

NATIONAL QUALITY FORUM

TO: Composites Technical Expert Panel
FR: Karen Pace, Senior Director
Karen Johnson, Senior Director
Elisa Munthali, Senior Project Manager
SU: Composite Measure Evaluation Guidance —Conference Call to Finalize Report
DA: February 14, 2013

Thank you for all of the feedback that you have provided so that we can begin to finalize the recommendations and report for the Composites Measure Evaluation Guidance project!

The Composites Technical Expert Panel will meet via conference call on Thursday, February 21, 2013. The purpose of this call is to:

- Review and discuss TEP feedback on draft report.
- Finalize recommendations and report.

Please let us know if you have any questions.

Expert Panel Action:

1. Review this briefing memo.
2. Review the TEP survey responses.
3. Review the latest version of the report. This is the posted report that includes redline changes that were sent to the TEP on 12/14 as well as additional changes.
4. Be prepared to sign off on the report or raise any unresolved issues.

Please use the following information to access the conference call line and online webinar:

Date/Time: Thursday, February 21, 2013, 3:00-5:00 pm ET

Dial-in number: 888-799-5160

Confirmation Code: 91821791

Webinar: <http://ngf.commpartners.com/se/Rd/Mt.aspx?170900>

The phone line will be open. Please place your phone on mute when not speaking. Do not put your phone on hold during the call.

NATIONAL QUALITY FORUM

On the January 3, 2013 post-comment conference call, the TEP indicated it needed more time to finish review of the suggested modifications to the report based on comments from NQF members, the public, TEP members, and the CSAC. As a result, the project timeline was delayed and the TEP was asked to send comments on the 12/31/12 draft and respond to a survey on the key issues. Eleven TEP members responded to the survey; NQF did not receive any suggested changes to the 12/31/12 draft report.

A spreadsheet with the TEP's responses to the survey is available for review and is posted on the project SharePoint

site: <http://share.qualityforum.org/Projects/Composite%20Measures%20Guidance%20Evaluation/SitePages/Home.aspx>. Below, we identify the major results from the TEP survey.

Definition (report p. 4)

There was agreement with the definition and a suggestion for a minor change, as follows:

A composite performance measure is a combination of two or more component measures, each of which individually reflects quality of care, into a single performance measure with a single score.

Quality Construct (report p. 9-10)

Because we will be requiring the quality construct to be submitted and evaluated in response to criterion 1d, the term should be defined. The TEP essentially agreed with the suggested language and offered some suggestions. Following is a revised version.

Quality of care is an abstract concept that is measured using observed variables. Composite measures are complex, multidimensional, and represent a higher order construct than the individual measures. The composite measure quality construct is a hypothetical concept of quality that includes:

- *the overall area of quality (e.g., quality of CABG surgery);*
- *the included component measures (e.g., pre-operative beta blockade; CABG using internal mammary artery; CABG risk-adjusted operative mortality);*
- *representativeness of the included component measures;*
- *the conceptual relationships between each component and the overall composite (e.g., components cause or define quality, components caused by or reflect quality); and*
- *the relationships among the component measures (e.g., correlated or not, process leads to outcome).*

Guidance for Criteria on Reliability and Validity (report p. 11-12; 15-16)

The TEP agreed with the guidance for reliability and validity testing of the composite measure, but suggested that requiring signal-to-noise analysis for reliability was too prescriptive and narrow. Instead, the text was modified so that signal-to-noise analysis is used as an example of one method of reliability testing. Someone questioned whether validity testing of the composite would ever be possible or even necessary if the component measures were valid. The prior TEP discussions noted that even if the component measures were reliable and valid, the construction of the composite may not result in reliable and valid measurement and the focus for testing should be on the composite measure as a whole, not the components. Validity testing is not restricted to testing against some criterion measure, which is unlikely to exist. More explanation was added to the guidance.

NATIONAL QUALITY FORUM

2a2. For composite performance measures, reliability must be demonstrated for the composite measure score.

