures associated with (proprietary) commercial product lines in the past, it has yet to endorse an entire product suite or software system. The Panel expressed concern that the impact of NQF-endorsement on commercial episode grouper products could have unintended consequences for the market; NQF approval for one or some products versus others could sway the trends in the investment of some products over others. While the impetus for the endorsement of a publicly available episode grouper is evident, it is less clear as to whether the owners of commercial episode groupers will find benefit in seeking NQF-endorsement. Further, the evaluation of commercial episode groupers would require additional process considerations. Given these factors, additional guidance and direction is needed as to whether the endorsement of commercial episode groupers is in alignment with NQF's strategic vision and goals.

 - Should both commercial and public groupers be considered for NQF endorsement?



NEXT STEPS:

Using the guidance and recommendations from both the CSAC and the Expert Panel, NQF staff will develop a proposed process for the review and evaluation of episode groupers. Given the need to further explore the implications of the evaluation of measures (in addition to the grouper and the episodes) that result from grouper episodes, NQF staff will be seeking to continue this work to better understand the implications for evaluation of these types of measures. In particular, we seek to better understand how the evaluation of episode grouper measures might differ from the evaluation of stand-alone cost and resource use measures on which our current endorsement efforts focus.