Testing should demonstrate that measurement error is minimal relative to the quality signal. Examples of testing include signal-to-noise analysis {RAND}, interunit reliability {Zaslavsky}, intraclass correlation coefficient, **other?**

Reliability of the individual component measures is not sufficient, and in some cases, component measures that are not independently reliable can contribute to reliability of the composite measure.

2b2. For composite performance measures, validity should be demonstrated for the composite measure score. If not feasible at the time of initial endorsement, validity of the component measures must meet NQF criteria, and by endorsement maintenance, validity of the composite performance measure must be demonstrated. It is unlikely that a “gold standard” criterion exists, so validity testing generally will focus on construct validation – testing hypotheses based on the theory of the construct. Examples include testing the correlation with measures hypothesized to be related or not related; testing the difference in scores between groups known to differ on quality assessed by some other measure.

Composite-Specific Evaluation Criteria

The TEP agreed on the inclusion of the two composite-specific evaluation criteria. Members also agreed that these would be must-pass subcriteria for composite performance measures.

1d. Under Importance to Measure and Report (report p. 9-10, 14)

A few edits were made to be consistent with the above description of the quality construct. Someone suggested that the model be identified. In prior discussions, the TEP had opted to not use the model terms and instead focus on the relationship of the component measures to one another and to the overall composite.

Composite 1d. For composite performance measures, the following must be clearly stated and logical:

- 1) The quality construct, including the overall area of quality; included component measures; representativeness of the component measures; and the relationship of the component measures to the overall composite and to each other; and
- 2) The rationale for constructing a composite measure, including how the composite provides a distinctive or additive value over the component measures individually; and
- 3) How the aggregation and weighting of the component measures are consistent with and representative of the stated quality construct and rationale.

2d. Under Scientific Acceptability of Measure Properties (report p. 12, 17)

The TEP agreed with the wording of 2d. However, some members questioned whether empirical analyses should be required, if they are relevant for all types of composites, or necessary if reliability and validity of the composite measure was demonstrated. A member pointed out that these criteria are an extension of reliability and validity. Although that is the case, a composite-specific criterion helps communicate expectations. It is not unlike having the risk adjustment criterion that applies only to

NATIONAL QUALITY FORUM

outcome measures and also is an extension of validity. Analysis of missing data may be appropriate for all performance measures but the issue is compounded for composite measures and it is not currently in the NQF criteria.

We did not receive any additional examples of analyses and if we cannot describe what we would expect, then the criterion may need to be revised. If empirical analyses are not needed then is 1d along with the usual criteria for reliability and validity sufficient? Please review the criterion and examples.

Composite 2d. For composite performance measures, empirical analyses to support the composite construction and demonstrate that:

- 1) *the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and*
- 2) *the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible; and*
- 3) *the extent of missing data and how the specified handling of missing data minimizes bias (i.e., achieves scores that are an accurate reflection of quality).*

Examples of analyses:

1) **If components are correlated** - *analyses based on shared variation (e.g., factor analysis, item-total correlation, inter-item correlation).*

If components are not correlated - *analyses demonstrating the contribution of each component to the composite score; correlation of the individual component measures to a common outcome measure*

If empirical analysis does not provide an adequate result, other justification must be provided.

2) *Ideally, sensitivity analyses of the effect of considered aggregation and weighting rules and rationale for selected rules; at a minimum, a discussion of pros and cons of considered approaches and rationale for selected rules*

3) *Overall frequency of missing data and distribution across providers*

Ideally, sensitivity analysis of effect of various rules for handling missing data and rationale for selected rules; at a minimum, a discussion of pros and cons of considered approaches and rationale for selected rules

Identification of Composite Performance Measures (report p. 4-6)

For purposes of NQF measure submission, evaluation, and endorsement, only the types of measures with substantial agreement by the TEP will be identified as composites. Lack of consensus indicates that we are not ready to make a recommendation that will trigger the review of a composite. Additional text was added indicating that this will need to be reviewed again, as was a section discussing the issues with specific types of measures. The TEP did not reach agreement on identifying two types of measures as composites and they appear in the list of measures not considered composted: 1) measures of one concept that combine information from a provider with information on average performance, whether an overall average or an average based on some grouping, should be considered composites; and 2) measures including multiple linked steps in one care process assessed for each patient.

NATIONAL QUALITY FORUM

The following will be classified as composite performance measures requiring submission of specific composite information and evaluation for purposes of NQF endorsement.

- *Measures with two or more individual performance measure scores combined into one score for a provider.*
- *Measures with two or more individual measure components assessed for each patient, then aggregated into one score for a provider. These include all-or-none measures (e.g., all essential care processes received) or any-or-none measures (e.g., any or none of a list of complications).*

*The following will **not** be classified as composite performance measures that require specific composite information and evaluation for purposes of NQF endorsement.*

- *Single performance measures, even if the data are patient scores from a composite instrument or scale (e.g., single performance measure on communication with doctors, computed as the percentage of patients with an average score for four survey questions about communication with doctors is equal or greater than 3).*
- *Measures with multiple measure components assessed for each patient, but result in multiple scores for a provider, rather than a single score. These generally should be submitted as separate measures and indicated as paired/grouped measures.*
- *Measures of multiple linked steps in one care process assessed for each patient. These measures focus on one care process (e.g., immunization) but may include multiple steps (e.g., assess immunization status, counsel patient, and administer vaccination). These are distinguished from all-or-none composites that capture multiple care processes (e.g., foot care, eye care, glucose control).*
- *Single performance measures of one concept (e.g., mortality) but specified with a method or adjustment that combines information from the provider with information on average performance of all providers or groups of providers.*

Decisionmaking Context as Part of Rationale (report p. 10)

There were two questions on the survey about the decisionmaking context – one about the text in the report and one on whether the TEP wanted to make a recommendation to NQF to consider endorsement for specific accountability applications rather than the current policy of endorsement for use in both quality improvement and any accountability application. Two TEP members were completely against the recommendation. Several TEP members were unable to access the report to review the text, though 4 completely agreed and 4 indicated they could live with it. Ultimately, endorsement for specific accountability applications represents a departure from current NQF policy and would require CSAC and Board review and approval. Based on the comments received, we made some revisions and noted the various positions. We also moved the examples into footnotes to make this section more concise.

The TEP acknowledged that NQF endorses performance measures intended for both accountability and performance improvement and does not endorse measures for a specific accountability application (e.g., payment vs. public reporting). However, the TEP discussed that the decisionmaking context could influence the composite measure construction, particularly the component measures included or aggregation and weighting rules. (1) The decision-making context also could influence whether a composite measure is more useful than individual performance measures. (2) Additionally, multiple composite measures for the same quality construct, even if addressing different decisionmaking motivations, will trigger an evaluation of competing measures and the rationales may be an important

NATIONAL QUALITY FORUM

aspect of determining whether multiple endorsed measures are justified. Therefore, if relevant, the rationale for the composite could include the intended decision-making context (e.g., select a provider for surgery, payment incentives to direct resources for improvement).

Some TEP members thought the decisionmaking context is a unique aspect of composite performance measures where choices about which components are included and how they are aggregated and weighted can affect interpretation and use. Other members expressed concern that it is inconsistent with NQF's current policy to endorse measures suitable for both performance improvement and any accountability application, or that it is unnecessary because all composites should be a valid reflection of quality.

¹ *For example, hospital performance on two related sets of measures (A and B) may be important to patients, but failure on group A measures may entail additional costs to the hospital (e.g., longer mean LOS for Medicare fee-for-service patients) whereas failure on group B measures may not entail such additional costs. Composites intended to inform patient choice should include both sets of measures, whereas a pay-for-performance program might use a composite limited to B measures, because the hospital already has a financial incentive to improve on A measures, and therefore the financial reward should be targeted to stimulate improvement on B measures.*

¹ *For example, a composite performance measure that includes multiple surgical mortality measures may be useful for assessing overall surgical quality, whereas the individual performance measures are more useful for selecting a hospital for a specific surgical procedure.*